

دانشکده تجارت و بازرگانی

ارائه مدل پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل و نقل

رشته اقتصاد و تجارت الکترونیک

استاد راهنما:

دکتر ایوب محمدیان

دکتر الناز نصیرزاده

استاد مشاور:

دکتر نوید محمدی

**دانشجو:**



مجوز بهره‌برداري از پايان‌نامه

بهره‌برداري از اين پايان‌نامه در چهارچوب مقررات كتابخانه و با توجه به محدوديتي كه توسط استاد راهنما به شرح زير تعيين مي‌شود، بلامانع است:

🞎 بهره‌برداري از اين پايان‌نامه/ رساله براي همگان بلامانع است.

🞎 بهره‌برداري از اين پايان‌نامه/ رساله با اخذ مجوز از استاد راهنما، بلامانع است.

🞎 بهره‌برداري از اين پايان‌نامه/ رساله تا تاريخ .................................... ممنوع است.

نام استاد يا اساتيد راهنما:

تاريخ:

امضا:

تشكر و قدرداني: (اختياري)

..........................................................................................................

چکيده

سرعت پیشرفت‌ هوش مصنوعی و یادگیری ماشین فرصت‌های بی‌سابقه‌ای را برای ارتقای عملکرد صنایع و مشاغل مختلف، از جمله بخش حمل‌ونقل فراهم می‌کند. در این زمینه، هوش مصنوعی برای غلبه بر چالش‌های مختلفی مانند آلودگی محیط زیست، افزایش تقاضای سفر، خودروهای خودران و نگرانی‌های ایمنی رانندگی به کار گرفته شده است و در سال‌های اخیر کسب و کارهای گوناگونی برای گسترش این حوزه شکل گرفته‌اند. پیش‌بینی موفقیت استارت‌آپ‌ها یک وظیفه حیاتی برای کارآفرینان و سرمایه‌گذاران استارت‌آپ است و این توانایی یک مزیت رقابتی ارزشمند برای سرمایه‌گذاران خطرپذیر که در جستجوی فرصت مناسب سرمایه‌گذاری هستند، ایجاد می‌کند. در این راستا مهم است مطالعاتی در خصوص موفقیت استارتاپ‌ها صورت بگیرد. در این پژوهش موفقیت استارتاپ‌های هوش مصنوعی در حوزه حمل و نقل بررسی خواهد شد. در این تحقیق برای جمع آوری داده از پایگاه داده وبسایت کرانچ بیس استفاده خواهد شد. این وبسایت اطلاعات دقیقی در مورد استارتاپ‌ها، سرمایه‌ی جذب شده و روندهای صنعت ارائه می‌دهد. این مطالعه با بهره‌گیری از انبوه داده‌های موجود در وبسایت کرانچ بیس، تجزیه و تحلیلی سیستماتیک برای بررسی ویژگی‌های مختلفی مانند دورهای تأمین مالی، ادغام و تملیک، عرضه اولیه سهام و روندهای بازار مرتبط با استارت‌آپ‌های هوش مصنوعی در صنعت حمل‌ونقل را ارائه خواهد داد. برای این منظور به دنبال ایجاد یک مدل پیش‌بینی‌کننده، از طریق یادگیری نظارت‌شده و با به کارگیری الگوریتم‌های داده‌کاوی مانند درخت تصمیم، شبکه‌های عصبی مصنوعی و جنگل تصادفی ، برای پیش بینی موفقیت استارتاپ‌های فعال این حوزه هستیم. این الگوریتم‌ها توانایی تجزیه و تحلیل الگوهای پیچیده در داده‌ها، کشف روابط پنهان میان متغیرها و پیش‌بینی دقیق در مورد آینده استارتاپ‌ها را ارائه خواهند داد. هدف این مطالعه، ایجاد بینشی دقیق در این حوزه است که به پیشرفت و پایداری سرمایه‌گذاری‌های نوآورانه کمک خواهند کرد.

واژه‌هاي كليدي:هوش مصنوعی، مدل پیش بینی موفقیت، صنعت حمل و نقل، داده کاوی

فهرست مطالب

[فصل1 کلیات و مفاهیم اولیه 11](#_Toc191112111)

[1-1- مقدمه 12](#_Toc191112112)

[1-2- معرفی موضوع 14](#_Toc191112113)

[1-3- بیان مساله 15](#_Toc191112114)

[1-4- اهداف تحقیق 16](#_Toc191112115)

[1-5- سوالات تحقیق 17](#_Toc191112116)

[1-6- ضرورت انجام تحقیق 17](#_Toc191112117)

[1-7- روش تحقیق **Error! Bookmark not defined.**](#_Toc191112118)

[1-8- سهم دانش افزایی و نوآوری تحقیق 20](#_Toc191112119)

[1-9- واژگان تحقیق 21](#_Toc191112120)

[1-10- جمع بندی و ساختار تحقیق 23](#_Toc191112121)

[فصل2 مروري بر منابع 24](#_Toc191112122)

[2-1- مقدمه 25](#_Toc191112123)

[2-2- مباني نظري 26](#_Toc191112124)

[2-2-1. هوش مصنوعی 26](#_Toc191112125)

[2-2-2. یادگیری ماشین و داده کاوی 27](#_Toc191112126)

[2-2-3. الگوریتم‌های داده کاوی 28](#_Toc191112127)

[2-2-4. استارتاپ 29](#_Toc191112128)

[2-2-5. دوره‌های تامین مالی 30](#_Toc191112129)

[2-2-6. موفقیت کسب و کار 31](#_Toc191112130)

[2-2-7. هوش مصنوعی در صنعت حمل و نقل 32](#_Toc191112131)

[2-2-8. صنعت حمل و نقل 33](#_Toc191112132)

[2-2-9. هوش مصنوعی در صنعت حمل و نقل 34](#_Toc191112133)

[2-2-10. کرانچ بیس 35](#_Toc191112134)

[2-3- پیشینه پژوهش 36](#_Toc191112135)

[2-4- شکاف‌های تحقیقات پیشین 47](#_Toc191112136)

[2-5- جمع بندی 49](#_Toc191112137)

[فصل3 روش تحقیق 51](#_Toc191112138)

[3-1- مقدمه 52](#_Toc191112139)

[3-2- روش تحقیق 53](#_Toc191112140)

[3-2-1. علت انتخاب روش 53](#_Toc191112141)

[3-2-2. تشريح كامل روش تحقيق 54](#_Toc191112142)

[فصل4 نتایج و تفسیر آن‌ها 73](#_Toc191112143)

[4-1- مقدمه 74](#_Toc191112144)

[4-2- توصیف داده‌ها 75](#_Toc191112145)

[4-3- آمار توصیفی 88](#_Toc191112146)

[4-3-1. حضور فعال در شبکه‌های اجتماعی 89](#_Toc191112147)

[4-3-2. کل سرمایه‌گذاری سهام 90](#_Toc191112148)

[4-3-3. تعداد سرمایه‌گذاران 91](#_Toc191112149)

[4-3-4. درآمد تخمینی 92](#_Toc191112150)

[4-3-5. تعداد بنیان‌گذاران 94](#_Toc191112151)

[4-3-6. بازدید از صفحه 95](#_Toc191112152)

[4-3-7. تعداد فناوری‌های فعال 96](#_Toc191112153)

[4-3-8. تعداد اپلیکیشن‌ها 98](#_Toc191112154)

[4-3-9. تعداد محصولات فعال 99](#_Toc191112155)

[4-3-10. تعداد پتنت‌های ثبت شده 100](#_Toc191112156)

[4-3-11. تعداد علائم تجاری ثبت شده 101](#_Toc191112157)

[4-3-12. هزینه‌های فناوری اطلاعات 102](#_Toc191112158)

[4-3-13. سن شرکت 104](#_Toc191112159)

[4-3-14. تعداد کارکنان 105](#_Toc191112160)

[4-4- نتایج داده‌ها 108](#_Toc191112161)

[4-4-1. عوامل موثر بر موفقیت شرکت‌های هوش مصنوعی حوزه حمل و نقل 108](#_Toc191112162)

[4-4-2. موفقیت صنایع مختلف حوزه حمل و نقل **Error! Bookmark not defined.**](#_Toc191112163)

[4-4-3. الگوریتم‌های داده کاوی 114](#_Toc191112164)

[ ماتریس سردرگمی: 115](#_Toc191112165)

[4-5- جمع‌بندی 118](#_Toc191112166)

[فصل5 120](#_Toc191112167)

[جمع‌بندي و پيشنهادها 120](#_Toc191112168)

[5-1- مقدمه 121](#_Toc191112169)

[5-2- پاسخ به سوالات پژوهش 122](#_Toc191112170)

[5-3- مقایسه با تحقیقات گذشته 125](#_Toc191112171)

[5-4- نوآوری پژوهش 127](#_Toc191112172)

[5-5- بحث 128](#_Toc191112173)

[5-6- محدودیت‌های پژوهش 130](#_Toc191112174)

[5-7- پیشنهادات برای تحقیقات آتی 130](#_Toc191112175)

[5-8- پیشنهادات کاربردی 131](#_Toc191112176)

[مراجع 132](#_Toc191112177)

[پيوست‌ها 137](#_Toc191112178)

فهرست اشکال

فهرست جداول

# کلیات و مفاهیم اولیه

## مقدمه

در این پژوهش به پیش‌بینی موفقیت شرکت‌ها و استارتاپ‌ها پرداخته شده است که به عنوان یکی از موضوعات مهم در حوزه‌های کارآفرینی و سرمایه‌گذاری در دهه‌های اخیر به طور گسترده‌ای مورد توجه سرمایه گذاران و کارآفرینان قرار گرفته است. این اهمیت به ویژه در صنعت حمل و نقل که یکی از صنایع حیاتی اقتصاد جهانی محسوب می‌شود، بیشتر به چشم می‌آید. صنعت حمل‌ونقل یکی از زیربناهای اساسی توسعه اقتصادی و اجتماعی محسوب می‌شود که نقش مهمی در تسهیل تجارت، کاهش هزینه‌های مبادلاتی و افزایش بهره‌وری ایفا می‌کند. بهبود زیرساخت‌های حمل‌ونقل موجب افزایش دسترسی به بازارهای داخلی و بین‌المللی، کاهش هزینه‌های لجستیکی و رشد سرمایه‌گذاری می‌شود. علاوه بر این، حمل‌ونقل کارآمد به بهبود کیفیت زندگی، افزایش اشتغال و توزیع عادلانه‌تر فرصت‌های اقتصادی کمک کرده و نقش مهمی در توسعه پایدار شهری و کاهش آلودگی‌های زیست‌محیطی دارد(Banister & Berechman, 2001). همچنین با توجه به افزایش جمعیت و تقاضای روزافزون برای حمل و نقل سریع، ایمن و کارآمد، استفاده از فناوری‌های نوین از جمله هوش مصنوعی[[1]](#footnote-1) و یادگیری ماشین[[2]](#footnote-2) به منظور بهبود عملکرد و کارایی این صنعت به یکی از الزامات ضروری تبدیل شده است.

سرعت پیشرفت‌ هوش مصنوعی و یادگیری ماشین فرصت‌های بی‌سابقه‌ای را برای ارتقای عملکرد صنایع و مشاغل مختلف، از جمله بخش حمل‌ونقل فراهم می‌کند. در این زمینه، هوش مصنوعی برای غلبه بر چالش‌های مختلفی مانند آلودگی محیط زیست، افزایش تقاضای سفر، خودروهای خودران و نگرانی‌های ایمنی رانندگی به کار گرفته شده است و در سال‌های اخیر کسب و کارهای گوناگونی برای گسترش این حوزه شکل گرفته‌اند.

یکی از چالش‌های مهم پیش روی کارآفرینان و سرمایه‌گذاران استارتاپ‌ها، پیش‌بینی موفقیت این کسب‌وکارها است. این امر به ویژه در صنعت حمل و نقل با توجه به تغییرات سریع تکنولوژیکی و نیاز به سرمایه‌گذاری‌های کلان، از اهمیت بیشتری برخوردار است. پیش‌بینی موفقیت استارت‌آپ‌ها یک وظیفه حیاتی برای کارآفرینان و سرمایه‌گذاران استارت‌آپ است و این توانایی یک مزیت رقابتی ارزشمند برای سرمایه‌گذاران خطرپذیر که در جستجوی فرصت مناسب سرمایه‌گذاری هستند، ایجاد می‌کند.

در این راستا، مطالعات مختلفی در خصوص موفقیت استارتاپ‌ها صورت گرفته است. از جمله این مطالعات می‌توان به تحقیقاتی اشاره کرد که با استفاده از داده‌های جمع‌آوری شده از منابع مختلف، الگوهای موفقیت استارتاپ‌ها را شناسایی و مدل‌های پیش‌بینی کننده‌ای را ارائه داده‌اند. یکی از منابع مهم برای جمع آوری داده در این زمینه، وبسایت کرانچ بیس[[3]](#footnote-3) است که اطلاعات دقیقی در مورد استارتاپ‌ها، سرمایه‌ی جذب شده و روندهای صنعت ارائه می‌دهد.

این پژوهش با هدف ارائه مدلی برای پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی در حوزه حمل و نقل، با بهره‌گیری از داده‌های موجود در وبسایت کرانچ بیس انجام شده است. این مطالعه با بهره‌گیری از انبوه داده‌های موجود در این وبسایت، تجزیه و تحلیلی سیستماتیک برای بررسی ویژگی‌های مختلفی مانند دورهای تأمین مالی، ادغام و تملیک، عرضه اولیه سهام و روندهای بازار مرتبط با استارت‌آپ‌های هوش مصنوعی در صنعت حمل‌ونقل را ارائه خواهد داد.

برای این منظور، مدل‌های پیش‌بینی‌کننده از طریق یادگیری نظارت‌شده و با به کارگیری الگوریتم‌های داده‌کاوی مانند درخت تصمیم[[4]](#footnote-4)، شبکه‌های عصبی مصنوعی[[5]](#footnote-5)، جنگل تصادفی[[6]](#footnote-6)، الگوریتم ماشین بردار پشتیبان[[7]](#footnote-7)، الگوریتم رگرسیون لجستیک[[8]](#footnote-8)، الگوریتم تقویت گرادیان[[9]](#footnote-9) و الگوریتم بیز ساده[[10]](#footnote-10) ایجاد خواهند شد. این الگوریتم‌ها توانایی تجزیه و تحلیل الگوهای پیچیده در داده‌ها، کشف روابط پنهان میان متغیرها و پیش‌بینی دقیق در مورد آینده استارتاپ‌ها را فراهم می‌کنند. هدف این مطالعه، ایجاد بینشی دقیق در این حوزه است که به پیشرفت و پایداری سرمایه‌گذاری‌های نوآورانه کمک خواهند کرد.

در فصل اول این پژوهش، به بررسی کلیات و مفاهیم اولیه پژوهش پرداخته می‌شود. ابتدا به معرفی موضوع و بیان مسئله پرداخته خواهد شد تا خواننده با زمینه پژوهش و ضرورت انجام آن آشنا شود. سپس اهداف تحقیق و سوالات تحقیق مطرح می‌شوند که به دنبال یافتن پاسخ آن‌ها هستیم. در ادامه، ضرورت انجام تحقیق و اهمیت آن برای صنعت حمل و نقل و سرمایه‌گذاران بررسی می‌شود.

روش تحقیق و نحوه جمع‌آوری داده‌ها و تحلیل آن‌ها نیز در این فصل توضیح داده خواهد شد. همچنین به سهم دانش افزایی و نوآوری تحقیق پرداخته می‌شود تا نشان دهد چگونه این پژوهش می‌تواند به پیشرفت علمی و عملی در این حوزه کمک کند. در پایان، واژگان تخصصی مورد استفاده در تحقیق معرفی و تعریف خواهند شد تا خواننده با اصطلاحات و مفاهیم اصلی آشنا شود.

## معرفی موضوع

هوش مصنوعی توانایی یک ماشین برای انجام عملکردهای شناختی مانند ادراک، استدلال، یادگیری و حل مسئله است که از دو دهه گذشته، به دلیل در دسترس بودن حجم عظیمی از داده‌های تولید شده از طریق اینترنت، در سراسر جهان مورد توجه قرار گرفته است. پردازش این داده‌ها با استفاده از الگوریتم‌های پیشرفته در گذشته نه چندان دور، سود زیادی برای دولت‌ها و کسب و کارها داشته است. نمونه‌هایی از روش‌های هوش مصنوعی که امروزه به حوزه حمل‌ونقل راه پیدا کرده‌اند عبارتند از: شبکه‌های عصبی مصنوعی، ماشین بردار پشتیبان، تبرید شبیه‌سازی شده، سیستم ایمنی مصنوعی، بهینه‌ساز کلونی مورچه‌ها، و مدل منطق فازی. این الگوریتم‌ها برای ارائه خدمات مختلفی در صنعت حمل‌ونقل به کار می‌روند که توانمندسازی کاربر را ارتقا داده و از تعاملات انسان و ماشین پشتیبانی می‌کند (Nguyen et al., 2018; Nikitas et al., 2020).

در طول صد سال گذشته، صنعت حمل و نقل شاهد دگرگونی‌ها و انقلاب‌های متعددی بوده است و اکنون در نقطه‌ای هستیم که پیشرفت‌های قابل توجهی در قالب هوش مصنوعی در حال انجام است. هوش مصنوعی توجه مدیران حمل و نقل در سراسر جهان را به خود جلب کرده است، خواه از طریق اتومبیل‌های خودران برای اطمینان بیشتر، نظارت بر وضعیت جاده برای افزایش ایمنی (Olugbade et al., 2022)، تجزیه و تحلیل جریان ترافیک برای کارایی بیشتر (Akhtar & Moridpour, 2021) و یا هوشمندسازی اینترنت اشیا برای جلوگیری از تصادفات جاده‌ای (Khan et al., 2022; Zantalis et al., 2019). هوش مصنوعی اکنون بیش از هر زمان دیگری در دسترس است. فعالان صنعت حمل و نقل و هوش مصنوعی را کاملا با هم ترکیب کرده‌اند، از جمله خودرو سازان، پارکینگ داران، خطوط کروز و موارد دیگر (Abduljabbar et al., 2019; Whig et al., 2024). همچنین مسئله شهرهای هوشمند آینده و تحول تکنولوژیک سیستم‌های جاده‌ای شهری نیز با رویکرد پایداری سه گانه مرتبط است: هزینه، محیط زیست و تعاملات اجتماعی (Hasan et al., 2019; Okrepilov et al., 2022).

طبق آمارهای منتشر شده در رابطه با جذب سرمایه حوزه‌های مختلف، مشاهده شده است که جذب سرمایه در برخی از صنایع حمل و نقل عمومی روند رو به کاهشی را تجربه می‌کنند (Glasner, 2023). به علاوه، با توجه به نوآوری‌های جدید هوش مصنوعی در حوزه‌های مختلف و عدم شناسایی ریسک‌های بالقوه بازار، مانند رشد بسیار سریع فناوری‌های هوش مصنوعی، احتمال شکست استارتاپ‌ها در این حوزه‌ها افزایش می‌یابد و در چندین موارد این پیچیدگی‌ها منجر به شکست استارتاپ شده است. بنابراین سرمایه‌گذاران برای سرمایه‌گذاری در استارتاپ‌های ترکیب این دو حوزه با هم، استارتاپ‌های هوش مصنوعی در حوزه حمل و نقل، با مشکلات متعددی روبرو می‌شوند. برای مثال، استارتاپ فناوری خودروهای خودمختار آرگو[[11]](#footnote-11) در آستانه تعطیلی قرار دارد. در سال 2017، این شرکت یک میلیارد دلار سرمایه گذاری از کمپانی فورد دریافت کرد و توسط فولکس واگن نیز حمایت می‌شد، اما در نهایت به علت پیش بینی ناپذیری زمان دستیابی به فناوری سطح 4 خودروهای خودران موفق به جذب سرمایه نشد و به علت عدم جذب سرمایه گذار مجبور به توقف فعالیت خود شد(RAMEY, 2022).

در این راستا، باید پژوهش‌های کمی و کیفی مختلفی برای شناسایی عوامل موفقیت و همچنین ارائه مدل‌های پیش‌بینی موفقیت کسب و کارهای حوزه حمل و نقل در صنایع مختلف انجام شود. برای این منظور دو روش مختلف به کار برده می‌شود: تحقیقی و تجربی. مطالعات تحقیقی بر شناسایی عواملی که از نظر آماری با موفقیت سرمایه گذاری مرتبط هستند تمرکز می‌کنند. مطالعات تجربی از الگوریتم‌های یادگیری ماشین برای شناسایی عوامل مرتبط با موفقیت شرکت و متعاقباً پیش بینی موفقیت شرکت‌ها استفاده می‌کنند. نتایج این مطالعات نشان دهنده پتانسیل پیش‌بینی خودکار موفقیت شرکت‌ها برای مقاصد سرمایه گذاری است. (Bento, 2018)

هدف این تحقیق، بررسی عوامل موفقیت و شکست کسب و کارهای هوش مصنوعی در حوزه حمل و نقل است. برای این منظور از داده‌های سایت کرانچ بیس و مدل‌های ماشین لرنینگ برای پیش‌بینی موفقیت کسب و کارهای فعال در این حوزه استفاده می‌شود. در این پژوهش، کسب و کارهای هوش مصنوعی شامل شرکت‌هایی است که محصولات یا خدمات را با استفاده از فناوری‌هایی مانند خودروهای خودران، خودروهای اشتراکی[[12]](#footnote-12) ، خدمات ارسال[[13]](#footnote-13) ، خودروهای برقی، و غیره ارائه می‌کنند.

## بیان مساله

صنعت حمل‌ونقل با نقش حیاتی در اقتصاد جهانی، جابه‌جایی افراد و کالاها و بهبود کیفیت زندگی، تحت تاثیر چالش‌های پیچیده‌ای مانند افزایش تقاضا، آلودگی محیط زیست و مشکلات ایمنی قرار دارد. این چالش‌ها نیازمند راه‌حل‌های نوآورانه و کارآمد است. یکی از رویکردهای نوین برای مواجهه با این چالش‌ها، استفاده از هوش مصنوعی و یادگیری ماشین است. این فناوری‌ها به دلیل توانایی تحلیل حجم بالایی از داده‌ها و استخراج الگوهای پنهان، فرصت‌های جدیدی را برای بهبود عملکرد و کارایی صنعت حمل‌ونقل ایجاد کرده‌اند(Abduljabbar et al., 2019).

با این حال، یکی از مشکلات اساسی که صنعت حمل‌ونقل با آن روبرو است، ناتوانی در پیش‌بینی دقیق موفقیت استارتاپ‌ها در این حوزه است. استارتاپ‌ها به عنوان محرکان اصلی نوآوری، نقش مهمی در توسعه فناوری‌های جدید ایفا می‌کنند. با این وجود، احتمال شکست این کسب‌وکارها نیز بالاست و نیاز به ابزارهای دقیق برای پیش‌بینی موفقیت آنها احساس می‌شود(Kim et al., 2023). پیش‌بینی موفقیت استارتاپ‌ها می‌تواند به کارآفرینان در بهبود مدل‌های کسب‌وکار و جذب سرمایه، و به سرمایه‌گذاران در شناسایی فرصت‌های مناسب سرمایه‌گذاری و کاهش ریسک کمک کند (Bento, 2017).

در این راستا تحقیقات نشان داده‌اند که استفاده از هوش مصنوعی و یادگیری ماشین می‌تواند به بهبود دقت پیش‌بینی موفقیت استارتاپ‌ها کمک کند. این فناوری‌ها قادرند با تحلیل داده‌های تاریخی و بررسی روابط پیچیده میان متغیرها، مدل‌هایی ایجاد کنند که به صورت دقیق‌تری موفقیت یا شکست استارتاپ‌ها را پیش‌بینی کنند (Nguyen et al., 2018). به عنوان مثال، الگوریتم‌های یادگیری ماشین مانند درخت تصمیم، شبکه‌های عصبی مصنوعی و جنگل تصادفی می‌توانند به شناسایی الگوهای موفقیت در داده‌های استارتاپ‌ها بپردازند و بر اساس آن پیش‌بینی‌های دقیقی ارائه دهند (Behrooz & Hayeri, 2022).

یکی از چالش‌های اصلی در پیش‌بینی موفقیت استارتاپ‌ها، دسترسی به داده‌های جامع و دقیق است. وبسایت کرانچ بیس به عنوان یکی از بزرگترین پایگاه‌های داده اطلاعات استارتاپ‌ها، می‌تواند منبع ارزشمندی برای این منظور باشد. این وبسایت اطلاعات جامعی در مورد استارتاپ‌ها، سرمایه‌های جذب شده و روندهای بازار ارائه می‌دهد که می‌تواند به عنوان پایه‌ای برای تحلیل و پیش‌بینی موفقیت استارتاپ‌ها استفاده شود (Kim et al., 2023).

در پژوهش‌های پیشین، روش‌های مختلفی برای پیش‌بینی موفقیت استارتاپ‌ها پیشنهاد شده است. به عنوان مثال، استفاده از داده‌های تامین مالی، تعداد و نوع سرمایه‌گذاران، ادغام و تملیک، عرضه اولیه سهام و روندهای بازار به عنوان شاخص‌های موفقیت مطرح شده‌اند (Kim et al., 2023). همچنین، استفاده از الگوریتم‌های مختلف یادگیری ماشین مانند درخت تصمیم، شبکه‌های عصبی مصنوعی و جنگل تصادفی به منظور تحلیل داده‌ها و پیش‌بینی موفقیت استارتاپ‌ها مورد بررسی قرار گرفته است (Abduljabbar et al., 2019; Bento, 2017).

با وجود تلاش‌های بسیاری که در این زمینه انجام شده، هنوز هم نیاز به مدل‌های دقیق‌تر و کاربردی‌تر احساس می‌شود. مدل‌های فعلی اغلب به دلیل نداشتن داده‌های کافی یا استفاده از روش‌های تحلیل نادرست، دقت کافی را ندارند. همچنین در این مدل‌ها سهم نوع صنعت در موفقیت استارتاپ‌ها نادیده گرفته شده است. از این رو، نیاز است که با استفاده از روش‌های پیشرفته‌تر و داده‌های جامع‌تر، مدل‌هایی توسعه یابند که بتوانند با دقت بالاتری موفقیت استارتاپ‌ها را پیش‌بینی کنند (Akhtar & Moridpour, 2021; Nguyen et al., 2018).

هدف این پژوهش، توسعه مدلی برای پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل با استفاده از داده‌های جمع‌آوری شده از وبسایت کرانچ بیس است. این مدل با استفاده از الگوریتم‌های یادگیری ماشین و تجزیه و تحلیل داده‌های مختلفی مانند دورهای تأمین مالی، ادغام و تملیک، عرضه اولیه سهام و روندهای بازار، به شناسایی عواملی که بر موفقیت استارتاپ‌ها تاثیر می‌گذارند، می‌پردازد. این پژوهش می‌تواند به کارآفرینان و سرمایه‌گذاران در اتخاذ تصمیمات بهتر و هوشمندانه‌تر کمک کند و به توسعه فناوری‌های نوین و بهبود عملکرد کسب‌وکارها یاری رساند (Hasan et al., 2019).

در نهایت، این پژوهش تلاش دارد تا با ارائه یک مدل پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل، به بهبود و پایداری سرمایه‌گذاری‌های نوآورانه کمک کند و به کارآفرینان و سرمایه‌گذاران در اتخاذ تصمیمات بهتر و هوشمندانه‌تر یاری رساند. این مدل می‌تواند به عنوان یک ابزار قدرتمند برای تحلیل و پیش‌بینی موفقیت استارتاپ‌ها در سایر صنایع نیز مورد استفاده قرار گیرد و به توسعه فناوری‌های نوین و بهبود عملکرد کسب‌وکارها کمک کند (Khan et al., 2022; Okrepilov et al., 2022).

## اهداف تحقیق

* تعیین متغیرهای موثر بر موفقیت کسب و کارهای هوش مصنوعی در حوزه حمل و نقل
* ارائه مدل ارزیابی و یادگیری ماشین با توجه به داده‌های در دسترس
* پیاده‌سازی مدل و اعتبار سنجی آن

## سوالات تحقیق

* چه عواملی بیشترین تاثیر را بر موفقیت شرکت‌های هوش مصنوعی در صنعت حمل و نقل دارند؟
* چگونه می‌توان با استفاده از داده‌های موجود، موفقیت شرکت‌های هوش مصنوعی را پیش‌بینی کرد؟
* کدام الگوریتم‌های داده‌کاوی بهترین عملکرد را در پیش‌بینی موفقیت این شرکت‌ها دارند؟

## ضرورت انجام تحقیق

توسعه هوش مصنوعی و یادگیری ماشین در سال‌های اخیر صنعت حمل و نقل را به کلی متحول کرده است، به طوری‌که مطالعات قبلی در این حوزه نیاز به بازبینی مجدد خواهند داشت. تحقیق در زمینه پیش‌بینی موفقیت کسب و کارهای هوش مصنوعی در حوزه حمل و نقل ابعاد جدیدی از فرایند تصمیم‌گیری در این صنعت را آشکار می‌سازد و می‌تواند به توسعه نظریه‌های مرتبط با عوامل کلیدی موفقیت در این حوزه کمک کند.

استارتاپ‌های هوش مصنوعی نقش مهمی در معرفی نوآوری‌ها و فناوری‌های جدید در صنعت حمل و نقل ایفا می‌کنند. این شرکت‌های نوپا با ایده‌ها و راه‌حل‌های خلاقانه خود، می‌توانند مسائل پیچیده‌ای مانند خودروهای خودران، مدیریت ترافیک، و بهینه‌سازی مسیرها را حل کنند. اما موفقیت این استارتاپ‌ها نیازمند تحلیل دقیق و پیش‌بینی‌های صحیح است تا بتوان منابع مالی و انسانی را به درستی تخصیص داد.

مخاطبان و ذی نفعان اصلی این تحقیق کارآفرینان و کسب و کارهای حوزه حمل و نقل هستند تا بتوانند اولا با کاربردها و شکل جدید صنعت حمل و نقل با وجود گسترش هوش مصنوعی آشنا شوند و ثانیا از مدل‌هاي پیش‌بینی در مورد موفقیت یا شکست ایده‌هاي تجاري خود استفاده کنند، زیرا می‌توانند در مواقع بحرانی با تصمیم‌گیری‌های خود تغییراتی را در مدل‌های تجاری خود اتخاذ کنند. سایر ذی‌نفعان مهم کسب و کارهای نوپا شامل سرمایه‌گذاران فرشته، صندوق‌هاي سرمایه‌ي اولیه، سرمایه‌گذاران خطرپذیر و غیره هستند. این نتایج کاربردی می‌توانند به شرکت‌ها و سازمان‌های حمل و نقل شهری در بهبود عملکرد خود، مدیریت بهینه منابع و افزایش کارایی کمک کند. پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی می‌تواند به کاهش ریسک‌های سرمایه‌گذاری کمک کند. با تحلیل دقیق داده‌ها و شناسایی عوامل کلیدی موفقیت، سرمایه‌گذاران می‌توانند ریسک‌های مرتبط با سرمایه‌گذاری در شرکت‌های نوپا را به حداقل برسانند. این امر نه تنها به افزایش بازده سرمایه‌گذاری منجر می‌شود بلکه به توسعه پایدار و نوآوری در صنعت حمل و نقل نیز کمک می‌کند.

از سمت دیگر مطالعه کافی در این زمینه از نظر قانونی و برنامه‌ای دارای اهمیت زیادی است، زیرا به تدوین دستاوردهایی می‌پردازد که با سیاست‌های کلان کشور و برنامه‌های توسعه هم‌خوانی دارند. این تحقیق می‌تواند به تعیین راهبردها و سیاست‌های متناسب با توسعه صنعت حمل و نقل شهری اساسی باشد و از نظر قانونی تطابق مؤثری با مقررات بالادستی ارائه دهد. با توجه به تاکید مقامات کشوری بر توسعه حوزه هوش مصنوعی مهم است که همه‌ی صنایع مرتبط و نوآوری‌های این حوزه بررسی گردند.)

با توجه به موارد فوق، انجام این تحقیق ضروری و دارای اهمیت بالایی است. نتایج آن می‌تواند تأثیرات مثبتی بر صنعت حمل و نقل و سرمایه‌گذاری‌های مرتبط با هوش مصنوعی داشته باشد و به بهبود و توسعه پایدار این حوزه کمک کند.

## روش تحقیق

در این پژوهش از چارچوب روش‌شناسی CRISP-DM برای فرآیند تحقیق استفاده می‌شود. این چارچوب که در سال 1996 توسعه یافت، به‌طور گسترده‌ای در حوزه داده‌کاوی و یادگیری ماشین برای ایجاد ساختار منظم و قابل تکرار به کار می‌رود (Wirth & Hipp, 2000). این روش شامل شش مرحله است که در ادامه هر مرحله بیان شده است.

1. درک کسب‌وکار[[14]](#footnote-14)

در مرحله نخست، هدف تحقیق شناسایی و تعریف می‌شود. این تحقیق با هدف توسعه مدلی برای پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل و نقل انجام می‌شود. موفقیت شرکت‌ها در این زمینه به عوامل مختلفی از جمله میزان سرمایه جذب‌شده، تعداد دورهای تأمین مالی و نوع سرمایه‌گذاران وابسته است.

تعریف دقیق مسأله کسب‌وکار و اهداف اصلی در این مرحله از اهمیت زیادی برخوردار است. هدف این پژوهش ارائه مدلی است که بتواند موفقیت شرکت‌ها را بر اساس داده‌های جمع‌آوری شده پیش‌بینی کند. علاوه بر این، شاخص‌های کلیدی عملکرد[[15]](#footnote-15)مانند رشد درآمد، نرخ بقا، و موفقیت در خروج (عرضه اولیه سهام یا تملک) به عنوان معیارهای ارزیابی موفقیت استفاده می‌شوند (Hughes & Morgan, 2007).

1. درک داده‌ها[[16]](#footnote-16)

داده‌ها از پایگاه داده کرانچ‌بیس استخراج خواهند شد. این پایگاه داده اطلاعات جامعی از جمله تأمین مالی استارتاپ‌ها، نوع سرمایه‌گذاران، تاریخچه شرکت‌ها، و وضعیت فعلی آنها را در اختیار قرار می‌دهد. تحلیل اولیه داده‌ها برای شناسایی ویژگی‌های کلیدی، بررسی توزیع داده‌ها و شناسایی الگوهای اولیه از این داده‌ها ضروری است.

برخی از این داده‌ها ممکن است به صورت مستقیم با موفقیت شرکت مرتبط نباشند، اما از طریق تحلیل مقدماتی داده‌ها و ایجاد فرضیات اولیه می‌توان الگوهای پنهان را شناسایی کرد. این مرحله شامل موارد زیر است:

* بررسی جامع داده‌ها برای درک کامل از انواع متغیرها و شناسایی ناهنجاری‌های احتمالی.
* تحلیل توزیع متغیرها برای شناسایی الگوهای مقدماتی و شناخت همبستگی‌های بالقوه میان متغیرها.

در این مرحله، شناخت دقیق از ساختار و ویژگی‌های داده‌ها به مدل‌سازی موفق در مراحل بعدی کمک خواهد کرد (Witten et al., 2011).

3- آماده‌سازی داده‌ها[[17]](#footnote-17)

داده‌های جمع‌آوری شده نیاز به پیش‌پردازش و آماده‌سازی دارند. این مرحله از اهمیت زیادی برخوردار است، چرا که کیفیت داده‌های ورودی بر روی کیفیت مدل تأثیر مستقیم دارد.

مراحل آماده‌سازی داده‌ها شامل موارد زیر است:

* پاکسازی داده‌ها: حذف مقادیر گمشده یا ناقص و تصحیح داده‌های نادرست. همچنین داده‌های پرت[[18]](#footnote-18) شناسایی و مدیریت خواهند شد تا نتایج نهایی بهینه‌تر باشند (García et al., 2015).
* نرمال‌سازی داده‌ها: تبدیل داده‌ها به مقیاس‌های قابل مقایسه برای استفاده در الگوریتم‌های ماشین لرنینگ.
* ایجاد ویژگی‌های جدید: از جمله ویژگی‌های زمان‌بندی سرمایه‌گذاری‌ها، نوع سرمایه‌گذاران، یا میزان رشد شرکت که می‌توانند تأثیر بسزایی در موفقیت داشته باشند.

این مرحله همچنین شامل مهندسی ویژگی‌ها است که از طریق آن ویژگی‌های پیچیده‌تری استخراج می‌شود که به افزایش دقت مدل‌ها کمک می‌کند (Guyon & Elisseeff, 2003).

4- مدل‌سازی[[19]](#footnote-19)

پس از آماده‌سازی داده‌ها، مدل‌های مختلف داده‌کاوی برای پیش‌بینی موفقیت شرکت‌ها استفاده خواهند شد. برای این تحقیق، الگوریتم‌های مختلفی از جمله درخت تصمیم[[20]](#footnote-20)، شبکه‌های عصبی مصنوعی[[21]](#footnote-21)، جنگل تصادفی[[22]](#footnote-22)، الگوریتم ماشین بردار پشتیبان[[23]](#footnote-23)، الگوریتم رگرسیون لجستیک[[24]](#footnote-24)، الگوریتم تقویت گرادیان[[25]](#footnote-25) و الگوریتم بیز ساده به کار خواهند رفت.

هر کدام از این الگوریتم‌ها نقاط قوت و ضعف خود را دارند که در فصل روش تحقیق به تفصیل شرح داده می‌شوند.

5- ارزیابی[[26]](#footnote-26)

برای ارزیابی عملکرد مدل‌ها از اعتبارسنجی متقابل[[27]](#footnote-27) و معیارهای مختلفی مانند دقت[[28]](#footnote-28)، صحت[[29]](#footnote-29)، بازخوانی[[30]](#footnote-30) و امتیاز F1 استفاده خواهد شد.

روش اعتبارسنجی متقابل به ما اجازه می‌دهد که داده‌ها را به چندین زیرمجموعه تقسیم کرده و مدل‌ها را بر روی زیرمجموعه‌های مختلف ارزیابی کنیم تا از صحت نتایج اطمینان حاصل شود.

امتیاز F1 به‌خصوص در مسائلی که کلاس‌ها نامتعادل هستند (یعنی تعداد نمونه‌های یک کلاس بیشتر از کلاس دیگر است) بسیار اهمیت دارد و معیار مهمی برای ایجاد تعادل بین دقت و بازخوانی است.

6- استقرار[[31]](#footnote-31)

در این مرحله، مدل نهایی انتخاب شده و برای استفاده عملی آماده می‌شود. این مدل می‌تواند به عنوان یک ابزار تصمیم‌گیری برای سرمایه‌گذاران و کارآفرینان در صنعت حمل و نقل هوش مصنوعی به کار گرفته شود. همچنین، نتایج و پیشنهادات به‌صورت گزارش‌های کاربردی و نمودارهای بصری ارائه خواهند شد. این مرحله اهمیت زیادی در تبدیل نتایج تحقیق به ابزارهای کاربردی دارد.

7- نتیجه‌گیری

در نهایت، نتایج پژوهش به صورت خلاصه ارائه می‌شود. مدل پیشنهادی این تحقیق می‌تواند به سرمایه‌گذاران و کارآفرینان کمک کند تا تصمیمات بهتری در مورد سرمایه‌گذاری بر روی شرکت‌های هوش مصنوعی در صنعت حمل و نقل اتخاذ کنند. همچنین، ویژگی‌های کلیدی که بر موفقیت شرکت‌ها تأثیر دارند شناسایی شده و به عنوان توصیه‌هایی برای بهبود فرآیندهای سرمایه‌گذاری و رشد استارتاپ‌ها پیشنهاد می‌شوند.

## سهم دانش افزایی و نوآوری تحقیق

در این پژوهش متغیرهای مختلف که تا بحال در کنار یکدیگر و در یک صنعت خاص در کنار یکدیگر مدل‌سازی نشده‌اند، مورد بررسی قرار خواهند گرفت. بنابراین تاثیر هر یک از متغیرها در موفقیت کسب و کارهای حوزه هوش مصنوعی در صنعت حمل و نقل مشخص خواهد شد.

این تحقیق با بررسی جامع و سیستماتیک استارتاپ‌های هوش مصنوعی در صنعت حمل و نقل و ایجاد مدل‌های پیش‌بینی‌کننده، به جنبه‌های مختلفی از دانش و نوآوری می‌افزاید. سهم دانش افزایی این پژوهش به شرح زیر است:

* + 1. **توسعه مدل‌های پیش‌بینی موفقیت**

این تحقیق با به کارگیری الگوریتم‌های پیشرفته داده‌کاوی نظیر درخت تصمیم، شبکه‌های عصبی مصنوعی و جنگل تصادفی، مدل‌هایی را برای پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل و نقل توسعه می‌دهد. این مدل‌ها به شناسایی الگوها و روابط پنهان در داده‌ها کمک می‌کنند و با دقت بالا قادر به پیش‌بینی موفقیت استارتاپ‌ها هستند. این جنبه از تحقیق می‌تواند به کارآفرینان و سرمایه‌گذاران در تصمیم‌گیری‌های بهتر و شناسایی فرصت‌های سرمایه‌گذاری کمک کند.

* + 1. **تحلیل عوامل کلیدی موفقیت**

این پژوهش به شناسایی و تحلیل عوامل کلیدی موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل و نقل می‌پردازد. با بررسی داده‌های جمع‌آوری‌شده از منابع معتبر مانند کرانچ‌بیس، این تحقیق به درک عمیق‌تری از ویژگی‌های مهم و تأثیرگذار بر موفقیت استارتاپ‌ها دست می‌یابد. این تحلیل‌ها می‌توانند به راهنمایی استارتاپ‌ها در جهت بهبود عملکرد و افزایش شانس موفقیت کمک کنند.

* + 1. **ارائه روش‌شناسی جامع**

روش‌شناسی به کار رفته در این تحقیق، که شامل جمع‌آوری، پیش‌پردازش و تحلیل داده‌ها و همچنین انتخاب و ارزیابی مدل‌های داده‌کاوی است، می‌تواند به عنوان یک چارچوب مفید برای سایر پژوهشگران در حوزه‌های مشابه مورد استفاده قرار گیرد. این روش‌شناسی با تأکید بر دقت و جامعیت، می‌تواند به بهبود کیفیت و دقت تحقیقات آینده در زمینه پیش‌بینی موفقیت استارتاپ‌ها کمک کند.

* + 1. **کاربردهای عملی و صنعتی**

نتایج این تحقیق، علاوه بر ارائه دانش تئوریک، دارای کاربردهای عملی در صنعت حمل و نقل است. با ارائه مدل‌های پیش‌بینی‌کننده موفقیت، این تحقیق می‌تواند به مدیران و تصمیم‌گیران صنعتی کمک کند تا استراتژی‌های بهتری برای توسعه و نوآوری در این صنعت تدوین کنند. این مدل‌ها همچنین می‌توانند به بهبود فرآیندهای مدیریتی و کاهش ریسک‌های مرتبط با سرمایه‌گذاری در استارتاپ‌های هوش مصنوعی کمک کنند.

* + 1. **نوآوری در تحلیل داده‌ها**

این تحقیق با استفاده از الگوریتم‌های پیشرفته داده‌کاوی و تحلیل‌های پیچیده، نوآوری‌هایی در روش‌های تحلیل داده‌ها ارائه می‌دهد. این نوآوری‌ها می‌توانند به بهبود دقت و کارایی مدل‌های پیش‌بینی‌کننده و همچنین کشف روابط پیچیده‌تر و غیرخطی در داده‌ها منجر شوند. این جنبه از تحقیق به توسعه روش‌های جدید و پیشرفته‌تر در زمینه داده‌کاوی و تحلیل داده‌ها کمک می‌کند.

در مجموع، این تحقیق با ارائه مدل‌های پیش‌بینی‌کننده دقیق و تحلیل‌های جامع، به افزایش دانش و نوآوری در حوزه‌های هوش مصنوعی، داده‌کاوی و صنعت حمل و نقل کمک می‌کند و نقش مهمی در پیشرفت و پایداری سرمایه‌گذاری‌های نوآورانه ایفا می‌کند.

## واژگان تحقیق

* **هوش مصنوعی:**

هوش مصنوعی به شاخه‌ای از علم کامپیوتر گفته می‌شود که هدف آن ایجاد سیستم‌هایی است که بتوانند وظایفی را که به‌طور طبیعی به هوش انسانی نیاز دارند، انجام دهند. این وظایف شامل یادگیری، استدلال، حل مسئله، درک زبان طبیعی و بینایی کامپیوتری است

* **مدل پیش‌بینی موفقیت[[32]](#footnote-32):**

مدل‌های پیش‌بینی موفقیت به الگوریتم‌ها و روش‌هایی اطلاق می‌شود که برای پیش‌بینی احتمال موفقیت یک کسب و کار یا استارتاپ بر اساس مجموعه‌ای از ویژگی‌ها و داده‌ها طراحی شده‌اند. این مدل‌ها می‌توانند شامل الگوریتم‌های یادگیری ماشینی مانند رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، و شبکه‌های عصبی باشند

* **صنعت حمل و نقل:**

صنعت حمل و نقل گروهی از شرکت ها هستند که خدمات را برای جا به جایی افراد یا کالا ها و هم چنین زیرساخت های حمل و نقل ارائه می دهند. صنعت حمل و نقل از چندین بخش تشکیل شده است که شامل موارد زیر هستند: هوایی، خطوط هوایی، دریایی، جاده‌ای، ریلی و غیره. استفاده از هوش مصنوعی در این صنعت می‌تواند به بهبود مدیریت ترافیک، ایمنی، و کاهش انتشار گازهای گلخانه‌ای کمک کند

* **داده کاوی:**

داده‌کاوی فرآیند استخراج اطلاعات مفید و الگوهای پنهان از مجموعه‌های بزرگ داده است. این فرآیند با استفاده از تکنیک‌های آماری، یادگیری ماشینی و پایگاه‌داده‌ها انجام می‌شود. داده‌کاوی می‌تواند به درک بهتر الگوهای موجود در داده‌ها و پیش‌بینی روندهای آینده کمک کند.

* **وبسایت کرانچ بیس:**

یک پایگاه داده آنلاین است که اطلاعات جامع و دقیقی درباره استارتاپ‌ها، دورهای تامین مالی، ادغام و تملیک، عرضه اولیه سهام، و روندهای بازار را ارائه می‌دهد. این وب‌سایت یکی از منابع اصلی جمع‌آوری داده برای تحقیقات مرتبط با موفقیت استارتاپ‌ها است.

* **سرمایه‌گذاری خطرپذیر [[33]](#footnote-33)**

نوعی از سرمایه‌گذاری است که در استارتاپ‌ها و شرکت‌های کوچک با پتانسیل رشد بالا انجام می‌شود. سرمایه‌گذاران خطرپذیر منابع مالی و مشاوره مدیریتی به این شرکت‌ها ارائه می‌دهند. هدف این سرمایه‌گذاری‌ها، دستیابی به بازده بالا از طریق رشد سریع و موفقیت بلندمدت است.

* **نرخ مثبت واقعی[[34]](#footnote-34):**

**م**عیاری است که نشان می‌دهد چه تعداد از موارد واقعی مثبت به درستی توسط مدل پیش‌بینی شده‌اند. که به عنوان حساسیت یا یادآوری نیز شناخته می‌شود، نسبت موارد مثبت درست پیش‌بینی شده به کل موارد واقعی مثبت است. این معیار برای ارزیابی کارایی مدل‌های پیش‌بینی در تشخیص صحیح موارد مثبت استفاده می‌شود.

* **نرخ مثبت کاذب[[35]](#footnote-35):**

**م**عیاری است که نشان می‌دهد چه تعداد از موارد واقعی منفی به اشتباه به عنوان مثبت پیش‌بینی شده‌اند. نسبت موارد منفی که به اشتباه به عنوان مثبت پیش‌بینی شده‌اند به کل موارد واقعی منفی است. این معیار برای ارزیابی دقت مدل‌های پیش‌بینی در تشخیص صحیح موارد منفی استفاده می‌شود.

* **ادغام و تملیک[[36]](#footnote-36):**

فرآیندهایی است که در آن شرکت‌ها با هم ترکیب می‌شوند یا یکی از آنها دیگری را خریداری می‌کند. ادغام و تملک به شرکت‌ها امکان می‌دهد منابع و بازارهای جدید را به دست آورند، هزینه‌ها را کاهش دهند و توانایی‌های خود را افزایش دهند. این فرآیندها به استراتژی‌های رشد و توسعه شرکت‌ها کمک می‌کنند.

* **عرضه اولیه عمومی[[37]](#footnote-37):**

فرآیندی است که در آن یک شرکت برای اولین بار سهام خود را به عموم عرضه می‌کند. عرضه اولیه عمومی به شرکت‌ها امکان می‌دهد سرمایه جدید جذب کنند و نقدینگی برای سهامداران فعلی فراهم کنند. این فرآیند شامل انتشار اسناد مالی و افشای اطلاعات مرتبط با شرکت است.

## جمع بندی و ساختار تحقیق

در این فصل به بررسی کلیات و مفاهیم اولیه تحقیق در زمینه پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل و نقل پرداختیم. چالش‌های موجود در این صنعت از جمله افزایش تقاضا، آلودگی محیط زیست، و مسائل ایمنی نیازمند راه‌حل‌های نوآورانه و کارآمد است. هوش مصنوعی و یادگیری ماشین با توانایی تحلیل داده‌های بزرگ و استخراج الگوهای پنهان، امکانات جدیدی را برای بهبود عملکرد و کارایی این صنعت فراهم کرده‌اند. هدف این پژوهش توسعه مدلی است که بتواند موفقیت استارتاپ‌های هوش مصنوعی در حوزه حمل و نقل را، با استفاده از داده‌های موجود در وبسایت کرانچ بیس و الگوریتم‌های پیشرفته یادگیری ماشین پیش‌بینی کند.

این تحقیق در چندین فصل تنظیم شده است که هر یک به جنبه‌های مختلف موضوع می‌پردازند. فصل اول به بررسی کلیات و مفاهیم اولیه اختصاص دارد، جایی که موضوع تحقیق، مسئله، اهداف، سوالات تحقیق، ضرورت انجام تحقیق، روش تحقیق، سهم دانش افزایی و نوآوری تحقیق، و واژگان تخصصی مورد بررسی قرار گرفتند.

در فصل دوم، مرور ادبیات صورت خواهد گرفت. این فصل به بررسی مطالعات و تحقیقات پیشین در زمینه پیش‌بینی موفقیت استارتاپ‌ها و کاربردهای هوش مصنوعی در صنعت حمل و نقل می‌پردازد. این مرور جامع به درک بهتر پیشینه علمی موضوع و شناسایی شکاف‌های تحقیقاتی کمک می‌کند.

فصل سوم به روش شناسی تحقیق اختصاص دارد. در این فصل، روش‌های تحقیق، نحوه جمع‌آوری داده‌ها، ابزارها و تکنیک‌های مورد استفاده به تفصیل تشریح می‌شوند. این بخش از تحقیق، پایه و اساس علمی پژوهش را تشکیل می‌دهد و تضمین می‌کند که نتایج تحقیق معتبر و قابل اعتماد باشند.

در فصل چهارم، تحلیل داده‌ها و ارائه مدل صورت می‌گیرد. داده‌های جمع‌آوری شده مورد تحلیل قرار می‌گیرند و مدل پیش‌بینی‌کننده با استفاده از الگوریتم‌های یادگیری ماشین توسعه و ارزیابی می‌شود. نتایج حاصل از این تحلیل‌ها، اطلاعات مفیدی را برای فهم بهتر عوامل مؤثر بر موفقیت استارتاپ‌ها فراهم می‌آورند.

فصل پنجم به بحث و نتیجه‌گیری اختصاص دارد. در این فصل، نتایج تحقیق تفسیر و با مطالعات پیشین مقایسه می‌شوند. همچنین، محدودیت‌های تحقیق مورد بحث قرار می‌گیرند و پیشنهادات برای تحقیقات آینده ارائه می‌شوند. این بخش از تحقیق به جمع‌بندی کلی دستاوردهای پژوهش می‌پردازد و به ارائه راهکارهای عملی برای کارآفرینان و سرمایه‌گذاران کمک می‌کند.

در نهایت، فصل ششم به منابع و مراجع اختصاص دارد. در این فصل، تمامی منابع و مراجعی که در طول تحقیق مورد استفاده قرار گرفته‌اند، به ترتیب حروف الفبا ذکر می‌شوند. این فهرست منابع به خوانندگان کمک می‌کند تا به اطلاعات بیشتری در مورد موضوع تحقیق دسترسی پیدا کنند و از دقت و جامعیت پژوهش اطمینان حاصل کنند.

# 

# مروري بر منابع

* 2. مقدمه

سرعت پیشرفت‌های فناوری در حوزه‌های مختلف، به ویژه در زمینه هوش مصنوعی و یادگیری ماشین، فرصت‌های بی‌سابقه‌ای را برای ارتقای عملکرد صنایع و مشاغل فراهم کرده است. یکی از صنایعی که از این پیشرفت‌ها بهره‌برداری کرده، صنعت حمل‌ونقل است. استفاده از هوش مصنوعی در این صنعت، بهبود کارایی، کاهش هزینه‌ها، افزایش ایمنی و بهبود تجربه مشتریان را ممکن کرده است. هوش مصنوعی در این حوزه به عنوان یک ابزار قدرتمند برای حل مسائل پیچیده و بهینه‌سازی فرآیندها مطرح شده است.

هوش مصنوعی می‌تواند به روش‌های متعددی در حمل‌ونقل به کار گرفته شود. از جمله این کاربردها می‌توان به بهبود سیستم‌های مدیریت ترافیک، توسعه خودروهای خودران، پیش‌بینی تراکم ترافیک و بهینه‌سازی مسیرها اشاره کرد. این تکنولوژی‌ها می‌توانند به کاهش آلودگی محیط زیست، افزایش کارایی انرژی و کاهش زمان سفر کمک کنند. به علاوه، هوش مصنوعی می‌تواند به افزایش ایمنی در حمل‌ونقل کمک کند، به طوری که با تحلیل داده‌های جمع‌آوری شده از حسگرها و سیستم‌های مختلف، خطرات احتمالی را پیش‌بینی کرده و از وقوع حوادث جلوگیری کند.

در سال‌های اخیر، استارت‌آپ‌های زیادی در حوزه هوش مصنوعی و حمل‌ونقل شکل گرفته‌اند. این شرکت‌های نوپا با استفاده از فناوری‌های نوین سعی در حل مسائل موجود در صنعت حمل‌ونقل و ارائه راهکارهای نوآورانه دارند. با این حال، پیش‌بینی موفقیت این استارت‌آپ‌ها یک چالش بزرگ است. سرمایه‌گذاران و کارآفرینان نیاز به ابزارهایی دارند که بتوانند با استفاده از آن‌ها، موفقیت یا شکست استارت‌آپ‌ها را پیش‌بینی کنند و تصمیم‌گیری‌های سرمایه‌گذاری بهتری داشته باشند.

پایگاه داده کرانچ بیس به عنوان یکی از معتبرترین منابع اطلاعاتی در زمینه استارت‌آپ‌ها و سرمایه‌گذاری، اطلاعات جامعی در مورد شرکت‌ها، تأمین مالی، سرمایه‌گذاران و روندهای صنعت ارائه می‌دهد. این پایگاه داده به پژوهشگران و تحلیل‌گران امکان می‌دهد تا با استفاده از داده‌های بزرگ و تجزیه و تحلیل سیستماتیک، به بررسی و پیش‌بینی موفقیت استارت‌آپ‌ها بپردازند.

در این پژوهش، با استفاده از داده‌های موجود در پایگاه داده وبسایت کرانچ بیس و به کارگیری الگوریتم‌های داده‌کاوی، به دنبال ایجاد یک مدل پیش‌بینی‌کننده موفقیت استارت‌آپ‌های هوش مصنوعی در حوزه حمل‌ونقل هستیم. این مدل با تحلیل ویژگی‌های مختلفی مانند دورهای تأمین مالی، ادغام و تملیک، عرضه اولیه سهام و روندهای بازار مرتبط با استارت‌آپ‌ها، سعی در پیش‌بینی دقیق موفقیت یا شکست این شرکت‌ها دارد. الگوریتم‌های داده‌کاوی مانند درخت تصمیم، شبکه‌های عصبی مصنوعی و جنگل تصادفی توانایی تجزیه و تحلیل الگوهای پیچیده در داده‌ها، کشف روابط پنهان میان متغیرها و ارائه پیش‌بینی‌های دقیق در مورد آینده استارت‌آپ‌ها را دارند.

هدف این فصل، مرور منابع و بیان مبانی نظری برای ایجاد بینشی دقیق در مورد عوامل مؤثر بر موفقیت استارت‌آپ‌های هوش مصنوعی در صنعت حمل‌ونقل است که می‌تواند به کارآفرینان، سرمایه‌گذاران و سیاست‌گذاران کمک کند تا تصمیمات بهتری در زمینه سرمایه‌گذاری و توسعه فناوری‌های نوین اتخاذ کنند. با توجه به اهمیت و پیچیدگی موضوع، این پژوهش می‌تواند به عنوان یک مرجع ارزشمند برای مطالعات آینده در این حوزه مورد استفاده قرار گیرد.

* 1. مباني نظري

در این بخش به تعریف مفاهیم اساسی، اصول مرتبط و مبانی نظری مرتبط با موضوع تحقیق پرداخته خواهد شد. این مفاهیم شامل هوش مصنوعی، یادگیری ماشین، داده‌کاوی، استارت‌آپ و تأمین مالی می‌باشند. همچنین به اصول و مبانی نظری مرتبط با مدل‌های پیش‌بینی موفقیت استارت‌آپ‌ها و کاربردهای هوش مصنوعی در صنعت حمل‌ونقل پرداخته خواهد شد.

* + 1. هوش مصنوعی[[38]](#footnote-38)

هوش مصنوعی شاخه‌ای از علوم کامپیوتر است که هدف آن طراحی و توسعه سیستم‌ها و الگوریتم‌هایی است که بتوانند وظایف و فعالیت‌هایی که معمولاً نیاز به هوش انسانی دارند، مانند تشخیص صدا، تصمیم‌گیری، حل مسئله، یادگیری و درک زبان طبیعی، را انجام دهند (Russell & Norvig, 2016). این تکنولوژی با استفاده از روش‌های یادگیری ماشین و شبکه‌های عصبی مصنوعی، به سیستم‌ها این امکان را می‌دهد که از تجربیات گذشته خود بیاموزند، به داده‌های جدید واکنش نشان دهند و توانایی انجام وظایفی را داشته باشند که پیشتر تنها انسان‌ها قادر به انجام آن بودند (Goodfellow, 2016).

یکی از کاربردهای مهم هوش مصنوعی در صنعت حمل و نقل است که می‌تواند به افزایش ایمنی و کارایی سیستم‌های حمل و نقل کمک کند. سیستم‌های هوشمند مبتنی بر هوش مصنوعی قادرند الگوهای ترافیک را تجزیه و تحلیل کنند و با بهینه‌سازی مسیرها، کاهش زمان سفر و بهبود جریان ترافیک نقش مؤثری ایفا کنند. این سیستم‌ها می‌توانند از داده‌های زنده برای پیش‌بینی وضعیت ترافیک و ارائه پیشنهادهایی برای رانندگان استفاده کنند. برای نمونه، سیستم‌های تشخیص تصادف با استفاده از داده‌های جمع‌آوری شده می‌توانند تصادفات را به سرعت شناسایی کرده و اقدامات اضطراری لازم را پیشنهاد دهند (Schwarting et al., 2018).

حمل و نقل هوشمند یکی از کاربردهای اساسی هوش مصنوعی است که با بهره‌گیری از هوش مصنوعی می‌تواند وضعیت ترافیک را بهبود بخشیده و بهینه‌سازی مسیرها را انجام دهد. سیستم‌های حمل و نقل هوشمند با استفاده از داده‌های حسگرها و دوربین‌ها، وضعیت ترافیک را در زمان واقعی تجزیه و تحلیل می‌کنند و توصیه‌های لازم برای بهبود جریان ترافیک را ارائه می‌دهند (Chen & Englund, 2015).

یکی از پیشرفته‌ترین کاربردهای هوش مصنوعی در صنعت حمل و نقل، توسعه خودروهای خودران است که با استفاده از الگوریتم‌های یادگیری ماشین و داده‌های حسگرها قادر به حرکت بدون راننده انسانی هستند. این خودروها می‌توانند با شناسایی محیط اطراف خود و پردازش داده‌های حسگرها و دوربین‌ها، تصمیم‌گیری‌های لازم برای حرکت را انجام دهند. این فناوری، ضمن کاهش تصادفات، به افزایش کارایی و ایمنی سیستم‌های حمل و نقل نیز کمک می‌کند (Litman, 2017).

با وجود مزایای چشمگیر هوش مصنوعی در حمل و نقل، چالش‌هایی نیز وجود دارد. با این وجود، توسعه هوش مصنوعی فرصتی برای بهبود کیفیت زندگی، کاهش هزینه‌ها و افزایش ایمنی در حمل و نقل فراهم می‌آورد.

* + 1. یادگیری ماشین[[39]](#footnote-39) و داده کاوی[[40]](#footnote-40)

یادگیری ماشین یکی از شاخه‌های هوش مصنوعی است که به توسعه الگوریتم‌ها و مدل‌هایی می‌پردازد که قادر به یادگیری از داده‌ها و انجام وظایف بدون نیاز به برنامه‌نویسی صریح هستند. این الگوریتم‌ها از تجربیات گذشته برای بهبود عملکرد و انجام وظایف جدید استفاده می‌کنند. یادگیری ماشین به سیستم‌ها این امکان را می‌دهد که به طور خودکار از داده‌ها بیاموزند و تصمیم‌گیری کنند. الگوریتم‌های یادگیری ماشین به سه دسته کلی تقسیم می‌شوند: نظارت‌شده، بدون نظارت و تقویتی (Alpaydin, 2020).

یادگیری نظارت‌شده شامل آموزش یک مدل با استفاده از داده‌هایی است که شامل ورودی‌ها و خروجی‌های شناخته شده هستند. مدل با استفاده از این داده‌ها الگوهایی را می‌آموزد که به آن کمک می‌کند تا بتواند خروجی‌های جدید را پیش‌بینی کند(Goodfellow, 2016).

یادگیری بدون نظارت برخلاف یادگیری نظارت‌شده، شامل داده‌هایی است که تنها ورودی‌ها را شامل می‌شوند و مدل باید به طور خودکار الگوها و ساختارهای پنهان در داده‌ها را بیابد. خوشه‌بندی[[41]](#footnote-41) و کاهش ابعاد[[42]](#footnote-42) دو نمونه از تکنیک‌های یادگیری بدون نظارت هستند. خوشه‌بندی به فرآیند گروه‌بندی داده‌ها به دسته‌هایی بر اساس شباهت‌های موجود بین آنها گفته می‌شود. کاهش ابعاد به تکنیک‌هایی اطلاق می‌شود که به منظور کاهش تعداد ویژگی‌های موجود در یک مجموعه داده به کار می‌روند، در حالی که اطلاعات مهم حفظ می‌شوند (Hastie, 2009).

یادگیری تقویتی نوع دیگری از یادگیری ماشین است که در آن یک عامل با محیط تعامل دارد و از طریق پاداش‌ها و تنبیه‌ها می‌آموزد که چگونه بهترین اقدامات را انجام دهد. در یادگیری تقویتی، تلاش می‌کند تا با حداکثر کردن پاداش کل، یک سیاست بهینه برای انجام وظایف خود بیابد. این روش به طور گسترده‌ای در رباتیک و بازی‌های رایانه‌ای مورد استفاده قرار می‌گیرد (Sutton & Barto, 1998).

داده‌کاوی به فرآیند استخراج اطلاعات مفید و کشف الگوهای پنهان از مجموعه داده‌های بزرگ اشاره دارد. داده‌کاوی شامل مجموعه‌ای از تکنیک‌ها و الگوریتم‌ها است که برای تحلیل و کشف اطلاعات از داده‌ها به کار می‌روند. این تکنیک‌ها می‌توانند شامل دسته‌بندی، خوشه‌بندی و رگرسیون باشند. داده‌کاوی به سازمان‌ها این امکان را می‌دهد که از داده‌های خود برای بهبود تصمیم‌گیری‌ها و کشف فرصت‌های جدید استفاده کنند (Han et al., 2006).

داده‌کاوی و یادگیری ماشین به عنوان دو ابزار قدرتمند در تحلیل داده‌ها و استخراج دانش شناخته می‌شوند. این دو حوزه به هم وابسته هستند و اغلب به طور مشترک برای حل مسائل پیچیده و استخراج اطلاعات مفید از داده‌ها به کار می‌روند. به عنوان مثال، در صنعت حمل و نقل، داده‌کاوی و یادگیری ماشین می‌توانند به تحلیل الگوهای ترافیک، پیش‌بینی تقاضای سفر و بهبود مدیریت ترافیک کمک کنند (Chen & Englund, 2015).

یکی از کاربردهای مهم یادگیری ماشین و داده‌کاوی در پیش‌بینی موفقیت استارتاپ‌ها است. با تحلیل داده‌های تاریخی و استفاده از الگوریتم‌های یادگیری ماشین، می‌توان مدل‌هایی را توسعه داد که قادر به پیش‌بینی موفقیت یا شکست استارتاپ‌ها بر اساس معیارهای مختلف باشند. این مدل‌ها می‌توانند به سرمایه‌گذاران و کارآفرینان کمک کنند تا تصمیمات بهتری در مورد سرمایه‌گذاری و توسعه کسب‌وکارهای خود بگیرند (Kim et al., 2023).

* + 1. الگوریتم‌های داده کاوی

الگوریتم‌های داده‌کاوی ابزارهایی هستند که برای استخراج اطلاعات مفید و الگوها از مجموعه‌های بزرگ داده به کار می‌روند. این الگوریتم‌ها به کسب و کارها، محققان و تحلیل‌گران کمک می‌کنند تا داده‌های خام را به دانش ارزشمندی تبدیل کنند که می‌تواند برای تصمیم‌گیری‌های استراتژیک مورد استفاده قرار گیرد. در اینجا به بررسی برخی از انواع اصلی الگوریتم‌های داده‌کاوی و کاربردهای آن‌ها می‌پردازیم.

یکی از الگوریتم‌های معروف داده‌کاوی، الگوریتم‌های دسته‌بندی هستند. این الگوریتم‌ها برای تخصیص داده‌های ورودی به یکی از چند دسته از پیش تعریف شده استفاده می‌شوند. به عنوان مثال، جنگل تصادفی ، ماشین بردار پشتیبان ، و شبکه‌های عصبی مصنوعی از جمله این الگوریتم‌ها هستند. این الگوریتم‌ها در کاربردهایی نظیر تشخیص اسپم در ایمیل‌ها، پیش‌بینی بیماری‌ها در پزشکی و طبقه‌بندی مشتریان در بازاریابی مورد استفاده قرار می‌گیرند. دسته‌بندی به کسب و کارها این امکان را می‌دهد که مشتریان خود را بهتر بشناسند و خدمات یا محصولات خود را به صورت دقیق‌تر و هدفمندتر به آن‌ها ارائه دهند (Kotsiantis et al., 2006).

خوشه‌بندی یکی دیگر از انواع الگوریتم‌های داده‌کاوی است. در خوشه‌بندی، داده‌ها به گروه‌هایی تقسیم می‌شوند که داده‌های هر گروه به هم شبیه‌تر هستند. الگوریتم‌هایی مانند K-Means، خوشه‌بندی سلسله‌مراتبی و DBSCAN برای این منظور استفاده می‌شوند. این الگوریتم‌ها در کاربردهایی نظیر تقسیم‌بندی بازار، تحلیل رفتار مشتریان و تشخیص الگوهای ناهنجار به کار می‌روند.

الگوریتم‌های رگرسیون برای مدل‌سازی و تحلیل روابط بین متغیرهای وابسته و مستقل استفاده می‌شوند. رگرسیون خطی ، رگرسیون لجستیک و رگرسیون ریدج از جمله الگوریتم‌های رایج در این حوزه هستند. این الگوریتم‌ها در پیش‌بینی قیمت‌ها، تحلیل روندها و ارزیابی ریسک‌های مالی کاربرد دارند (Kotsiantis et al., 2006).

در نهایت، الگوریتم‌های مبتنی بر یادگیری عمیق شامل شبکه‌های عصبی کانولوشن و شبکه‌های عصبی بازگشتی نیز بخش مهمی از داده‌کاوی مدرن را تشکیل می‌دهند. این الگوریتم‌ها به ویژه در تحلیل تصاویر، پردازش زبان طبیعی و تشخیص گفتار کاربرد دارند. به عنوان مثال، شبکه‌های عصبی کانولوشن می‌توانند در تشخیص چهره و اشیاء در تصاویر ویدئویی به کار روند (Wang & Perez, 2017).

در مجموع، الگوریتم‌های داده‌کاوی ابزارهای قدرتمندی برای استخراج دانش از داده‌ها هستند که می‌توانند در زمینه‌های مختلفی از جمله بازاریابی، پزشکی، مالی و علمی کاربرد داشته باشند. این الگوریتم‌ها به کسب و کارها و محققان کمک می‌کنند تا تصمیمات بهتری بگیرند و فرآیندهای خود را بهینه‌سازی کنند.

* + 1. استارتاپ[[43]](#footnote-43)

استارتاپ یک نهاد انسانی است که با هدف ایجاد محصول یا خدمات نوآورانه در شرایط عدم قطعیت بالا شکل می‌گیرد. این نهادها توسط کارآفرینان تأسیس شده و تلاش می‌کنند تا به مدل‌های کسب‌وکار تکرارپذیر و مقیاس‌پذیر دست یابند. به طور معمول، استارتاپ‌ها با منابع محدود مالی و انسانی آغاز می‌شوند و با تکیه بر نوآوری و ارائه راه‌حل‌های جدید سعی در تأثیرگذاری بر بازار دارند.

از ویژگی‌های برجسته استارتاپ‌ها، مواجهه با ریسک‌های بالا است؛ چرا که بسیاری از آن‌ها ایده‌ها و محصولات جدیدی را توسعه می‌دهند که هنوز در بازار آزموده نشده‌اند. با این حال، موفقیت در این مسیر می‌تواند منجر به دستاوردهای اقتصادی و اجتماعی قابل توجهی شود (Kim et al., 2023).

استارتاپ‌ها معمولاً چرخه‌ای شامل ایده‌پردازی، تأسیس، رشد و خروج را طی می‌کنند. در مرحله ایده‌پردازی، شناسایی نیازهای بازار و طراحی راه‌حل‌های مرتبط انجام می‌شود. در مراحل بعدی، منابع مالی اولیه، معمولاً از طریق سرمایه‌گذاران فرشته یا صندوق‌های سرمایه‌گذاری خطرپذیر، جمع‌آوری می‌شوند تا توسعه محصول آغاز شود. مرحله رشد با تمرکز بر افزایش سهم بازار و جذب مشتریان همراه است، در حالی که مرحله خروج شامل فروش شرکت یا عرضه عمومی سهام است.

یکی از مهم‌ترین عوامل موفقیت استارتاپ‌ها، دسترسی به سرمایه و مدیریت آن است. تأمین مالی در مراحل مختلف (مانند سری‌های A، B و C) امکان توسعه و گسترش کسب‌وکار را فراهم می‌کند. علاوه بر این، انطباق با تغییرات تکنولوژیکی و بازار از جمله ویژگی‌های حیاتی برای بقای استارتاپ‌ها به شمار می‌رود.

استارتاپ‌ها نقش اساسی در نوآوری، ایجاد اشتغال و رشد اقتصادی ایفا می‌کنند. این نهادها با توسعه فناوری‌های نوین و تغییر الگوهای سنتی کسب‌وکار، تأثیر قابل توجهی بر صنایع مختلف دارند. با این حال، برای موفقیت نیاز به تیم‌های مدیریتی قوی، نوآوری مستمر و برنامه‌ریزی دقیق مالی دارند (Bento, 2018).

* + 1. دوره‌های تامین مالی[[44]](#footnote-44)

دوره‌های تامین مالی فرآیندهایی هستند که طی آن استارتاپ‌ها سرمایه مورد نیاز خود را از سرمایه‌گذاران جذب می‌کنند تا بتوانند رشد کنند، محصولات جدیدی توسعه دهند و بازارهای خود را گسترش دهند. این دوره‌ها معمولاً به ترتیب خاصی انجام می‌شوند و شامل مراحل مختلفی هستند که هر یک اهداف و ویژگی‌های خاص خود را دارند. مراحل اصلی تامین مالی شامل تامین مالی اولیه، سری A، سری B، سری C و مراحل بعدی، ادغام و تملیک، تامین مالی عمومی و تامین مالی جمعی می‌باشد.

* تامین مالی اولیه

این اولین مرحله از تامین مالی برای بسیاری از استارتاپ‌ها است که معمولاً از منابع شخصی بنیان‌گذاران، خانواده، دوستان یا سرمایه‌گذاران فرشته تأمین می‌شود. هدف این مرحله، جمع‌آوری سرمایه اولیه برای توسعه ایده، ساخت نمونه اولیه محصول، و انجام تحقیقات بازار است. این سرمایه معمولاً به صورت نقدی یا خدماتی ارائه می‌شود و به استارتاپ کمک می‌کند تا اولین قدم‌ها را بردارد و برای مراحل بعدی آمادگی کسب کند.

* سری A

پس از عبور از مرحله اولیه، استارتاپ‌ها به سرمایه بیشتری برای گسترش بازار و جذب مشتریان نیاز دارند. سرمایه‌گذاری سری A معمولاً توسط صندوق‌های سرمایه‌گذاری مخاطره‌آمیز انجام می‌شود. در این مرحله، استارتاپ‌ها باید ثابت کنند که مدل کسب‌وکارشان قابلیت مقیاس‌پذیری دارد و می‌تواند سودآور باشد. هدف اصلی در این مرحله، رشد سریع و افزایش سهم بازار است.

* سری B

سری B معمولاً به شرکت‌هایی اختصاص دارد که در حال حاضر به مرحله عملیاتی رسیده و مدل کسب‌وکار خود را اثبات کرده‌اند. این نوع سرمایه‌گذاری کمک می‌کند تا کسب‌وکارها به بازارهای جدید ورود کنند و عملیات خود را گسترش دهند. سرمایه‌گذاری سری B معمولاً توسط سرمایه‌گذاران بزرگتری که به دنبال شرکت‌های پایدارتری هستند، ارائه می‌شود.

* سری C و مراحل بعدی

این مرحله برای شرکت‌هایی است که به بلوغ بیشتری رسیده‌اند و به دنبال توسعه بیشتر، ورود به بازارهای بین‌المللی یا ادغام و تملیک هستند. هدف اصلی این مرحله، تقویت برند و افزایش سهم بازار است. این مراحل معمولاً سرمایه‌گذاران بزرگتر و بانک‌های سرمایه‌گذاری را به خود جذب می‌کنند.

* ادغام و تملیک[[45]](#footnote-45)

ادغام و تملیک فرآیندهایی هستند که طی آن شرکت‌ها برای افزایش ارزش و کارایی خود با سایر شرکت‌ها ادغام می‌شوند یا آن‌ها را خریداری می‌کنند. این فرآیندها می‌توانند به صورت داوطلبانه یا غیرداوطلبانه انجام شوند و معمولاً به منظور به دست آوردن مزایای استراتژیک، افزایش سهم بازار یا کسب فناوری‌های جدید انجام می‌شوند.

6. تامین مالی عمومی[[46]](#footnote-46)

آخرین مرحله تامین مالی برای بسیاری از استارتاپ‌ها، عرضه عمومی سهام است. در این مرحله، شرکت سهام خود را در بازار بورس عرضه می‌کند و به سرمایه‌گذاران عمومی اجازه می‌دهد تا سهام شرکت را خریداری کنند. عرضه عمومی سهام به استارتاپ‌ها کمک می‌کند تا سرمایه قابل توجهی جذب کنند و به مرحله جدیدی از رشد وارد شوند.

* تامین مالی جمعی[[47]](#footnote-47)

تامین مالی جمعی یک روش نسبتاً جدید برای جمع‌آوری سرمایه است که استارتاپ‌ها از طریق پلتفرم‌های آنلاین اقدام به جمع‌آوری سرمایه از تعداد زیادی از افراد می‌کنند. این روش به خصوص در مراحل اولیه و برای پروژه‌های خاص یا محصولات نوآورانه بسیار موثر است و به استارتاپ‌ها کمک می‌کند تا علاوه بر جذب سرمایه، علاقه مشتریان و بازار را نیز ارزیابی کنند.

* + 1. موفقیت کسب و کار

موفقیت کسب و کار به عنوان یک مفهوم چندوجهی و پیچیده، نیازمند تحلیل و بررسی دقیق عوامل متعددی است که هر کدام به نحوی می‌توانند در دستیابی به موفقیت نقش داشته باشند و نمایانگر عملکرد مالی و تجاری آن‌ها می‌باشد. معیارهای اصلی که به‌طور گسترده برای اندازه‌گیری موفقیت این نوع استارتاپ‌ها استفاده می‌شوند شامل عرضه عمومی سهام، ادغام و تملیک و تبدیل شدن به شرکت‌های یونیکورن است. این معیارها نه تنها بازتاب دهنده وضعیت اقتصادی شرکت هستند، بلکه بیانگر توانایی آن‌ها در ایجاد ارزش پایدار برای سرمایه‌گذاران و توسعه‌دهندگان آن‌ها نیز می‌باشند.

عرضه عمومی سهام یکی از معیارهای برجسته موفقیت است که نشان‌دهنده پذیرش عمومی و اعتبار بالای استارتاپ در بازارهای مالی است. شرکت‌هایی که وارد فرآیند عرضه عمومی سهام می‌شوند، از توانایی جمع‌آوری سرمایه قابل‌توجه برای توسعه بیشتر برخوردارند. این رویداد معمولاً با افزایش قابلیت اطمینان به استارتاپ همراه است و می‌تواند محرکی قوی برای رشد آتی باشد (Hagman, 2024).

ادغام و تملیک نیز یکی دیگر از نشانه‌های موفقیت است. این فرآیند نشان‌دهنده ارزشمندی فناوری یا خدمات استارتاپ برای سایر شرکت‌ها است. از آنجایی که بسیاری از استارتاپ‌ها توسط شرکت‌های بزرگ‌تر خریداری می‌شوند، این فرآیند فرصتی را برای دسترسی به منابع مالی، بازارهای جدید و دانش فنی گسترده‌تر فراهم می‌کند. بر اساس تحقیقات، ادغام و تملیک در استارتاپ‌های موفق اغلب بازگشت سرمایه قابل‌توجهی را برای سرمایه‌گذاران اولیه به همراه دارد (Ezangina & Malovichko, 2021).

تبدیل شدن به یونیکورن، که به استارتاپ‌هایی با ارزش بیش از یک میلیارد دلار اطلاق می‌شود، نیز معیار دیگری است که نشان‌دهنده پایداری و موفقیت تجاری استارتاپ است. این نوع استارتاپ‌ها اغلب دارای مدل‌های کسب‌وکار خلاقانه‌ای هستند که توانسته‌اند سرمایه‌گذاران بزرگ را جذب کنند و سهم عمده‌ای از بازار را به دست آورند (Simon, 2016). به‌طور خاص در حوزه هوش مصنوعی، یونیکورن‌ها به‌عنوان نمونه‌هایی از تحول فناوری و نوآوری‌های کاربردی در صنایع مختلف از جمله حمل‌ونقل مطرح هستند (Lee & Lin, 2020).

علاوه بر این، عوامل متعدد دیگری مانند تعداد و کیفیت دورهای تأمین مالی، توانایی تیم مدیریتی، و استفاده از فناوری‌های نوین، در دستیابی به این معیارهای موفقیت نقش دارند. تیم‌های مدیریتی توانمند و مبتکر می‌توانند با ارائه راهبردهای مؤثر، چالش‌های مختلف را مدیریت کرده و شرکت را به اهداف تعیین‌شده برسانند. همچنین، سرمایه‌گذاری بر روی نوآوری‌های فناوری نظیر یادگیری ماشین و تحلیل داده‌های بزرگ، ابزارهای کلیدی برای پیش‌بینی موفقیت استارتاپ‌ها محسوب می‌شوند (Kim et al., 2023).

در نهایت، موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل ترکیبی از دستیابی به معیارهای عرضه عمومی سهام ، ادغام و تملیک و یونیکورن شدن است. این پژوهش با استفاده از داده‌های بزرگ و روش‌های یادگیری ماشین، به ارائه الگوهایی برای پیش‌بینی موفقیت این استارتاپ‌ها کمک کرده‌اند که می‌تواند برای سرمایه‌گذاران و کارآفرینان ابزاری مؤثر جهت کاهش ریسک و افزایش بازدهی باشد (Kaunisto, 2024).

* + 1. هوش مصنوعی در صنعت حمل و نقل

هوش مصنوعی (AI) نقش مهمی در تحول صنعت حمل‌ونقل ایفا کرده و با ارائه راهکارهای هوشمند، باعث افزایش کارایی، ایمنی و پایداری این بخش شده است. یکی از مهم‌ترین کاربردهای AI در این حوزه، توسعه خودروهای خودران است که با استفاده از یادگیری ماشین، بینایی کامپیوتری و حسگرهای پیشرفته قادر به شناسایی موانع، عابران پیاده و سایر وسایل نقلیه هستند و می‌توانند بدون نیاز به راننده انسانی، مسیر خود را طی کنند (Abduljabbar et al., 2019). این فناوری علاوه بر کاهش تصادفات جاده‌ای، منجر به بهبود مدیریت ترافیک و کاهش مصرف سوخت می‌شود.

هوش مصنوعی همچنین در سیستم‌های مدیریت هوشمند ترافیک کاربرد گسترده‌ای دارد. با تحلیل داده‌های ترافیکی در زمان واقعی، سیستم‌های مبتنی بر AI می‌توانند مسیرهای بهینه را پیشنهاد داده، زمان‌بندی چراغ‌های راهنمایی را تنظیم کرده و ازدحام شهری را کاهش دهند (Hasan et al., 2021). این فناوری با ترکیب اینترنت اشیا (IoT) و تحلیل داده‌های بزرگ، اطلاعات دقیقی از جریان ترافیک ارائه می‌دهد که منجر به تصمیم‌گیری‌های سریع‌تر و مؤثرتر می‌شود.

در لجستیک و حمل‌ونقل کالا، هوش مصنوعی به بهینه‌سازی مسیرها، زمان‌بندی تحویل و کاهش هزینه‌ها کمک می‌کند. سیستم‌های مبتنی بر AI می‌توانند با تحلیل تقاضای بازار و شرایط جاده‌ای، مسیرهای بهینه را تعیین کرده و زمان تحویل را به حداقل برسانند. علاوه بر این، این فناوری در نگهداری پیشگیرانه وسایل نقلیه کاربرد دارد، به‌گونه‌ای که با تحلیل داده‌های حسگرها، می‌توان مشکلات احتمالی را قبل از وقوع شناسایی و از خرابی‌های ناگهانی جلوگیری کرد (Abduljabbar et al., 2019).

در حمل‌ونقل عمومی، هوش مصنوعی به بهبود خدمات و تجربه مسافران کمک می‌کند. با استفاده از تحلیل داده‌های مسافرتی، سیستم‌های AI می‌توانند الگوهای رفت‌وآمد را شناسایی کرده و ظرفیت ناوگان را در ساعات اوج مصرف تنظیم کنند. همچنین، پیش‌بینی تأخیرها و ارائه اطلاعات بلادرنگ به مسافران از دیگر مزایای استفاده از AI در این بخش است (Hasan et al., 2021).

با وجود این مزایا، چالش‌هایی همچون نیاز به زیرساخت‌های مناسب، مسائل مربوط به حریم خصوصی و امنیت داده‌ها همچنان وجود دارند. با این حال، با توسعه فناوری و سیاست‌گذاری‌های صحیح، می‌توان از پتانسیل‌های هوش مصنوعی برای ایجاد یک سیستم حمل‌ونقل پایدارتر، ایمن‌تر و هوشمندتر بهره برد (Abduljabbar et al., 2019; Hasan et al., 2021).

* + 1. صنعت حمل و نقل

حمل‌ونقل به‌عنوان یکی از زیربناهای اصلی توسعه اقتصادی و اجتماعی در جوامع مختلف نقش حیاتی دارد. این صنعت به‌عنوان یکی از ارکان اساسی فرآیندهای تجاری و اقتصادی، تأثیر قابل توجهی در تسهیل جابه‌جایی کالاها و افراد، کاهش هزینه‌های مبادلاتی و افزایش بهره‌وری دارد. بهبود زیرساخت‌های حمل‌ونقل موجب دسترسی بهتر به بازارهای داخلی و بین‌المللی، کاهش هزینه‌های لجستیکی و رشد سرمایه‌گذاری می‌شود. علاوه بر این، حمل‌ونقل کارآمد به بهبود کیفیت زندگی، افزایش اشتغال و توزیع عادلانه‌تر فرصت‌های اقتصادی کمک می‌کند و نقش مهمی در توسعه پایدار شهری و کاهش آلودگی‌های زیست‌محیطی دارد (Hasan et al., 2019).

در سال‌های اخیر، با افزایش جمعیت و نیاز به تحرک بیشتر، سیستم‌های حمل‌ونقل در اکثر نقاط جهان با چالش‌های زیادی روبه‌رو هستند. مشکلاتی مانند ترافیک سنگین، آلودگی هوا، ایمنی و هزینه‌های بالای حمل‌ونقل، نیاز به راه‌حل‌های نوآورانه برای بهبود عملکرد این سیستم‌ها را اجتناب‌ناپذیر کرده است. در این راستا، استفاده از فناوری‌های نوین و به‌ویژه هوش مصنوعی به‌عنوان ابزاری کارآمد و مؤثر، تحول بزرگی در این صنعت به وجود آورده است (Nikitas et al., 2020).

* + 1. هوش مصنوعی در صنعت حمل و نقل

در حوزه حمل‌ونقل، هوش مصنوعی با ارائه راه‌حل‌هایی برای بهینه‌سازی مصرف منابع، کاهش تصادفات، بهبود ایمنی و تسهیل حرکت، تاثیر شگرفی دارد (Hasan et al., 2019).

هوش مصنوعی می‌تواند در بسیاری از جنبه‌های حمل‌ونقل به‌کار گرفته شود. یکی از مهم‌ترین کاربردهای آن در مدیریت ترافیک است. سیستم‌های هوشمند حمل‌ونقل می‌توانند با تجزیه و تحلیل داده‌های ترافیکی و پیش‌بینی شرایط آینده، زمان‌بندی چراغ‌های راهنمایی را بهینه کنند، مسیرهای پر ترافیک را شناسایی کرده و به رانندگان اطلاع دهند تا از وقوع ترافیک‌های سنگین جلوگیری کنند. این تکنولوژی‌ها می‌توانند موجب کاهش زمان سفر، کاهش مصرف سوخت و بهبود شرایط ترافیکی شوند (Abduljabbar et al., 2019).

از دیگر کاربردهای هوش مصنوعی در حمل‌ونقل می‌توان به خودروهای خودران اشاره کرد. خودروهای خودران که به‌طور کامل با استفاده از الگوریتم‌های هوش مصنوعی کنترل می‌شوند، قادرند بدون دخالت انسان به رانندگی بپردازند. این خودروها می‌توانند با تحلیل دقیق داده‌ها از دوربین‌ها، حسگرها و سیستم‌های راداری، شرایط جاده و ترافیک را پیش‌بینی کرده و بهترین مسیر را انتخاب کنند. استفاده از خودروهای خودران در کاهش تصادفات، افزایش ایمنی و بهینه‌سازی مصرف انرژی نقش مهمی دارد (Nikitas et al., 2020).

همچنین، هوش مصنوعی در بهبود سیستم‌های حمل‌ونقل عمومی مانند اتوبوس‌ها و قطارها نیز کاربرد دارد. این سیستم‌ها می‌توانند با استفاده از داده‌های واقعی و پیش‌بینی‌های هوشمند، مسیرهای بهینه‌تر و زمان‌بندی‌های دقیق‌تری را برای سرویس‌دهی به مسافران ایجاد کنند. به‌عنوان مثال، سیستم‌های هوشمند می‌توانند پیش‌بینی کنند که در ساعت‌های خاصی از روز، کدام مسیرها بیشترین ترافیک را خواهند داشت و بر اساس آن، تعداد وسایل نقلیه یا تعداد ایستگاه‌ها را تنظیم کنند تا از شلوغی و تاخیر جلوگیری کنند (Hasan et al., 2019).

یکی دیگر از حوزه‌هایی که هوش مصنوعی در حمل‌ونقل به‌طور فزاینده‌ای مورد توجه قرار گرفته است، بهینه‌سازی مسیرها و مدیریت پارکینگ است. با استفاده از داده‌های جمع‌آوری شده از سیستم‌های حمل‌ونقل و الگوریتم‌های هوش مصنوعی، می‌توان به رانندگان پیشنهاد داد که از کدام مسیرها برای رسیدن به مقصد خود استفاده کنند یا در کدام پارکینگ‌ها فضای خالی وجود دارد. این امر می‌تواند به کاهش زمان جستجوی پارکینگ و کاهش ترافیک منجر شود (Abduljabbar et al., 2019).

به‌طور کلی، هوش مصنوعی در حمل‌ونقل به‌عنوان ابزاری است که می‌تواند ضمن کاهش هزینه‌ها و زمان سفر، به بهبود کیفیت خدمات حمل‌ونقل و افزایش رضایت مشتریان کمک کند. این سیستم‌ها با تحلیل داده‌ها و پیش‌بینی روندهای آینده، می‌توانند تصمیم‌گیری‌های بهتری را در مورد مدیریت ترافیک، سیستم‌های حمل‌ونقل عمومی و خودروهای خودران اتخاذ کنند (Nikitas et al., 2020).

* + 1. کرانچ بیس

کرانچ بیس یک پایگاه داده جامع است که اطلاعات دقیقی در مورد استارت‌آپ‌ها، سرمایه‌گذاران، دورهای تأمین مالی و روندهای صنعت ارائه می‌دهد. این پایگاه داده به پژوهشگران و تحلیل‌گران امکان می‌دهد تا با استفاده از داده‌های بزرگ و تجزیه و تحلیل سیستماتیک، به بررسی و پیش‌بینی موفقیت استارت‌آپ‌ها بپردازند (Bento, 2018; Kim et al., 2023).

در وبسایت کرانچ بیس، حوزه حمل و نقل به 33 زیر صنعت تقسیم می‌شوند که در جدول 2-1 آورده شده است.

**جدول 2- 1- حوزه‌های مختلف صنعت حمل و نقل**

|  |  |  |
| --- | --- | --- |
| 1 | حمل و نقل هوایی | Air Transportation |
| 2 | صنعت خودرو | Automotive |
| 3 | وسایل نقلیه خودران | Autonomous Vehicles |
| 4 | اشتراک‌گذاری خودرو | Car Sharing |
| 5 | زیرساخت‌های شارژ | Charging Infrastructure |
| 6 | خدمات پیک | Courier Service |
| 7 | خدمات تحویل | Delivery Service |
| 8 | خودروی برقی | Electric Vehicle |
| 9 | خدمات کشتی‌رانی | Ferry Service |
| 10 | مدیریت ناوگان | Fleet Management |
| 11 | تحویل غذا | Food Delivery |
| 12 | خدمات باربری | Freight Service |
| 13 | حمل و نقل آخرین مایل | Last Mile Transportation |
| 14 | خدمات لیموزین | Limousine Service |
| 15 | لجستیک | Logistics |
| 16 | حمل و نقل دریایی | Marine Transportation |
| 17 | پارکینگ | Parking |
| 18 | بنادر و لنگرگاه‌ها | Ports and Harbors |
| 19 | تأمین کالا | Procurement |
| 20 | حمل و نقل عمومی | Public Transportation |
| 21 | راه‌آهن | Railroad |
| 22 | وسایل نقلیه تفریحی | Recreational Vehicles |
| 23 | اشتراک‌گذاری سفر | Ride Sharing |
| 24 | تحویل در همان روز | Same Day Delivery |
| 25 | حمل و نقل دریایی | Shipping |
| 26 | کارگزار حمل و نقل | Shipping Broker |
| 27 | سفرهای فضایی | Space Travel |
| 28 | مدیریت زنجیره تأمین | Supply Chain Management |
| 29 | خدمات تاکسی | Taxi Service |
| 30 | حمل و نقل | Transportation |
| 31 | خودکارسازی انبار | Warehouse Automation |
| 32 | انبارداری | Warehousing |
| 33 | حمل و نقل آبی | Water Transportation |

* 1. پیشینه پژوهش

تحقیقات زیادی در زمینه پیش بینی موفقیت استارتاپ‌ها و همچنین کاربردهای هوش مصنوعی در صنعت حمل‌ونقل انجام شده است. در مقاله "پیش‌بینی موفقیت استارتاپ‌ها با استفاده از یادگیری ماشین" که توسط بنتو (Bento, 2017) نوشته شده است، تلاش شده تا مدل پیش‌بینی موفقیت استارتاپ‌ها با استفاده از داده‌های سایت کرانچ بیس توسعه یابد. این مطالعه به بررسی استفاده از الگوریتم‌های یادگیری ماشین مانند جنگل تصادفی، ماشین بردار پشتیبان و رگرسیون لجستیک می‌پردازد تا بتواند با دقت بالا استارتاپ‌های موفق را از استارتاپ‌های ناموفق تشخیص دهد. داده‌های مورد استفاده شامل اطلاعات بیش از 140,000 استارتاپ از سال‌های مختلف است که به تحلیل و پردازش آن‌ها پرداخته شده است. یکی از نکات کلیدی این پژوهش، تاکید بر اهمیت داده‌های مالی و سرمایه‌گذاری در پیش‌بینی موفقیت استارتاپ‌ها است. همچنین، میزان سرمایه‌گذاری و تعداد دورهای سرمایه‌گذاری انجام شده از جمله متغیرهای مهم در مدل پیش‌بینی بودند. علاوه بر این، مقاله نشان می‌دهد که استفاده از الگوریتم جنگل تصادفی توانسته است به نرخ مثبت واقعی[[48]](#footnote-48) 94 درصد دست یابد که این میزان بالاترین دقت ثبت شده در این حوزه با استفاده از داده‌های کرانچ بیس است. این مطالعه با ارائه مدل‌هایی که می‌توانند به عنوان سیستم‌های پشتیبان تصمیم برای سرمایه‌گذاران و کارآفرینان استفاده شوند، نشان می‌دهد که چگونه می‌توان از تکنیک‌های داده‌کاوی و یادگیری ماشین برای بهبود فرآیند تصمیم‌گیری و افزایش دقت پیش‌بینی استفاده کرد. نتایج این تحقیق می‌تواند به بهبود فرآیندهای انتخاب استارتاپ‌های موفق و افزایش بازده سرمایه‌گذاری‌ها کمک کند (Bento, 2017).

عبدالجبار و همکاران (Abduljabbar et al., 2019) به بررسی کاربردهای مختلف هوش مصنوعی در حمل‌ونقل پرداخته و تأثیرات مثبت آن بر کاهش ترافیک و افزایش ایمنی را تحلیل کرده‌اند. این مطالعه نشان می‌دهد که استفاده از هوش مصنوعی می‌تواند به بهبود سیستم‌های مدیریت ترافیک و توسعه خودروهای خودران کمک کند. در بررسی پیش‌بینی تراکم ترافیک با استفاده از الگوریتم‌های هوش مصنوعی نشان داده‌اند که این روش‌ها دقت بالایی دارند و می‌توانند به کاهش زمان سفر و بهبود تجربه مشتریان کمک کنند.

خاویر آرورو و همکاران(Arroyo et al., 2019) به بررسی توانایی مدل‌های یادگیری ماشین در پشتیبانی از تصمیم‌گیری در سرمایه‌گذاری‌های اولیه می‌پردازد. این پژوهش با استفاده از داده‌های کرانچ بیس بیش از 120,000 شرکت نوپا را تحلیل کرده و تلاش دارد تا مدلی جامع برای پیش‌بینی موفقیت این شرکت‌ها ایجاد کند.

در این مطالعه، مدل‌های مختلف یادگیری ماشین از جمله جنگل تصادفی، درختان تصمیم، و تقویت گرادیان مورد ارزیابی قرار گرفته‌اند. این مدل‌ها به پیش‌بینی رویدادهایی مانند جذب دور جدیدی از سرمایه‌گذاری، ادغام و تملیک، عرضه اولیه عمومی و حتی تعطیلی شرکت پرداخته‌اند. نتایج نشان می‌دهد که مدل تقویت گرادیان با دقت 82% بهترین عملکرد را داشته و می‌تواند به سرمایه‌گذاران در انتخاب شرکت‌های مناسب برای سرمایه‌گذاری کمک کند. ویژگی‌های مهم برای پیش‌بینی موفقیت شرکت‌ها شامل سن شرکت، تعداد بنیان‌گذاران، میزان سرمایه جذب شده و حضور شرکت در شبکه‌های اجتماعی مانند لینکدین می‌باشد. این مطالعه همچنین نشان می‌دهد که استفاده از مدل‌های یادگیری ماشین می‌تواند نرخ موفقیت سرمایه‌گذاران را افزایش دهد و به آن‌ها در ایجاد پرتفوی کم‌ریسک‌تر و با بازدهی بالاتر کمک کند. این پژوهش با ارائه یک رویکرد داده‌محور برای پیش‌بینی موفقیت شرکت‌های نوپا، نشان می‌دهد که یادگیری ماشین می‌تواند ابزار قدرتمندی برای بهبود فرآیند تصمیم‌گیری در سرمایه‌گذاری‌های خطرپذیر باشد(Arroyo et al., 2019).

ژبیکوفسکی و آنتوسیوک (Żbikowski & Antosiuk, 2021) بر روی توسعه مدلی برای پیش‌بینی موفقیت شرکت‌ها تمرکز دارد. این مطالعه به طور خاص به بررسی چالش‌های ناشی از تعصبات موجود در داده‌ها پرداخته و سعی دارد با جلوگیری از نفوذ اطلاعات ناشناخته در زمان تصمیم‌گیری، مدلی کاربردی و قابل استفاده در دنیای واقعی ارائه دهد. داده‌های مورد استفاده در این پژوهش از پلتفرم کرانچ بیس جمع‌آوری شده که شامل اطلاعات 213171 شرکت است. این پژوهش سه الگوریتم مختلف شامل رگرسیون لجستیک، ماشین بردار پشتیبانی، و الگوریتم XGBoost را برای پیش‌بینی موفقیت شرکت‌ها مقایسه می‌کند. نتایج نشان می‌دهد که الگوریتم XGBoost بهترین عملکرد را از نظر دقت و بازشناسی دارد، هرچند که هر سه مدل دقت بالایی را در پیش‌بینی موفقیت شرکت‌ها نشان داده‌اند. با این حال، مشکل اصلی که مدل‌ها با آن مواجه بودند، میزان پایین بازشناسی بود که منجر به نادیده گرفتن بسیاری از استارتاپ‌های موفق شد. مهم‌ترین یافته‌های این مطالعه شامل شناسایی عواملی مانند کشور، منطقه فعالیت شرکت و صنعت مرتبط به عنوان متغیرهای اصلی در پیش‌بینی موفقیت است. نتایج این پژوهش می‌تواند به عنوان یک سیستم پشتیبانی تصمیم برای صندوق‌های سرمایه‌گذاری مورد استفاده قرار گیرد و به بهبود عملکرد آن‌ها در انتخاب شرکت‌های موفق کمک کند(Żbikowski & Antosiuk, 2021).

کیم و همکاران (Kim et al., 2023) تلاش می‌کند تا شکاف پژوهشی موجود در بررسی موفقیت استارتاپ‌ها را پر کند. در حالی که مطالعات پیشین بیشتر بر شرایط داخلی استارتاپ‌ها متمرکز بوده‌اند، این مطالعه بر تاثیرات ویژگی‌های صنعت بر موفقیت استارتاپ‌ها نیز توجه دارد. داده‌های مورد استفاده از سایت کرانچ بیس جمع‌آوری شده و شامل اطلاعات 218,207 شرکت از سال 2011 تا 2021 می‌باشد. پس از پیش‌پردازش داده‌ها، از شش مدل یادگیری ماشین برای پیش‌بینی موفقیت استارتاپ‌ها استفاده شده و ویژگی‌های مهم برای پیش‌بینی شناسایی شده‌اند. نتایج نشان می‌دهد که عواملی مانند میزان نمایش رسانه‌ای، میزان تامین مالی، سطح همگرایی صنعت، و سطح وابستگی صنعت، تاثیر بسزایی در موفقیت استارتاپ‌ها دارند. این مطالعه مدل‌های یادگیری ماشینی مختلفی را برای پیش‌بینی موفقیت به کار می‌برد و نشان می‌دهد که استفاده از ویژگی‌های صنعت می‌تواند دقت پیش‌بینی را بهبود بخشد. به طور کلی، این پژوهش نشان می‌دهد که ترکیب ویژگی‌های داخلی و خارجی شرکت‌ها، به ویژه ویژگی‌های صنعت، می‌تواند ابزار قدرتمندی برای پیش‌بینی موفقیت استارتاپ‌ها فراهم کند (Kim et al., 2023).

یی فانک، تی و همکاران(Te et al., 2023)، محققان به بررسی عواملی می‌پردازند که بر موفقیت استارتاپ‌ها در جذب سرمایه‌گذاری‌های اولیه تأثیر می‌گذارند. این مطالعه به طور خاص از داده‌های دو پلتفرم معروف کرانچ بیس و لینکدین برای ایجاد مجموعه‌ای از ویژگی‌های پیش‌بینی استفاده کرده و دو مدل یادگیری ماشین با استفاده از روش تقویت گرادیان سبک[[49]](#footnote-49) آموزش داده است. مدل‌ها شامل یک مدل مستقل مبتنی بر داده‌های لینکدین و یک مدل ترکیبی از داده‌های کرانچ بیس و لینکدین هستند.

نتایج این پژوهش نشان می‌دهد که استفاده از داده‌های لینکدین به عنوان منبع مکمل می‌تواند دقت پیش‌بینی موفقیت استارتاپ‌ها را افزایش دهد، به طوری که مدل ترکیبی با دقت 84 درصد عملکرد بهتری نسبت به مدل مستقل مبتنی بر داده‌های کرانچ بیس دارد. این مطالعه همچنین با استفاده از روش شاپلی[[50]](#footnote-50) به تحلیل ویژگی‌های مهم برای پیش‌بینی موفقیت پرداخته و نشان داده است که ویژگی‌های مربوط به سرمایه‌گذاری و همچنین ویژگی‌های دموگرافیک و تجربه کاری مؤسسین تأثیر بسزایی در موفقیت استارتاپ‌ها دارند. این تحقیق با ترکیب داده‌های متنوع و استفاده از روش‌های پیشرفته یادگیری ماشین، می‌تواند به عنوان یک سیستم پشتیبان تصمیم برای سرمایه‌گذاران و کارآفرینان استفاده شود. نتایج نشان می‌دهد که داده‌های موجود در شبکه‌های اجتماعی حرفه‌ای مانند لینکدین می‌تواند اطلاعات ارزشمندی را برای پیش‌بینی موفقیت استارتاپ‌ها فراهم کند و به بهبود فرآیند غربالگری و ارزیابی اولیه سرمایه‌گذاری کمک کند (Te et al., 2023).

آنا دیاس و آلِساندرو ماگرینی (Deias & Magrini, 2023) به بررسی چگونگی تاثیر پویایی‌های تأمین مالی بر شانس موفقیت خروج یک شرکت نوپا پرداخته‌اند. این پژوهش بر اساس مدل رگرسیون لجستیک چندوجهی طراحی شده است که به تحلیل رویدادهای خروج و تعطیلی شرکت‌ها در رابطه با میزان سرمایه‌گذاری سهامی در دوره‌های مختلف پرداخته و همچنین متغیرهای مکانی، بخشی، سن و شبکه‌های ارتباطی را کنترل می‌کند. مطالعه نشان می‌دهد که پویایی‌های تامین مالی تأثیر قابل توجهی بر شانس خروج موفقیت‌آمیز و خطر تعطیلی شرکت‌ها دارند. برای مثال، افزایش سرمایه در مراحل اولیه (مانند سرمایه‌گذاری فرشته) خطر تعطیلی را کاهش می‌دهد اما تأثیر معناداری بر شانس خروج موفقیت‌آمیز ندارد. در مقابل، افزایش سرمایه در مراحل میانی (سری B) شانس خروج موفقیت‌آمیز و خطر تعطیلی را افزایش می‌دهد، در حالی که افزایش سرمایه در مراحل پایانی (سری D و بیشتر) شانس خروج موفقیت‌آمیز را افزایش می‌دهد بدون تاثیر معنادار بر خطر تعطیلی. این پژوهش با استفاده از داده‌های جامع کرانچ بیس ، دیدی کلی از سناریوهای مختلف در چرخه عمر یک شرکت نوپا ارائه می‌دهد و به بهبود تصمیم‌گیری سرمایه‌گذاران کمک می‌کند. یافته‌های این مطالعه بر اهمیت پویایی‌های تامین مالی سهامی در موفقیت شرکت‌ها تاکید دارد و نشان می‌دهد که توجه به نوع و زمان‌بندی سرمایه‌گذاری می‌تواند تأثیر قابل توجهی بر نتایج نهایی شرکت‌ها داشته باشد (Deias & Magrini, 2023).

مارک پوتانین و همکاران (Potanin et al., 2023) به بررسی پیش‌بینی موفقیت استارتاپ‌ها و شبیه‌سازی پرتفوی سرمایه‌گذاری خطرپذیر پرداخته‌اند. این پژوهش به بررسی موفقیت استارتاپ‌ها در مراحل سرمایه‌گذاری سری B و سری C می‌پردازد و تلاش دارد تا موفقیت‌های کلیدی نظیر دستیابی به عرضه اولیه عمومی، رسیدن به وضعیت یونیکورن، یا انجام موفقیت‌آمیز ادغام و تملک را پیش‌بینی کند.

محققان از مدل یادگیری عمیق نوآورانه‌ای استفاده کرده‌اند که انواع مختلفی از داده‌ها نظیر معیارهای تأمین مالی، ویژگی‌های بنیان‌گذاران و دسته‌بندی صنعت را ترکیب می‌کند. یکی از ویژگی‌های متمایز این پژوهش، استفاده از الگوریتم بک‌تست جامع[[51]](#footnote-51) است که فرآیند سرمایه‌گذاری خطرپذیر را شبیه‌سازی می‌کند. این شبیه‌سازی به ارزیابی عملکرد مدل در برابر داده‌های تاریخی کمک کرده و بینش‌های کاربردی را در زمینه استفاده عملی از آن در سرمایه‌گذاری‌های واقعی ارائه می‌دهد. نتایج این مطالعه نشان می‌دهد که مدل پیشنهادی با استفاده از داده‌های کرانچ بیس توانسته است رشد سرمایه 14 برابری را تحقق بخشد و استارتاپ‌های با پتانسیل بالا را در مراحل اولیه سرمایه‌گذاری شناسایی کند. یافته‌های تجربی نشان می‌دهند که استفاده از مجموعه داده‌های متنوع در بهبود دقت پیش‌بینی مدل تأثیر بسزایی دارد. این پژوهش نشان می‌دهد که مدل‌های یادگیری عمیق و داده‌های غیر ساختاریافته می‌توانند در پیش‌بینی موفقیت استارتاپ‌ها نقش مؤثری ایفا کنند و زمینه‌ساز پیشرفت‌های آینده در این حوزه باشند(Potanin et al., 2023).

در ادامه در جدول 2-2 پژوهش‌های موجود در این حوزه به صورت خلاصه داده شده‌اند:

**جدول 2- 2- پژوهش‌های پیشین و نتایج آن‌ها**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ردیف | محقق  (سال) | موضوع | روش تحلیل داده | جامعه آماری  و نمونه | فاکتورهای پیش بینی موفقیت | متغیرهای اصلی | یافته­های تحقیق |
| **1** | **(Kim et al., 2023)** | چگونه در بازار موفق شویم؟ پیش بینی موفقیت راه اندازی با استفاده از رویکرد یادگیری ماشینی | Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, Naïve Bayes | Crunchbase | **عرضه عمومی اولیه (IPO)** | **نوع شرکت (Company type)**  **محبوبیت شرکت (Company popularity)**  **جذابیت شرکت (Company attractiveness)**  **افزایش محبوبیت (Growth of popularity)**  **سن شرکت (Company age)**  **تعداد کارکنان (Employees)**  **موقعیت مقر اصلی شرکت (HQ location)**  **تعداد دفعات ظاهر شدن در رسانه‌ها (Exposure)**  **اخیر بودن تأمین مالی (Funding recency)**  **بسامد تأمین مالی (Funding frequency)**  **مقدار مالی تأمین شده (Funding monetary)**  **نوع سرمایه‌گذاران (Investor type)**  **مشارکت سرمایه‌گذاران برتر (Top investors' involvement)**  **تعداد سرمایه‌گذاران رهبر (Number of lead investor)**  **کل تعداد سرمایه‌گذاران (Total number of investors)**  **مقدار فناوری موجود (Amount of technology available)**  **تعداد برنامه‌ها (Number of applications)**  **تعداد دانلودهای برنامه‌ها (Total downloads of applications)**  **تعداد اختراعات ثبت شده (Number of patents)**  **تعداد علائم تجاری ثبت شده (Number of trademarks)**  **هزینه‌های IT (IT spends)** | دو عامل موفقیت استارتاپ‌ها: همگرایی با صنعت و در معرض رسانه قرار داشتن و یا نوع دیده شدن محصول |
| **2** | **(Bento, 2017)** | پیش‌بینی موفقیت استارت‌آپ‌ها با استفاده از یادگیری ماشین | * Random Forest * Support Vector Machine * Logistic Regression | Crunchbase | عرضه اولیه عمومی (IPO)  ادغام و خرید (M&A) | **تعداد دوره‌های تأمین مالی (Funding Rounds)**  **مجموع مبلغ تأمین مالی (Total Funding USD)**  **وضعیت استارت‌آپ (Status)**  **تأثیر سرمایه‌گذاری سرمایه داران مخاطره‌پذیر (Impact of Venture Capital)**  **تاریخ تأسیس (Age of the Company)**  **متغیرهای مربوط به جغرافیا (Geographical Variables)** | افزایش قابل توجه در دقت پیش‌بینی موفقیت استارت‌آپ‌ها با استفاده از الگوریتم‌های یادگیری ماشین و تعریف جدید برای متغیرهای هدف |
| **3** | **(Te et al., 2023)** | **تحلیل داده‌های کرانچ بیس و لینکدین برای پیش‌بینی موفقیت استارت‌آپ‌ها در جذب سرمایه‌ی سری A** | Light Gradient Boosting Machine - LGBM | Crunchbase  LinkedIn | موفقیت در جذب سرمایه‌ی سری A | **مبلغ کل سرمایه جذب شده (Total Funding Amount)**  **تعداد دوره‌های تأمین مالی (Number of Funding Rounds)**  **(Startup Status)**  **تاریخ تأسیس (Founding Date)**  **صنعت (Industry)**  **موقعیت جغرافیایی (Geographical Location)**  **تعداد کارمندان (Number of Employees)** | نشان دادن اینکه استفاده از داده‌های LinkedIn به عنوان منبع مکمل به داده‌های Crunchbase به بهبود نتایج پیش‌بینی کمک می‌کند و میانگین دقت زیر منحنی (AUC) به ۸۴٪ می‌رسد. |
| **4** | **(Deias & Magrini, 2023)** | **تأثیر پویایی تامین مالی سهام بر موفقیت کسب‌وکارها** | Multinomial Logistic Regression | Crunchbase | فروش سهام (Sale of Equity)  ادغام و تملیک (Mergers and Acquisitions) | **مبلغ جمع‌آوری‌شده در هر دور تأمین مالی (Amount Raised in Each Funding Round)**  **تعداد دوره‌های تأمین مالی (Number of Funding Rounds)**  **موقعیت جغرافیایی (Geographic Location)**  **بخش اقتصادی (Industry Sector)**  **سن شرکت (Age of the Company)**  **پیوندهای شبکه‌ای (Network Links)** | نشان داده است که دینامیک تامین مالی سهام تاثیر قابل توجهی بر احتمال خروج موفق کسب‌وکارها دارد و این اثرگذاری نه تنها در مراحل اولیه بلکه در مراحل پیشرفته‌تر توسعه کسب‌وکار نیز مشهود است |
| **5** | **(Potanin et al., 2023)** | **پیش‌بینی موفقیت استارت‌آپ‌ها و شبیه‌سازی پورتفولیو سرمایه‌گذاری مخاطره‌آمیز با استفاده از داده‌های کرانچ بیس** | deep learning models | Crunchbase | عرضه عمومی اولیه (IPO)  خرید و ادغام (Acquisition - ACQ)  وضعیت یونیکورن (Unicorn Status - UNIC) | **مبلغ جمع‌آوری‌شده در هر دور تأمین مالی (Amount Raised in Each Funding Round)**  **تعداد دوره‌های تأمین مالی (Number of Funding Rounds)**  **موقعیت جغرافیایی (Geographic Location)**  **بخش صنعتی (Industry Sector)**  **سن شرکت (Age of the Company)**  **ویژگی‌های بنیانگذاران (Founders Features)**  **ویژگی‌های سرمایه‌گذاران (Investors Features)**  **ویژگی‌های دوره‌های سرمایه‌گذاری (Rounds Features)** | یافته‌های تحقیق نشان می‌دهند که مدل‌های یادگیری عمیق می‌توانند دقیقاً موفقیت استارت‌آپ‌ها را پیش‌بینی کنند. ترکیب ویژگی‌های بنیان‌گذاران و سرمایه‌گذاران به پیش‌بینی مؤثر موفقیت کمک می‌کند و بازده سرمایه‌گذاری‌ها را بهبود می‌بخشد. |
| **6** | **(Arroyo et al., 2019)** | ارزیابی عملکرد یادگیری ماشین برای پشتیبانی تصمیم‌گیری در سرمایه‌گذاری‌های سرمایه خطر | Logistic Regression, Decision Tree, Random Forest, Gradient Boosting, Support Vector Machine, Naïve Bayes | Crunchbase | خریداری شدن شرکت (Acquisition)  جذب سرمایه دور بعدی (Next Round Funding)  بسته شدن شرکت (Closure)  عرضه عمومی اولیه (IPO) | **موقعیت جغرافیایی (Geographic Location)**  **بخش صنعتی (Industry Sector)**  **سن شرکت (Company Age)**  **تعداد بنیان‌گذاران (Number of Founders)**  **تعداد دوره‌های تأمین مالی (Number of Funding Rounds)**  **مبلغ کل سرمایه جذب شده (Total Amount Raised)**  **زمان سپری شده از آخرین دور تأمین مالی (Time Since Last Funding Round)**  **تعداد سرمایه‌گذاران برجسته (Number of Renowned Investors)**  **اطلاعات مربوط به شبکه‌های اجتماعی (Social Media Presence): حضور شرکت در شبکه‌های اجتماعی مانند LinkedIn، Twitter و Facebook.** | نشان دهنده‌ی اینکه مدل‌های یادگیری ماشین می‌توانند به سرمایه‌گذاران خطر کمک کنند تا فرصت‌های سرمایه‌گذاری را شناسایی کرده و ریسک سرمایه‌گذاری‌های بالقوه را بهتر ارزیابی کنند. |
| **7**  **44** | **(Prohorovs et al., 2019)** | **عوامل موفقیت استارتاپ در مرحله جذب سرمایه: دیدگاه بنیان‌گذاران** | **Factor Analysis, Binary Logistic Regression** | **بنیان‌گذاران 40 استارتاپ در لتونی و روسیه** | مهارت‌های مدیریتی تیم و بنیان‌گذار، ارتباطات مستقیم با سرمایه‌گذاران، آمادگی شرکت، پشتیبانی مدیران، مدل کسب و کار مالی | تخصص تیم مدیریتی، تجربه قبلی، دسترسی به منابع مالی، آمادگی برای توسعه | تیم‌های موفق شامل بنیان‌گذارانی با تجربه قبلی و مهارت‌های مدیریتی بودند که ارتباطات مستقیم با سرمایه‌گذاران داشتند. سرمایه‌گذاران به مدل کسب و کار مالی و آمادگی شرکت برای توسعه توجه زیادی کردند. |
| **8** | **(Weking et al., 2019)** | **آیا مدل کسب و کار بر موفقیت استارتاپ تأثیر دارد؟ تحلیل کمی** | **Contingency Analysis** | **500 استارتاپ از Crunchbase که در سال 2015 تأسیس شده‌اند، 250 استارتاپ موفق و 250 استارتاپ شکست خورده** | - | **مدل‌های کسب و کار (مانند Freemium، Subscription، Cross Selling و Hidden Revenue)** | مدل‌های Freemium و Subscription با افزایش احتمال بقا در ارتباط هستند؛ مدل‌های Cross Selling، Hidden Revenue، Layer Player و No Frills به احتمال شکست بیشتر استارتاپ‌ها منجر می‌شوند. این پژوهش اهمیت مدل کسب و کار برای موفقیت استارتاپ را تأیید می‌کند. |
| **9** | **(Żbikowski & Antosiuk, 2021)** | **پیش‌بینی موفقیت استارتاپ‌ها با استفاده از داده‌های Crunchbase به رویکرد بدون سوگیری** | **Logistic Regression, Support Vector Machine (SVM), Gradient Boosting Trees (XGBoost)** | **Crunchbase** | عرضه اولیه عمومی (IPO)  ادغام و خرید (M&A)  موفقیت در جذب سرمایه‌ی سری B | **کشور**  **منطقه**  **بخش صنعتی**  **جنسیت بنیان‌گذار**  **آیا بنیان‌گذار دارای مدرک تحصیلی است؟**  **تعداد سال‌های بین فارغ‌التحصیلی و تأسیس استارتاپ**  **تعداد سال‌های تحصیل بنیان‌گذار** | متغیرهای لوکیشن و صنعت بیشترین تاثیر را در تصمیم‌گیری مدل داشتند. استارتاپ‌هایی که در صنایع نرم‌افزار و خدمات اینترنتی فعالیت می‌کنند، موفقیت بیشتری داشتند. جنسیت و تحصیلات بنیان‌گذار نیز بر موفقیت استارتاپ تأثیرگذار بودند. |
| **10** | **(Ross et al., 2021)** | **پیش‌بینی موفقیت استارتاپ‌ها و پیش‌بینی خروج (IPO، خریداری شدن، یا شکست)** | **Logistic Regression, Random Forest, XGBoost, K-Nearest Neighbors** | **Crunchbase و USPTO** | موفقیت از طریق عرضه اولیه سهام (IPO)  ادغام و تملیک (Merger and Acquisition) | **میانگین زمان بین مراحل سرمایه‌گذاری (avg\_time\_between\_funding)**  **تعداد مراحل سرمایه‌گذاری (funding\_rounds)**  **تعداد سرمایه‌گذاران (num\_investors)**  **تعداد بنیان‌گذاران زن (num\_female\_founders)**  **تعداد بنیان‌گذاران مرد (num\_male\_founders)**  **تعداد پتنت‌ها (num\_patents)**  **نوع سرمایه‌گذاران (VC، Angel و غیره) (investor\_types)**  **کد کشور دفتر مرکزی (country\_code)**  **کد ایالت دفتر مرکزی (state\_code)**  **دسته‌بندی صنعت (industry category)**  **طول توضیحات شرکت (description\_length)**  **آیا دامنه وب‌سایت دارد؟ (has\_domain)**  **آیا آدرس ایمیل ثبت کرده است؟ (has\_email)**  **آیا حساب لینکدین دارد؟ (has\_linkedin)**  **آیا حساب فیسبوک دارد؟ (has\_facebook)**  **آیا حساب توییتر دارد؟ (has\_twitter)**  **تعداد اکتساب‌های انجام شده (num\_acquisitions)** | مدل یادگیری ماشین با استفاده از داده‌های عمومی، پیش‌بینی موفقیت استارتاپ‌ها را با دقت بالا انجام داده و به ویژه در مورد IPO، خریداری شدن، و شکست احتمالی دقیق بوده است. مدل توانسته بود نتایج خروج را با دقت 80 تا 90 درصد پیش‌بینی کند. |
| **11** | **(Dellermann et al., 2021)** | **یافتن یونیکورن: پیش‌بینی موفقیت استارتاپ‌های مرحله ابتدایی از طریق روش هوش ترکیبی** | **Logistic Regression (رگرسیون لجستیک)**  **Naive Bayes (نایو بیز)**  **Support Vector Machine (SVM) (ماشین بردار پشتیبان)**  **Artificial Neural Network (ANN) (شبکه عصبی مصنوعی)**  **Artificial Neural Network (ANN) (شبکه عصبی مصنوعی)** | **Crunchbase, Mattermark, and Dealroom** | دریافت سرمایه‌گذاری سری A | **تعداد مراحل سرمایه‌گذاری (Number of funding rounds)**  **میانگین زمان بین مراحل سرمایه‌گذاری (Average time between funding rounds)**  **تعداد بنیان‌گذاران زن (Number of female founders)**  **تعداد بنیان‌گذاران مرد (Number of male founders)**  **تعداد پتنت‌ها (Number of patents)**  **نوع سرمایه‌گذاران (Type of investors)**  **موقعیت دفتر مرکزی (Headquarters location - country and state)**  **دسته‌بندی صنعتی (Industry category)** | ترکیب هوش ماشینی و هوش جمعی (Hybrid Intelligence) برای پیش‌بینی موفقیت استارتاپ‌ها در شرایط عدم قطعیت بالا، نتایج دقیق‌تری ارائه می‌دهد. این روش توانست با استفاده از ویژگی‌های "سخت" و "نرم" استارتاپ‌ها، احتمال دریافت سرمایه‌گذاری سری A را با دقت بیشتری پیش‌بینی کند |
| **12** | **(Krishna et al., 2016)** | **پیش‌بینی نتیجه استارتاپ‌ها: شکست کمتر، موفقیت بیشتر** | **Naive Bayes (بیز ساده)**  **ADTrees (درختان تصمیم مبتنی بر AND/OR)**  **Bayesian Networks (شبکه‌های بیزی)**  **Lazy lb1 (روش نزدیک‌ترین همسایه)**  **Random Forest (جنگل تصادفی)**  **Simple Logistic (رگرسیون لجستیک ساده)** | **Crunchbase و TechCrunch** | موفقیت برای جذب سرمایه و یا تعطیلی استارتاپ | **میزان سرمایه اولیه (Seed funding amount)**  **زمان دریافت سرمایه اولیه (Time to get seed funding)**  **زمان دریافت سرمایه اولیه (Time to get seed funding)**  **میزان Burn rate**  **شاخص‌های شدت (Severity scores)**  **ارزش‌گذاری پس از هر مرحله سرمایه‌گذاری (Valuation after each round)**  **مدت زمان فعالیت شرکت (Months active)**  **کل سرمایه جذب شده (Total funds raised)** | ترکیب مدل‌های مختلف یادگیری ماشین به دقت بالا در پیش‌بینی موفقیت و شکست استارتاپ‌ها منجر شد. مدل‌ها توانستند نتایج پیش‌بینی با دقت بیش از 90% را بر اساس عوامل مهم مانند سرمایه‌گذاری و میزان Burn rate ارائه کنند​. |

* 1. شکاف‌های تحقیقات پیشین

با وجود تحقیقات گسترده‌ای که در زمینه پیش‌بینی موفقیت استارت‌آپ‌ها انجام شده است، هنوز برخی شکاف‌های مهم در این حوزه وجود دارد که نیاز به بررسی و مطالعه بیشتر دارند. پژوهش کیم و همکاران (Kim et al., 2023) به بررسی موفقیت استارتاپ‌ها با استفاده از داده‌های وبسایت کرانچ بیس و مدل‌های یادگیری ماشین پرداخته است. با وجود اهمیت این مطالعه در ارائه مدل‌های پیش‌بینی موفقیت استارتاپ‌ها، چندین شکاف تحقیقاتی مهم در این مقاله وجود دارد. اولاً، این مقاله بر ویژگی‌های عمومی استارتاپ‌ها تمرکز کرده و به طور خاص به تحلیل صنایع مختلف نپرداخته است. در حالی که ویژگی‌های صنعت می‌تواند تأثیر زیادی بر موفقیت استارتاپ‌ها داشته باشد، مقاله به بررسی جزئیات صنایع خاص نپرداخته و داده‌های خاص هر صنعت را در نظر نگرفته است. این کمبود می‌تواند به کاهش دقت مدل‌های پیش‌بینی منجر شود.

ثانیاً، مقاله به نحوه جمع‌آوری، تمیز کردن و تحلیل داده‌های کرانچ بیس به طور دقیق نپرداخته است. استفاده از داده‌های کرانچ بیس بدون توجه کافی به کیفیت و تمیزی داده‌ها می‌تواند به نتایج نادرست منجر شود. همچنین، مقاله به تحلیل زمان‌بندی و روندهای بلندمدت در صنعت‌های مختلف نپرداخته و بیشتر بر تحلیل کوتاه‌مدت متمرکز شده است. عدم توجه به تغییرات زمانی و روندهای بلندمدت می‌تواند به نادیده گرفتن عوامل مهمی منجر شود که بر موفقیت استارتاپ‌ها تأثیر می‌گذارند. این محدودیت‌ها می‌تواند به کاهش دقت و کاربردپذیری مدل‌های پیش‌بینی منجر شود.

ژبیکوفسکی و همکاران (Żbikowski & Antosiuk, 2021) به بررسی موفقیت کسب‌وکارها با استفاده از رویکردهای یادگیری ماشین بدون سوگیری پرداخته است. این مقاله با استفاده از داده‌های کرانچ بیس به تحلیل موفقیت استارتاپ‌ها می‌پردازد، اما چندین شکاف تحقیقاتی مهم در این مقاله مشاهده می‌شود. اولاً، این مقاله به‌طور کلی به بررسی کسب‌وکارها پرداخته و به ویژگی‌های خاص صنایع مختلف توجه نکرده است. در حالی که ویژگی‌های صنعت می‌تواند تأثیر بسزایی بر موفقیت کسب‌وکارها داشته باشد، این مقاله به تحلیل جزئیات صنایع مختلف نپرداخته است. این نقص می‌تواند به کاهش دقت مدل‌های پیش‌بینی منجر شود.

ثانیاً، مقاله به استفاده از داده‌های متنوع و ترکیبی در تحلیل‌ها به‌طور کامل نپرداخته است. استفاده از داده‌های متنوع و ترکیبی می‌تواند به بهبود دقت و کاربردپذیری مدل‌های پیش‌بینی کمک کند، اما مقاله به‌طور کافی به این موضوع نپرداخته است. این شکاف‌ها می‌تواند به کاهش دقت و کاربردپذیری مدل‌های پیش‌بینی منجر شود (Żbikowski & Antosiuk, 2021).

آنا دیاس و همکاران (Deias & Magrini, 2023)، به بررسی تأثیر دینامیک‌های تأمین مالی سهام بر موفقیت استارتاپ‌ها می‌پردازد. مارک پتانین و همکاران (Potanin et al., 2023) به ارزیابی عملکرد مدل‌های یادگیری ماشین برای پشتیبانی از تصمیم‌گیری در سرمایه‌گذاری‌های مخاطره‌آمیز می‌پردازد. این مطالعات با تحلیل داده‌های کرانچ بیس و بررسی عوامل مالی مرتبط با تأمین مالی سهام، به پیش‌بینی موفقیت استارتاپ‌ها پرداخته است. با این حال، چندین شکاف تحقیقاتی مهم در این مقاله وجود دارد. اولاً، مقاله بیشتر بر تحلیل عوامل مالی داخلی تمرکز کرده و به تحلیل جزئیات صنعتی و تأثیر ویژگی‌های خاص صنایع مختلف نپرداخته است. این محدودیت می‌تواند به نادیده گرفتن عوامل مهمی منجر شود که بر موفقیت استارتاپ‌ها در صنایع مختلف تأثیر می‌گذارند.

ثانیاً، مقاله به تحلیل زمان‌بندی و روندهای بلندمدت در تأمین مالی سهام توجه کافی نداشته است. در حالی که تغییرات زمانی و روندهای بلندمدت می‌تواند تأثیر بسزایی بر موفقیت استارتاپ‌ها داشته باشد، این مقاله بیشتر بر تحلیل کوتاه‌مدت و عوامل مالی داخلی تمرکز کرده است. این کمبود می‌تواند به کاهش دقت مدل‌های پیش‌بینی منجر شود. علاوه بر این، مقاله به استفاده از تکنیک‌های پیشرفته یادگیری ماشین توجه کافی نداشته و بیشتر به تکنیک‌های پایه‌ای پرداخته است. استفاده از تکنیک‌های پیشرفته می‌تواند به بهبود دقت پیش‌بینی‌ها کمک کند، اما این مقاله به‌طور کافی به این موضوع نپرداخته است.

بنابراین میتوان گفت یکی از اصلی‌ترین شکاف‌های تحقیقات پیشین، تمرکز بیشتر تحقیقات بر روی شرایط داخلی استارت‌آپ‌ها و نادیده گرفتن ویژگی‌ها و تأثیرات صنعت می‌باشد. بیشتر مطالعات گذشته بر روی ویژگی‌های داخلی مانند تیم مدیریتی، منابع مالی، و مدل‌های کسب‌وکار تمرکز داشته‌اند و به تأثیرات محیط بیرونی و صنعت توجه کمتری داشته‌اند.

همچنین، در تحقیقات پیشین حوزه حمل و نقل معمولاً از روش‌های داده‌کاوی ساده‌تر استفاده شده است و به کاربرد الگوریتم‌های پیشرفته‌تر و تحلیل‌های پیچیده‌تر مانند شبکه های عصبی کمتر توجه شده است. این موضوع باعث شده که پیش‌بینی‌های انجام شده دقت و کارایی کمتری داشته باشند.

در نهایت، بیشتر مطالعات بر روی داده‌های محلی یا محدود به یک منطقه خاص انجام شده‌اند و کمتر به تحلیل داده‌های بزرگ و جامع از منابع جهانی مانند کرانچ بیس پرداخته‌اند. استفاده از داده‌های جهانی و تحلیل جامع‌تر می‌تواند به بهبود دقت و کاربرد مدل‌های پیش‌بینی کمک کند.

این شکاف‌ها نشان‌دهنده نیاز به تحقیقات دقیق‌تر در زمینه پیش‌بینی موفقیت استارت‌آپ‌ها با در نظر گرفتن ویژگی‌های صنعت و استفاده از الگوریتم‌های پیشرفته داده‌کاوی می‌باشد. در این پژوهش سعی شده است با تمرکز بر روی صنعت حمل و نقل و استفاده از داده‌های خاص وب‌سایت کرانچ بیس، به طور دقیق‌تر و کاربردی‌تر به پیش‌بینی موفقیت استارتاپ‌ها در این صنعت بپردازد.

با در نظر گرفتن صنعت حمل ونقل، تغییرات زمانی و تحلیل روندهای بلندمدت، استفاده از تکنیک‌های پیشرفته‌تر ماشین لرنینگ، و در نظر گرفتن متغیرهای خارجی، مدل پیش‌بینی این پژوهش می‌تواند نتایج دقیق‌تری ارائه دهد و به تصمیم‌گیری بهتر برای سرمایه‌گذاران و کارآفرینان کمک کند. این باعث می‌شود که این پژوهش بتواند شکاف‌های موجود در تحقیقات پیشین را پر کند و به طور ملموسی در پیش‌بینی موفقیت استارتاپ‌ها در صنعت حمل و نقل موثر باشد. با توجه به شکاف‌های موجود در مقالات بررسی شده، این پژوهش می‌تواند با تمرکز بر ویژگی‌های خاص صنعت حمل و نقل و تحلیل دقیق داده‌های مربوط به این صنعت، به پر کردن این شکاف‌ها کمک کرده و نتایج بهتری ارائه دهد

* 1. جمع بندی

در این فصل، به بررسی پیشینه تحقیق و ادبیات موجود در زمینه پیش‌بینی موفقیت استارت‌آپ‌های هوش مصنوعی در صنعت حمل و نقل با استفاده از هوش مصنوعی و یادگیری ماشین پرداخته شد. پیشرفت‌های فناوری در حوزه‌های هوش مصنوعی و یادگیری ماشین، فرصت‌های بی‌سابقه‌ای برای ارتقای عملکرد صنایع و مشاغل فراهم کرده است. یکی از صنایعی که از این پیشرفت‌ها بهره‌برداری کرده، صنعت حمل‌ونقل است. استفاده از هوش مصنوعی در این صنعت، بهبود کارایی، کاهش هزینه‌ها، افزایش ایمنی و بهبود تجربه مشتریان را ممکن کرده است.

تحقیقات متعددی در زمینه کاربردهای هوش مصنوعی در صنعت حمل‌ونقل انجام شده است. این تحقیقات نشان می‌دهند که استفاده از هوش مصنوعی می‌تواند به بهبود سیستم‌های مدیریت ترافیک، توسعه خودروهای خودران، پیش‌بینی تراکم ترافیک و بهینه‌سازی مسیرها کمک کند. این تکنولوژی‌ها می‌توانند به کاهش آلودگی محیط زیست، افزایش کارایی انرژی و کاهش زمان سفر کمک کنند. همچنین، هوش مصنوعی می‌تواند به افزایش ایمنی در حمل‌ونقل کمک کند، به طوری که با تحلیل داده‌های جمع‌آوری شده از حسگرها و سیستم‌های مختلف، خطرات احتمالی را پیش‌بینی کرده و از وقوع حوادث جلوگیری کند. بنابراین ارائه مدلی برای پیش بینی موفقیت استارتاپ‌های این حوزه میتواند چراغ راهی برای سرمایه گذاران این حوزه باشد.

با این حال، پیش‌بینی موفقیت استارت‌آپ‌ها در صنعت حمل‌ونقل همچنان یک چالش بزرگ است. بررسی‌های انجام شده نشان می‌دهند که بیشتر مطالعات گذشته بر روی ویژگی‌های داخلی استارت‌آپ‌ها مانند تیم مدیریتی، منابع مالی و مدل‌های کسب‌وکار تمرکز داشته‌اند و به تأثیرات محیط بیرونی و ویژگی‌های خاص صنایع مختلف توجه کمتری داشته‌اند. این موضوع نشان‌دهنده یکی از شکاف‌های اصلی در تحقیقات پیشین است.

علاوه بر این، بیشتر مطالعات از روش‌های ساده‌تر داده‌کاوی استفاده کرده و به کاربرد الگوریتم‌های پیشرفته‌تر مانند شبکه‌های عصبی توجه کمتری داشته‌اند. این موضوع باعث شده که دقت و کارایی پیش‌بینی‌ها کاهش یابد. همچنین، بیشتر مطالعات بر روی داده‌های محلی یا محدود به یک منطقه خاص انجام شده‌اند و کمتر به تحلیل داده‌های بزرگ و جامع از منابع جهانی مانند کرانچ بیس پرداخته‌اند. استفاده از داده‌های جهانی و تحلیل جامع‌تر می‌تواند به بهبود دقت و کاربرد مدل‌های پیش‌بینی کمک کند.

با توجه به این شکاف‌ها، پژوهش حاضر با تمرکز بر روی صنعت حمل‌ونقل و استفاده از داده‌های جامع وب‌سایت کرانچ بیس، به دنبال ارائه مدل‌های دقیق‌تری برای پیش‌بینی موفقیت استارت‌آپ‌ها است. استفاده از تکنیک‌های پیشرفته‌تر یادگیری ماشین و تحلیل‌های دقیق‌تر داده‌ها می‌تواند به بهبود دقت پیش‌بینی‌ها و کمک به تصمیم‌گیری بهتر برای سرمایه‌گذاران و کارآفرینان منجر شود. این پژوهش با پر کردن شکاف‌های موجود در تحقیقات پیشین، می‌تواند به طور ملموسی در پیش‌بینی موفقیت استارت‌آپ‌ها در صنعت حمل‌ونقل موثر باشد و به عنوان یک مرجع ارزشمند برای مطالعات آینده در این حوزه مورد استفاده قرار گیرد.

## 

# روش تحقیق

## مقدمه

در دهه‌های اخیر، پیشرفت‌های قابل‌توجهی در زمینه فناوری‌های هوش مصنوعی صورت گرفته که توانسته‌اند بسیاری از صنایع، از جمله صنعت حمل و نقل را متحول کنند. با افزایش تقاضا برای راه‌حل‌های نوآورانه و هوشمند در این حوزه، شرکت‌های بسیاری در تلاش برای ارائه خدمات و محصولات جدید هستند که بتوانند کارایی و ایمنی را افزایش داده و هزینه‌ها را کاهش دهند (Abduljabbar et al., 2019). با این وجود، بازار رقابتی و چالش‌های مالی متعدد، پیش‌بینی موفقیت این شرکت‌ها را به یک موضوع مهم و پیچیده تبدیل کرده است.

هدف از این تحقیق، توسعه یک مدل پیش‌بینی است که با استفاده از تکنیک‌های پیشرفته ماشین لرنینگ، قادر به پیش‌بینی موفقیت شرکت‌های فعال در زمینه هوش مصنوعی در صنعت حمل و نقل باشد. این مدل با تحلیل داده‌های تاریخی و جاری از شرکت‌ها، می‌تواند به تصمیم‌گیری بهتر برای سرمایه‌گذاران و مدیران کسب‌وکار کمک کند (Bento, 2018).

پایگاه داده کرانچ‌بیس به‌عنوان یکی از جامع‌ترین منابع اطلاعاتی برای شرکت‌های فناوری، داده‌های گسترده‌ای از بیش از 4000 شرکت فعال در صنعت حمل و نقل فراهم می‌کند. این داده‌ها شامل اطلاعاتی همچون تعداد کارکنان، مراحل مختلف جذب سرمایه، مکان دفتر مرکزی، و اطلاعاتی درباره بنیان‌گذاران و سرمایه‌گذاران است (Zhang et al., 2021). استفاده از این داده‌ها به ما این امکان را می‌دهد که عوامل مؤثر بر موفقیت یا شکست شرکت‌ها را با دقت بیشتری شناسایی کنیم.

برای مدل‌سازی پیش‌بینی موفقیت، از روش‌های مختلف یادگیری ماشین استفاده می‌شود. این روش‌ها شامل الگوریتم‌های شناخته‌شده‌ای مانند رگرسیون لجستیک، جنگل تصادفی و ماشین بردار پشتیبان است. هر یک از این الگوریتم‌ها به‌طور خاص برای تحلیل داده‌های بزرگ و پیچیده طراحی شده‌اند و توانایی آن‌ها در پیش‌بینی نتایج کسب‌وکارها در تحقیقات پیشین اثبات شده است (Kim et al., 2023).

یکی از چالش‌های اساسی در تحلیل داده‌های کسب‌وکار، عدم تعادل در دسته‌بندی داده‌ها و وجود داده‌های گمشده است. برای مقابله با این چالش‌ها، از تکنیک‌های مختلف پیش‌پردازش داده‌ها مانند نرمال‌سازی، تبدیل دسته‌ای و پر کردن داده‌های گمشده با استفاده از الگوریتم‌های داده کاوی، استفاده خواهد شد.

در ادامه در این فصل به‌طور جامع به بررسی و توضیح روش‌های انتخاب‌شده برای تحقیق، تحلیل داده‌ها و مدل‌سازی پرداخته می‌شود و چارچوب کلی پژوهش را برای دستیابی به اهداف تعیین‌شده مشخص خواهد شد.

## روش تحقیق

### علت انتخاب روش

در این پژوهش، استفاده از چارچوب CRISP-DM[[52]](#footnote-52) به‌عنوان روش اصلی تحقیق انتخاب شده است. این چارچوب به‌عنوان یکی از استانداردهای جامع و پر استفاده در حوزه داده‌کاوی و پیش‌بینی در صنایع مختلف شناخته می‌شود و به دلیل ساختار گام‌به‌گام و رویکرد سیستماتیک آن، محققان را قادر می‌سازد تا داده‌ها را به‌طور مؤثری تحلیل کرده و نتایج دقیق و قابل اعتمادی به دست آورند. این چارچوب با تمرکز بر فرآیندهای داده‌محور و ایجاد ارتباط منطقی بین مراحل مختلف تحلیل، به‌ویژه در زمینه‌های صنعتی مانند پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل، بسیار کارآمد است.

یکی از دلایل اصلی انتخاب این روش، ساختار مرحله‌ای و جامع آن است که تمامی مراحل یک پروژه داده‌کاوی را از شناسایی مسئله کسب‌وکار تا پیاده‌سازی نهایی مدل‌ها پوشش می‌دهد. CRISP-DM به‌ویژه به دلیل توانایی آن در مدیریت داده‌های پیچیده و بزرگ، در این پژوهش به‌کار گرفته شده است. داده‌های مرتبط با شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل دارای ویژگی‌های متعدد و متنوعی هستند، از جمله تعداد کارکنان، مراحل مختلف جذب سرمایه، موقعیت جغرافیایی دفتر مرکزی و اطلاعات مربوط به بنیان‌گذاران و سرمایه‌گذاران. این داده‌ها می‌توانند پیچیدگی‌های زیادی در تحلیل ایجاد کنند و استفاده از CRISP-DM با توجه به ساختار مرحله‌ای آن، این پیچیدگی‌ها را به‌خوبی مدیریت می‌کند (Wirth & Hipp, 2000).

دلیل دیگر برای انتخاب این روش در این پژوهش، انعطاف‌پذیری بالای این چارچوب است. داده‌های مورد استفاده در این تحقیق از پایگاه داده کرانچ‌بیس استخراج شده‌اند که یکی از جامع‌ترین منابع اطلاعاتی برای شرکت‌های فناوری است. داده‌های استخراج شده شامل اطلاعات مرتبط با سرمایه‌گذاری‌ها، موقعیت جغرافیایی، و ویژگی‌های دیگر کسب‌وکارهای هوش مصنوعی است. CRISP-DM این امکان را فراهم می‌کند که داده‌های پیچیده و متنوع به‌صورت جامع و دقیق تحلیل شوند و الگوهای موفقیت شرکت‌ها شناسایی شوند.

سومین دلیل انتخاب این چارچوب، توانایی CRISP-DM در تطبیق با روش‌های مختلف یادگیری ماشین است. در این پژوهش، از 6 الگوریتم مختلف داده‌کاوی برای پیش‌بینی موفقیت شرکت‌ها استفاده شده است. هر یک از این الگوریتم‌ها دارای ویژگی‌های خاصی هستند که برای تحلیل داده‌های پیچیده مناسب هستند. برای مثال، الگوریتم جنگل تصادفی به دلیل توانایی‌اش در مدیریت داده‌های نامتوازن و نویزی، یکی از الگوریتم‌های اصلی در این تحقیق است. CRISP-DM به‌صورت طبیعی این امکان را فراهم می‌کند که از این الگوریتم‌ها به‌شکل مؤثری استفاده شود و به‌همین دلیل، این چارچوب به‌عنوان یکی از روش‌های کلیدی تحقیق انتخاب شده است (Bento et al., 2018؛ Lisin et al., 2017؛ Corea et al., 2019).

علاوه بر این، یکی از چالش‌های اساسی در تحلیل داده‌های مربوط به کسب‌وکارهای استارت‌آپی، وجود داده‌های گمشده و نامتوازن است. CRISP-DM با ارائه ساختار پیش‌پردازش داده‌ها، توانایی مقابله با این چالش‌ها را فراهم می‌کند. در این تحقیق، برای کاهش اثرات منفی داده‌های گمشده و نامتوازن، از تکنیک‌های نرمال‌سازی و پر کردن داده‌های گمشده استفاده شده است. این تکنیک‌ها به بهبود کیفیت داده‌ها و افزایش دقت پیش‌بینی کمک می‌کنند (Liu et al., 2024).

در نهایت، انتخاب این روش‌ها بر اساس موفقیت آن‌ها در مطالعات مشابه بوده است. تحقیقات پیشین نشان داده‌اند که استفاده از الگوریتم‌های یادگیری ماشین می‌تواند دقت بالایی در پیش‌بینی موفقیت شرکت‌ها ارائه دهد. با توجه به این نتایج مثبت و قابل‌اعتماد در پروژه‌های مشابه، CRISP-DM به‌عنوان چارچوب اصلی تحقیق انتخاب شده است تا بتواند به‌طور مؤثر به تحلیل و پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌و‌نقل بپردازد.

### تشريح كامل روش تحقيق

این پژوهش با هدف پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل طراحی و اجرا شده است. به منظور دستیابی به این هدف، از رویکرد CRISP-DM استفاده شده است. رویکرد CRISP-DM به‌عنوان یک استاندارد مرجع در داده‌کاوی، شامل شش مرحله اصلی است که به ترتیب از درک نیازهای کسب‌وکار تا پیاده‌سازی نتایج را دربرمی‌گیرد (شکل3-1).

**شکل 3-1**



شکل 3-8 فرآیند اجرای تحقیق

#### درک کسب‌وکار

در این مرحله هدف اصلی پژوهش، شناسایی و پیش‌بینی عوامل کلیدی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل است. این صنعت به دلیل ترکیب پیچیدگی‌های فناوری با شرایط خاص بازار حمل‌ونقل، به تحلیل دقیق نیاز دارد تا بتوان تصمیمات بهتری برای سرمایه‌گذاری و مدیریت شرکت‌ها اتخاذ کند. یکی از چالش‌های اصلی پیش ‌روی مدیران و سرمایه‌گذاران در این حوزه، ریسک بالای شکست یا موفقیت شرکت‌های فعال در این صنعت است، به‌خصوص که هوش مصنوعی به‌عنوان یک فناوری نوظهور با عدم قطعیت‌های زیادی مواجه است.

در این مرحله، هدف پژوهشگران شناسایی این عوامل و بررسی تأثیر آن‌ها بر موفقیت یا شکست شرکت‌ها است. با استفاده از داده‌های مربوط به شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل، متغیرهایی همچون تعداد کارکنان، مراحل مختلف جذب سرمایه، موقعیت جغرافیایی دفتر مرکزی، اطلاعات مربوط به بنیان‌گذاران و سرمایه‌گذاران، و میزان مواجهه شرکت با رسانه‌ها و مشتریان به‌عنوان عوامل تأثیرگذار در موفقیت شرکت‌ها در نظر گرفته شده‌اند. این متغیرها به‌ویژه با استفاده از تکنیک‌های یادگیری ماشین مورد تحلیل قرار می‌گیرند تا الگوهای نهان میان داده‌ها کشف و از آن‌ها برای پیش‌بینی موفقیت یا شکست شرکت‌ها استفاده شود (Kim et al., 2023).

به‌منظور اطمینان از درک صحیح نیازهای کسب‌وکار و استفاده از روش‌های مناسب پیش‌بینی، این مرحله به بررسی جامع صنعت حمل‌ونقل و جایگاه هوش مصنوعی در آن پرداخته و ریسک‌ها و فرصت‌های این صنعت را شناسایی کرده است. این بررسی‌ها به محققان کمک می‌کند تا معیارهای کلیدی موفقیت را به‌طور دقیق تعریف کرده و الگوریتم‌های مناسب برای پیش‌بینی را انتخاب کنند.

#### درک داده

در این مرحله، داده‌های مربوط به شرکت‌های فعال در حوزه هوش مصنوعی و حمل‌ونقل از پایگاه داده کرانچ بیس جمع‌آوری شده است. کرانچ بیس یکی از بزرگ‌ترین و جامع‌ترین منابع اطلاعاتی برای استارتاپ‌ها و شرکت‌های فناوری در جهان است و داده‌های آن شامل اطلاعات جامعی درباره مراحل مختلف جذب سرمایه، تعداد دورهای سرمایه‌گذاری، تعداد کارکنان، موقعیت جغرافیایی، بنیان‌گذاران و سرمایه‌گذاران، و اطلاعات مربوط به محصولات و فناوری‌های استفاده‌شده توسط شرکت‌ها می‌باشد (Kim et al., 2023). این اطلاعات به‌طور سیستماتیک از کرانچ بیس استخراج و در قالب‌های ساختاریافته برای تحلیل‌های بعدی ذخیره شده‌اند.

در این مرحله، توجه ویژه‌ای به تحلیل ویژگی‌های داده‌ها و متغیرهای موجود شده است.

داده‌های جمع‌آوری‌شده شامل چندین دسته از ویژگی‌ها هستند که هر کدام به‌طور مستقیم یا غیرمستقیم بر موفقیت شرکت‌ها تأثیرگذارند (جدول ). در این پژوهش از 42 ویژگی برای مدلسازی پیش‌بینی موفقیت شرکت‌ها استفاده شده است.

|  |  |  |  |
| --- | --- | --- | --- |
| * 1-Color | * 2-Director\_name | * **3-Num\_critic\_for\_reviews** | * **4-Duration** |
| * **5-Director\_facebook\_likes** | * **6-Actor\_3\_facebook\_likes** | * 7-Actor\_2\_name | * **8-Actor\_1\_facebook\_likes** |
| * **9-Gross** | * **10-Genres** | * 11-Actor\_1\_name | * 12-Movie\_title |
| * **13-Num\_voted\_users** | * **14-Cast\_total\_facebook\_likes** | * 15-Actor\_3\_name | * **16-Facenumber\_in\_poster** |
| * 17-Plot\_keywords | * 18-Movie\_imdb\_link | * **19-Num\_user\_for\_reviews** | * 20-Language |
| * 21-Country | * 22-Content\_rating | * **23-Budget** | * **24-Title\_year** |
| * **25-Actor\_2\_facebook\_likes** | * **26-Imdb\_score** | * **27-Aspect ratio** | * **28- Movie Facebook like** |

**جدول 3- 1- ویژگی‌های انتخاب شده برای مدلسازی پیش بینی موفقیت**

|  |  |  |  |
| --- | --- | --- | --- |
|  | گروه | ویژگی | دلیل انتخاب |
| 1 | ویژگی‌های مرتبط با شرکت | Number of Employees (تعداد کارکنان) | تعداد کارکنان می‌تواند نشان‌دهنده اندازه و ظرفیت رشد شرکت باشد. شرکت‌های با تعداد کارکنان بیشتر معمولاً توانایی اجرای پروژه‌های بزرگتر و مدیریت عملیات پیچیده‌تر را دارند. |
| 2 | Headquarters Location (موقعیت جغرافیایی دفتر مرکزی) | موقعیت جغرافیایی می‌تواند بر دسترسی به بازارها، منابع انسانی و شبکه‌های صنعتی تأثیر بگذارد. شرکت‌های مستقر در مراکز صنعتی مهم ممکن است دسترسی بهتری به سرمایه‌گذاران و مشتریان داشته باشند. |
| 3 | Headquarters Regions (مناطق دفتر مرکزی) | مناطق جغرافیایی می‌توانند نشان‌دهنده محیط‌های اقتصادی و قانونی مختلف باشند که بر موفقیت شرکت تأثیر می‌گذارند. برخی مناطق ممکن است مشوق‌های مالیاتی یا دسترسی به زیرساخت‌های بهتری داشته باشند. |
| 4 | Founded Date (تاریخ تأسیس) | تاریخ تأسیس شرکت نشان‌دهنده میزان تجربه و سابقه آن در بازار است. شرکت‌هایی که سابقه طولانی‌تری دارند، معمولاً تجربه بیشتری در مواجهه با 5چالش‌های بازار دارند و احتمالاً پایدارتر هستند. |
| 5 | CB Rank (Company) (رتبه شرکت در Crunchbase) | رتبه در کرانچ می‌تواند معیاری از عملکرد عمومی و شناخته‌شدن شرکت باشد. شرکت‌هایی با رتبه بهتر معمولاً به دلیل جذب سرمایه، تعداد کارکنان و دستاوردهایشان موفق‌تر دیده می‌شوند. |
| 6 | Estimated Revenue Range (محدوده درآمد تخمینی) | این ویژگی نشان‌دهنده توانایی مالی شرکت است. شرکت‌هایی که درآمد بیشتری دارند، احتمالاً از نظر مالی پایدارتر هستند و می‌توانند بهتر در رقابت‌های بازار موفق شوند. |
| 7 | Operating Status (وضعیت عملیاتی) | وضعیت عملیاتی شرکت (فعال، غیرفعال یا فروخته شده) یکی از عوامل کلیدی در ارزیابی موفقیت شرکت‌ها است. شرکت‌های فعال که هنوز در حال فعالیت هستند، معمولاً به‌عنوان شرکت‌های موفق‌تر در نظر گرفته می‌شوند. |
| 8 | Number of Founders (تعداد بنیان‌گذاران) | تعداد بنیان‌گذاران می‌تواند تأثیر زیادی بر پایداری و نوآوری شرکت داشته باشد. تیم بنیان‌گذاران قوی و متنوع می‌تواند منابع و دیدگاه‌های بیشتری برای موفقیت شرکت فراهم کند. |
| 9 | Similar Companies (شرکت‌های مشابه) | مقایسه شرکت با سایر شرکت‌های مشابه در همان صنعت می‌تواند به ارزیابی موقعیت شرکت در بازار کمک کند. شرکت‌هایی که در میان شرکت‌های مشابه در وضعیت خوبی قرار دارند، معمولاً موفقیت بیشتری دارند. |
| 10 | Number of Articles |  |
| 11 | ویژگی‌های مرتبط با جذب سرمایه | Number of Funding Rounds (تعداد دورهای جذب سرمایه) | تعداد دورهای جذب سرمایه نشان‌دهنده علاقه سرمایه‌گذاران و توانایی شرکت در جذب منابع مالی است. شرکت‌هایی که چندین دور سرمایه‌گذاری را به موفقیت پشت سر گذاشته‌اند، به‌طور معمول به‌عنوان شرکت‌های موفق شناخته می‌شوند. |
| 12 | Last Funding Type (نوع آخرین دور جذب سرمایه) | نوع آخرین دور سرمایه‌گذاری می‌تواند نشان‌دهنده مرحله توسعه شرکت باشد (مانند Series A، B، یا IPO). این ویژگی می‌تواند به پیش‌بینی وضعیت شرکت و سطح رشد آن کمک کند. |
| 13 | Last Funding Amount (in USD) (مقدار آخرین سرمایه‌گذاری به دلار آمریکا) | مقدار آخرین سرمایه‌گذاری به‌طور مستقیم بر رشد و توسعه شرکت تأثیر می‌گذارد. شرکت‌هایی که سرمایه‌گذاری‌های بزرگتری جذب می‌کنند، معمولاً در مراحل رشد بالاتر قرار دارند. |
| 14 | Total Funding Amount (in USD) (کل میزان سرمایه‌گذاری به دلار آمریکا) | کل سرمایه‌گذاری‌های جذب‌شده نشان‌دهنده میزان اعتماد سرمایه‌گذاران و قدرت مالی شرکت است. این ویژگی می‌تواند به‌عنوان یکی از شاخص‌های موفقیت شرکت استفاده شود. |
| 15 | Last Equity Funding Type | نوع آخرین سرمایه‌گذاری سهام، مرحله رشد و استراتژی‌های توسعه شرکت را نشان می‌دهد. این ویژگی نشان می‌دهد که شرکت در چه مرحله‌ای از جذب سرمایه است و آیا برنامه‌های توسعه بیشتری دارد یا خیر. |
| 16 | Last Equity Funding Amount (in USD) (مقدار آخرین سرمایه‌گذاری سهام) | مقدار آخرین سرمایه‌گذاری سهام نشان‌دهنده توانایی شرکت در جذب سرمایه از بازارهای سهامی است که به استحکام مالی شرکت اشاره دارد. |
| 17 | Total Equity Funding Amount (in USD) (کل سرمایه‌گذاری سهام) | کل میزان سرمایه‌گذاری سهامی جذب‌شده بیانگر موفقیت شرکت در جذب سرمایه‌های خارجی و افزایش ارزش آن است. |
| 18 | Funding Status (وضعیت جذب سرمایه) | وضعیت سرمایه‌گذاری شرکت نشان می‌دهد که آیا شرکت همچنان در حال جذب سرمایه است یا نه. شرکت‌هایی که به‌طور مداوم سرمایه جذب می‌کنند، معمولاً پتانسیل رشد بیشتری دارند. |
| 19 | Number of Exits (IPO) (تعداد خروج‌ها به شکل IPO) | تعداد خروج‌های IPO نشان‌دهنده توانایی شرکت در جذب سرمایه از بازار عمومی و افزایش ارزش آن است. |
| 20 | Acquisition Type (نوع اکتساب) | نوع اکتساب نشان‌دهنده چگونگی خرید یا ادغام شرکت با دیگر شرکت‌ها است. این ویژگی می‌تواند میزان جذابیت و ارزش شرکت برای خریداران و سرمایه‌گذاران را نشان دهد. شرکت‌هایی که به‌صورت موفق در ادغام یا اکتساب شرکت‌های دیگر مشارکت می‌کنند، معمولاً در بازار موفق‌تر هستند. |
| 21 | Hub Tags (If Unicorn) | این ویژگی مشخص می‌کند که آیا شرکت به‌عنوان یک "یونیکورن" (شرکت خصوصی با ارزش بالای 1 میلیارد دلار) شناخته می‌شود یا خیر. شرکت‌هایی که برچسب یونیکورن دارند، معمولاً شرکت‌هایی موفق و ارزشمند هستند که پتانسیل بالایی برای رشد دارند. |
| 22 | ویژگی‌های مرتبط با سرمایه‌گذاران | Number of Lead Investors (تعداد سرمایه‌گذاران اصلی) | تعداد سرمایه‌گذاران اصلی نشان‌دهنده اعتماد سرمایه‌گذاران بزرگ به شرکت است و معمولاً سرمایه‌گذاران معتبر، نشانه موفقیت و پتانسیل رشد شرکت هستند. |
| 23 | Number of Investors (تعداد سرمایه‌گذاران) | تعداد سرمایه‌گذاران می‌تواند نشان‌دهنده جذابیت شرکت برای سرمایه‌گذاران باشد. هرچه تعداد سرمایه‌گذاران بیشتر باشد، اعتماد به شرکت و پتانسیل رشد آن بیشتر است. |
| 24 | Top 5 Investors (پنج سرمایه‌گذار برتر) | حضور سرمایه‌گذاران بزرگ و معتبر می‌تواند به‌عنوان نشانه‌ای از اعتماد سرمایه‌گذاران به شرکت در نظر گرفته شود. |
| 25 | Investor Type (نوع سرمایه‌گذار) | نوع سرمایه‌گذار (مانند سرمایه‌گذاران خطرپذیر، صندوق‌های سرمایه‌گذاری) نشان‌دهنده سطح خطرپذیری و اهداف سرمایه‌گذاری شرکت است. |
| 26 | ویژگی‌های مرتبط با فناوری | Patents Granted (تعداد پتنت‌های ثبت‌شده) | تعداد پتنت‌های ثبت‌شده نشان‌دهنده سطح نوآوری و دارایی‌های معنوی شرکت است و به‌طور مستقیم بر ارزش آن در بازار تأثیر می‌گذارد. |
| 27 | Trademarks Registered (تعداد علائم تجاری ثبت‌شده) | ثبت علائم تجاری نشان‌دهنده توانایی شرکت در محافظت از برند و محصولات خود است که به موفقیت آن در بازار کمک می‌کند. |
| 28 | Total Products Active (تعداد محصولات فعال) | تعداد محصولات فعال نشان‌دهنده دامنه محصولات شرکت و توانایی آن در ارائه راهکارهای مختلف به مشتریان است. |
| 29 | IT Spend (هزینه‌های فناوری اطلاعات) | هزینه‌های فناوری اطلاعات نشان‌دهنده سطح سرمایه‌گذاری شرکت در فناوری است که می‌تواند بر توانایی نوآوری و بهبود عملکرد عملیاتی تأثیر بگذارد. |
| 30 | Number of Apps (تعداد برنامه‌ها) | تعداد برنامه‌های توسعه‌یافته نشان‌دهنده توانایی شرکت در ارائه محصولات دیجیتال و نوآوری در بازار است. |
| 31 | Downloads Last 30 Days (دانلودهای 30 روز گذشته) | این ویژگی نشان‌دهنده میزان استقبال کاربران از محصولات دیجیتال شرکت است. |
| 32 | Monthly visit (بازدید ماهانه) | تعداد بازدیدهای ماهانه نشان‌دهنده میزان محبوبیت و تعامل شرکت با کاربران و مشتریان است. |
| 33 | Visit duration (مدت زمان بازدید) | مدت زمان بازدید نشان‌دهنده میزان تعامل کاربران با محتوای شرکت است. هرچه مدت زمان بازدید بیشتر باشد، احتمالاً جذابیت محتوا بیشتر است. |
| 34 | Page views/visit (تعداد صفحات مشاهده شده در هر بازدید) | این ویژگی میزان عمق تعامل کاربران با سایت یا پلتفرم شرکت را نشان می‌دهد. |
| 35 | Bounce Rate (نرخ پرش) | نرخ پرش نشان می‌دهد که چه تعداد از کاربران پس از ورود به سایت، بدون تعامل آن را ترک کرده‌اند. نرخ پایین‌تر معمولاً نشان‌دهنده تعامل بالاتر و جذابیت بیشتر سایت است. |
| 36 | Global Traffic Rank (رتبه ترافیک جهانی) | رتبه ترافیک جهانی سایت نشان‌دهنده میزان بازدیدهای سایت در مقایسه با سایر سایت‌ها است و می‌تواند به پیش‌بینی موفقیت شرکت کمک کند. |
| 37 | اکانت Facebook | داشتن حساب فیسبوک و میزان تعامل در این پلتفرم می‌تواند به ارزیابی فعالیت‌های بازاریابی و تعاملات شرکت کمک کند. |
| 38 | Twitter | فعالیت‌های توییتر می‌تواند شاخصی از حضور دیجیتال و تعاملات شرکت با جامعه آنلاین باشد. |
| 39 | linkedIn | لینکدین نشان‌دهنده حضور شرکت در شبکه‌های حرفه‌ای و میزان تعامل با نیروی کار و مشتریان حرفه‌ای است. |
| 40 | Active tech count (تعداد فناوری‌های فعال) | تعداد فناوری‌های فعال نشان‌دهنده میزان استفاده شرکت از فناوری‌های مختلف برای بهبود عملیات و ارائه خدمات است. |
| 41 | ویژگی‌های مرتبط با صنعت | Industry Groups (گروه‌های صنعتی(کلی تر)) | گروه‌های صنعتی نشان‌دهنده دسته‌بندی‌های کلی‌تری از صنایع هستند که شرکت در آن‌ها فعالیت دارد و می‌تواند به تحلیل بازار هدف کمک کند. |
| 42 | Industries (صنایع (جزئی تر)) | این ویژگی دقیق‌تر است و نشان‌دهنده صنایعی است که شرکت به‌طور خاص در آن‌ها فعالیت می‌کند. |

از آنجایی‌که ویژگی‌های متنوعی برای بررسی انتخاب شده اند، برای تقسیم بندی راحت‌تر این ویژگی‌ها به پنج گروه کلی تقسیم می‌شوند:

1- ویژگی‌های مربوط به شرکت: این ویژگی‌ها شامل اطلاعات پایه‌ای در مورد هر شرکت است، مانند تعداد کارکنان، موقعیت جغرافیایی دفتر مرکزی و میزان مواجهه با مشتریان. مطالعات قبلی نشان داده‌اند که این ویژگی‌ها تأثیر بسزایی در موفقیت استارتاپ‌ها دارند، به‌خصوص در مراحل اولیه رشد شرکت‌ها (Kim et al., 2023).

1. ویژگی‌های مرتبط با جذب سرمایه: وضعیت مالی و میزان سرمایه‌گذاری‌های انجام‌شده در شرکت‌ها یکی از عوامل مهم در موفقیت یا شکست استارتاپ‌ها است. داده‌های مربوط به تعداد دورهای جذب سرمایه، مقدار سرمایه جذب‌شده و تاریخ‌های مربوط به این سرمایه‌گذاری‌ها جمع‌آوری و تحلیل شده است. این اطلاعات می‌توانند شاخص مهمی برای ارزیابی پتانسیل موفقیت شرکت‌ها باشند، به‌ویژه در مراحل اولیه رشد آن‌ها (Deias & Magrini, 2023; Sharchilev et al., 2018)

3- ویژگی‌های مرتبط با سرمایه‌گذاران: نوع سرمایه‌گذاران و میزان دخالت آن‌ها در شرکت می‌تواند تأثیر مستقیمی بر موفقیت شرکت‌ها داشته باشد. مطالعات قبلی نشان داده‌اند که سرمایه‌گذاران بر اساس تجربه، قدرت چانه‌زنی و شبکه‌های ارتباطی خود می‌توانند نقش مهمی در موفقیت استارتاپ‌ها ایفا کنند (Kim et al., 2023).

4- ویژگی‌های مربوط به فناوری: فناوری‌های مورد استفاده در استارتاپ‌های هوش مصنوعی و تعداد پتنت‌ها و علائم تجاری ثبت‌شده توسط شرکت‌ها به‌عنوان شاخصی از نوآوری و توانمندی‌های فنی شرکت‌ها در نظر گرفته شده‌اند. فناوری‌های پیشرفته می‌توانند منجر به نوآوری‌های مخرب شوند و به شرکت‌ها کمک کنند تا بر رقبا غلبه کنند و در بازار موفق شوند (Kim et al., 2023).

5- ویژگی‌های مربوط به صنعت: ویژگی‌های صنعت شامل میزان نوآوری در صنعت، پایداری و جذابیت صنعت است. این ویژگی‌ها نشان‌دهنده میزان فرصت‌ها و منابع موجود در یک صنعت خاص برای موفقیت استارتاپ‌ها هستند. صنعت‌هایی که دارای موانع کمتر و رشد بالاتری هستند، معمولاً زمینه بهتری برای موفقیت شرکت‌ها فراهم می‌کنند (Kim et al., 2023).

با توجه به این پنج گروه ویژگی، تحلیل اولیه داده‌ها نشان داد که برخی از ویژگی‌های مرتبط با رسانه‌ها و میزان مواجهه شرکت‌ها با مشتریان، تعداد دورهای جذب سرمایه و میزان دخالت سرمایه‌گذاران، نقش کلیدی در موفقیت یا شکست شرکت‌های هوش مصنوعی دارند. از این‌رو، این متغیرها به‌عنوان ورودی‌های اصلی برای مدل‌های یادگیری ماشین مورد استفاده قرار گرفتند (Kim et al., 2023). در ادامه 42 ویژگی مختلف که برای پیش بینی موفقیت استارتاپ‌ها به کار گرفته شده‌اند به همراه دلیل انتخاب این ویژگی‌ها آورده شده است (جدول 3-1).

برچسب برای داده کاوی نیز با استفاده از عرضه اولیه، یونیکورن شدن و ادغام و تملیک به عنوان متغیر خروجی تهیه شد.

#### آماده‌سازی داده

در این مرحله، داده‌های جمع‌آوری‌شده به منظور آماده‌سازی برای تحلیل‌های بعدی و مدل‌سازی، از نظر کیفیت و کامل بودن مورد بررسی قرار گرفتند. این فرآیند شامل پاکسازی داده‌ها، حذف مقادیر گمشده و نرمال‌سازی متغیرها بود تا اطمینان حاصل شود که داده‌های ورودی به مدل‌های یادگیری ماشین دقیق و معتبر هستند. داده‌های گمشده با استفاده از روش‌هایی مانند جایگذاری میانگین و تکنیک‌های پیشرفته‌تر جایگزین شده و متغیرها به مقیاس‌های استاندارد تبدیل شدند تا عملکرد مدل‌ها بهبود یابد (Johnson, 2018).

در پروژه یادگیری ماشین، فرآیند پیش‌پردازش داده‌ها بخش اساسی از مرحله آماده‌سازی داده‌ها برای مدل‌سازی است. این فرآیند به‌منظور اطمینان از دقت و کارایی مدل‌های یادگیری ماشین، از جمله انتخاب ویژگی‌های مناسب، تبدیل داده‌ها به فرمت قابل استفاده و حذف یا جایگزینی داده‌های گمشده، انجام می‌شود. پیش‌پردازش داده‌ها شامل مجموعه‌ای از گام‌هاست که هدف آن‌ها تبدیل داده‌ها به شکلی است که برای آموزش مدل‌های یادگیری ماشین قابل استفاده باشد. این فرآیند معمولاً شامل حذف ویژگی‌های غیرضروری، کدگذاری داده‌های غیر عددی به فرمت عددی و انتخاب ویژگی‌های مناسب است. برای مدیریت داده‌های گمشده در این پژوهش داده‌های بدون ویژگی‌های صنعت حذف شدند. پس از شناسایی مقادیر گمشده، برای پرکردن آن‌ها از روش‌های مختلفی استفاده می‌شود. این روش‌ها به‌طور معمول شامل جایگزینی مقادیر گمشده با میانگین، میانه یا مد متغیرهای مربوطه است. در برخی از موارد، از مدل‌های پیش‌بینی KNN و یا جنگل تصادفی برای تخمین مقادیر گمشده استفاده شده است (Kim et al., 2023).

در این پژوهش، برای بررسی فعالیت در شبکه‌های اجتماعی فیسبوک، توییتر و لینکدین مقادیر صفر به ازای نداشتن اکانت و یک به ازای فعال بودن در فضای مجازی جایگزینی شد. در قدم بعدی، تعداد کارمندان که به صورت بازه در وبسایت کرانچ بیس داده شده بود، به دو ستون حد بالا و پایین تقسیم شد. برای پر کردن مقادیر گمشده از مدل KNN استفاده شده است. این کار برای درآمد تخمینی شرکت نیز انجام شد (جدا کردن حد بالا و پایین و پر کردن مقادیر گمشده با استفاده از مدل پیش بینی KNN). برای برخی ویژگی‎های فناوری، برای مثال تعداد بازدید روزانه، تعداد فناوری‌های فعال، تعداد مقالات، بازدید ماهانه، پتنت‌های ثبت شده، تعداد اپلیکیشن‌ها و غیره داده‌های گمشده صفر در نظر گرفته شدند (به این معنا که این شرکت در این حوزه‌ها فعالیتی ندارد). مقادیر گمشده در بخش هزینه‌ها، مانند هزینه فناوری، مقادیر سرمایه‌گذاری، و سایر ویژگی‌های مالی نیز با مدل پیش‌بینی KNN پر شده‌اند. در نهایت ویژگی‌های عرضه سهام اولیه، ادغام و تملیک و یونیکورن شدن به عنوان معیارهای موفقیت تعریف شدند.

با استناد به کیم و همکاران (Kim et al., 2023) داده‌هایی که ویژگی‌های مربوط به صنعت یا تعداد کارمندان را نداشتند، حذف شدند (253 داده). پس از حذف این ویژگی‌ها 3879 ویژگی برای بررسی باقی ماند.

برای مقادیری که در برخی متغیرها (مثل تعداد دانلود اپلیکیشن‌ها، تعداد پتنت‌ها و هزینه‌های IT) از دست رفته بودند، از K-Nearest Neighbors (KNN) imputation برای جایگزینی استفاده شد.

در برخی موارد مجموعه داده دارای عدم تعادل شدید بود (391 شرکت موفق در مقابل 3488 شرکت ناموفق). برای رفع این مشکل، از SMOTE[[53]](#footnote-53) استفاده شد که با ایجاد داده‌های مصنوعی برای کلاس‌های کم‌نمونه، تعادل نسبی را برقرار کرد.

#### مدل‌سازی

برای پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل و نقل، از چندین الگوریتم پیشرفته یادگیری ماشین استفاده شده است. هر یک از این الگوریتم‌ها دارای ویژگی‌ها و قابلیت‌های خاصی هستند که آن‌ها را برای تحلیل داده‌های بزرگ و پیچیده مناسب می‌سازد. برای جداسازی داده‌های آزمون و داده‎های آموزشی از روش K-Fold استفاده شده است. این روش یکی از تکنیک‌های پرکاربرد در ارزیابی مدل‌های یادگیری ماشین است که به کاهش واریانس در ارزیابی عملکرد مدل کمک می‌کند. در این روش، مجموعه داده به K **قسمت مساوی** تقسیم شده و مدل K **بار** آموزش داده می‌شود، به‌طوری که در هر تکرار یکی از بخش‌ها به‌عنوان **مجموعه تست** و باقی بخش‌ها به عنوان **مجموعه آموزش** در نظر گرفته می‌شوند. این فرآیند باعث استفاده بهینه از داده‌ها شده و ارزیابی مدل را مستقل از نحوه‌ی تصادفی تقسیم‌بندی داده‌ها می‌کند (Kohavi, 1995). از جمله مزایای این روش می‌توان به **کاهش وابستگی به نمونه‌های خاص و افزایش دقت ارزیابی مدل** اشاره کرد، اما از طرفی اجرای چندباره مدل منجر به **افزایش زمان محاسباتی** می‌شود. مقدار K معمولاً 5 یا 10 انتخاب می‌شود، زیرا تعادل مناسبی بین **دقت ارزیابی و هزینه‌ی محاسباتی** ایجاد می‌کند (Bishop & Nasrabadi, 2006). در این پژوهش مقدار k، 10 در نظر گرفته شده است.

در این تحقیق، از هفت الگوریتم استفاده شده است: رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، تقویت گرادیان، بیز ساده، شبکه عصبی مصنوعی، و ماشین بردار پشتیبان.

* **رگرسیون لجستیک:** یک مدل آماری برای پیش‌بینی احتمال وقوع یک نتیجه دسته‌بندی‌شده است. این مدل از یک تابع لجستیک (سیگموید) برای تبدیل ورودی‌های خطی به احتمالات استفاده می‌کند. به طور معمول، رگرسیون لجستیک در مسائل دسته‌بندی باینری مانند تشخیص بیماری یا تشخیص اسپم استفاده می‌شود.
* **درخت تصمیم:** رخت تصمیم یک مدل پیش‌بینی است که با تقسیم داده‌ها به زیرمجموعه‌ها بر اساس ویژگی‌ها، یک درخت از تصمیمات می‌سازد. هر گره درخت یک شرط بر روی ویژگی‌هاست و برگ‌ها نتایج پیش‌بینی هستند. این مدل ساده و قابل تفسیر است و به خوبی می‌تواند دسته‌بندی و رگرسیون انجام دهد.
* **جنگل تصادفی:** این الگوریتم یک مدل پیشرفته است که از ترکیب چندین درخت تصمیم‌گیری استفاده می‌کند تا دقت پیش‌بینی را افزایش دهد. جنگل تصادفی به‌خصوص در مواجهه با داده‌های نویزی و نامتوازن کارایی بالایی دارد و توانایی کاهش خطاهای ناشی از پیچیدگی‌های موجود در داده‌ها را داراست .
* **تقویت گرادیان:** یک الگوریتم یادگیری ماشینی است که در آن مدل‌ها به صورت ترتیبی ساخته می‌شوند. در هر مرحله، مدل جدید به گونه‌ای آموزش داده می‌شود که خطاهای مدل‌های قبلی را اصلاح کند. این روش از تابع هزینه برای هدایت فرآیند آموزش استفاده می‌کند و معمولاً دقت بالایی در پیش‌بینی دارد.
* بیز ساده: یک مدل دسته‌بندی است که بر اساس قاعده بیز و فرض استقلال ویژگی‌ها، احتمال هر کلاس را محاسبه می‌کند. این روش به‌ویژه برای مسائل با داده‌های با ابعاد بالا مانند متون کاربرد دارد. هرچند فرض استقلال ویژگی‌ها ممکن است در دنیای واقعی به طور کامل صحیح نباشد، اما الگوریتم به‌طور غیرمنتظره‌ای خوب عمل می‌کند.
* **شبکه عصبی مصنوعی:** شبکه‌های عصبی مصنوعی با الهام از ساختار مغز انسان برای شناسایی الگوهای پیچیده در داده‌ها طراحی شده‌اند. این مدل با استفاده از چندین لایه پنهان، روابط غیرخطی پیچیده را شناسایی می‌کند و می‌تواند در مواجهه با حجم بالای داده‌ها عملکرد بسیار خوبی داشته باشد .
* **ماشین بردار پشتیبان:** یک الگوریتم دسته‌بندی است که با پیدا کردن بهترین مرز (هیپراسپیس) بین دسته‌ها، پیش‌بینی می‌کند. این مدل به‌ویژه در داده‌های با ابعاد بالا و دسته‌بندی‌های پیچیده مفید است. ماشین بردار پشتیبان می‌تواند به‌طور موثر از کِرنِل‌ها برای شبیه‌سازی مرزهای غیرخطی استفاده کند و عملکرد خوبی در مشکلات مختلف دسته‌بندی دارد.

#### ارزیابی عملکرد مدل‌ها

پس از ایجاد مدل‌های مختلف، مرحله ارزیابی عملکرد آن‌ها آغاز می‌شود. در این پژوهش، برای ارزیابی دقت و کارایی مدل‌ها از معیارهای مختلفی استفاده شده است. این معیارها شامل ماتریس سردرگمی، دقت، بازده، فراخوانی، و امتیاز F1 هستند. هر یک از این معیارها جنبه‌های مختلفی از عملکرد مدل را اندازه‌گیری می‌کنند و به ما کمک می‌کنند تا مدل‌ها را براساس توانایی آن‌ها در پیش‌بینی موفقیت یا شکست شرکت‌ها ارزیابی کنیم.

* **دقت**: دقت یکی از مهم‌ترین معیارهای ارزیابی مدل‌های یادگیری ماشین است و نشان‌دهنده نسبت تعداد پیش‌بینی‌های صحیح به کل پیش‌بینی‌ها است. دقت معمولاً به‌عنوان معیار کلی برای ارزیابی مدل‌ها در نظر گرفته می‌شود، اما در مواجهه با داده‌های نامتوازن ممکن است کافی نباشد.
* ماتریس سردرگمی: ماتریس سردرگمی، تعداد پیش‌بینی‌های صحیح و نادرست را در دسته‌های مختلف نمایش می‌دهد. این ماتریس به شما کمک می‌کند تا نحوه عملکرد مدل را در شناسایی هر کلاس بررسی کنید. چهار مقدار اصلی در ماتریس سردرگمی عبارتند از:

True Positives (TP): پیش‌بینی صحیح برای کلاس مثبت.

True Negatives (TN): پیش‌بینی صحیح برای کلاس منفی.

False Positives (FP): پیش‌بینی اشتباه برای کلاس مثبت.

False Negatives (FN): پیش‌بینی اشتباه برای کلاس منفی.

* **صحت**: صحت معیاری است که نشان می‌دهد از تمامی پیش‌بینی‌های مثبت مدل، چه درصدی واقعاً صحیح بوده‌اند. به عبارت دیگر، صحت اندازه‌گیری می‌کند که مدل چقدر در پیش‌بینی‌های مثبت خود دقیق است. فرمول آن به صورت زیر است:
* **فراخوانی**: فراخوانی معیاری است که نشان می‌دهد از تمامی نمونه‌های واقعی مثبت، چه درصدی توسط مدل شناسایی و پیش‌بینی شده‌اند. به عبارت دیگر، فراخوانی اندازه‌گیری می‌کند که مدل چقدر توانسته است نمونه‌های مثبت واقعی را شناسایی کند. فرمول آن به صورت زیر است:
* **امتیاز** **F1**: امتیاز F1 ترکیبی از صحت و فراخوانی است و تعادل بین این دو معیار را اندازه‌گیری می‌کند. امتیاز F1 به‌ویژه زمانی اهمیت دارد که نیاز به یک معیار کلی برای ارزیابی مدل‌ها باشد که هم دقت و هم پوشش‌دهی را در نظر بگیرد .
* اعتبارسنجی متقابل: از این روش برای ارزیابی عمومی‌پذیری مدل‌ها و جلوگیری از وقوع اورفیتینگ[[54]](#footnote-54) استفاده شده است. در این روش، داده‌ها به چندین بخش تقسیم می‌شوند و مدل چندین بار با داده‌های متفاوت آموزش داده می‌شود و در هر بار، عملکرد آن ارزیابی می‌شود. اعتبارسنجی متقابل به‌ویژه برای ارزیابی مدل‌هایی که با داده‌های پیچیده سر و کار دارند، بسیار کارآمد است.

#### تحلیل نتایج

پس از ارزیابی مدل‌ها، مرحله تحلیل نتایج آغاز می‌شود. در این مرحله، نتایج به‌دست‌آمده از مدل‌ها بررسی می‌شوند تا الگوها و عوامل کلیدی موفقیت شناسایی شوند. تحلیل نتایج به شناسایی ویژگی‌هایی که بیشترین تأثیر را بر موفقیت دارند، کمک می‌کند و همچنین می‌تواند به ارائه راهکارهایی برای بهبود استراتژی‌های کسب‌وکار در شرکت‌های نوپا منجر شود .

در این تحقیق، عوامل کلیدی موفقیت شامل تعداد کارکنان، دورهای سرمایه‌گذاری، میزان مواجهه شرکت با رسانه‌ها و مشتریان، فناوری‌های مورد استفاده، و میزان دخالت سرمایه‌گذاران شناسایی شده‌اند. این ویژگی‌ها در پیش‌بینی موفقیت یا شکست شرکت‌ها نقشی اساسی داشته‌اند و از طریق تحلیل‌های آماری و یادگیری ماشین به‌عنوان مهم‌ترین عوامل شناسایی شده‌اند.

## جمع بندی

در این بخش، مراحل مختلف پژوهش از جمله درک کسب‌وکار، درک داده، پیش‌پردازش داده‌ها، مدل‌سازی، و ارزیابی عملکرد مدل‌ها به‌طور مفصل بررسی شده است.

ابتدا در مرحله درک کسب‌وکار، پژوهش بر شناسایی عوامل کلیدی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل تمرکز داشته است. با توجه به پیچیدگی‌های این صنعت و ریسک‌های بالای آن، هدف اصلی این بخش شناسایی و تحلیل عواملی بود که می‌توانند موفقیت یا شکست شرکت‌ها را پیش‌بینی کنند. این عوامل شامل ویژگی‌هایی مانند تعداد کارکنان، مراحل جذب سرمایه، موقعیت جغرافیایی، و میزان تعامل شرکت با رسانه‌ها و مشتریان بودند که در مراحل بعدی برای پیش‌بینی عملکرد شرکت‌ها به‌کار گرفته شدند.

در مرحله درک داده، داده‌ها از پایگاه اطلاعاتی کرانچ‌بیس استخراج شدند. این پایگاه شامل اطلاعاتی از قبیل مراحل جذب سرمایه، تعداد کارکنان، ویژگی‌های جغرافیایی و بنیان‌گذاران و سرمایه‌گذاران است که می‌تواند تأثیر زیادی بر موفقیت یا شکست استارتاپ‌ها داشته باشد. پس از جمع‌آوری داده‌ها، این داده‌ها به پنج گروه کلی تقسیم شدند: ویژگی‌های شرکت، ویژگی‌های جذب سرمایه، ویژگی‌های سرمایه‌گذاران، ویژگی‌های فناوری و ویژگی‌های صنعت. این تقسیم‌بندی به محققان کمک کرد تا بتوانند متغیرهای کلیدی را شناسایی کرده و مدل‌های پیش‌بینی موفقیت را با دقت بیشتری طراحی کنند.

در مرحله پیش‌پردازش داده‌ها، داده‌ها از نظر کیفیت بررسی شده و مقادیر گمشده با استفاده از روش‌های مختلف جایگزینی مانند مدل KNN و میانگین پر شدند. همچنین داده‌های نامتوازن حذف شده و ویژگی‌های مختلف به مقیاس‌های استاندارد تبدیل شدند تا الگوریتم‌ها بتوانند به‌طور مؤثرتری آن‌ها را تحلیل کنند. این پیش‌پردازش‌ها بهبود دقت و کارایی مدل‌ها را در مراحل بعدی فراهم آورد.

در مرحله مدل‌سازی، از هفت الگوریتم مختلف یادگیری ماشین شامل رگرسیون لجستیک، درخت تصمیم، جنگل تصادفی، تقویت گرادیان، بیز ساده، شبکه عصبی مصنوعی و ماشین بردار پشتیبان استفاده شد. این الگوریتم‌ها به‌طور خاص برای پیش‌بینی موفقیت یا شکست استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل طراحی شدند. برای ارزیابی مدل‌ها، از روش K-Fold برای تقسیم داده‌ها به بخش‌های مختلف استفاده شد که به کاهش واریانس و افزایش دقت ارزیابی مدل‌ها کمک کرد. علاوه بر این، ارزیابی عملکرد مدل‌ها با استفاده از معیارهایی مانند ماتریس سردرگمی، دقت، بازده، فراخوانی، و امتیاز F1 انجام شد تا اطمینان حاصل شود که مدل‌ها توانایی پیش‌بینی صحیح موفقیت یا شکست شرکت‌ها را دارند.

در نهایت، این تحقیق به‌طور جامع به تحلیل داده‌ها و پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل می‌پردازد. با استفاده از روش‌های مختلف پیشرفته یادگیری ماشین و تحلیل دقیق ویژگی‌های مرتبط با کسب‌وکار، نتایج این پژوهش می‌تواند به مدیران و سرمایه‌گذاران کمک کند تا تصمیمات بهتری در خصوص سرمایه‌گذاری در این صنعت اتخاذ کنند. همچنین، این پژوهش توانسته است مدل‌هایی مؤثر برای پیش‌بینی موفقیت شرکت‌ها در صنعت حمل‌ونقل ارائه دهد که می‌توانند در پروژه‌های آینده و تحلیل‌های بیشتر مورد استفاده قرار گیرند.

# نتایج و تفسیر آن‌ها

## مقدمه

در این فصل به تحلیل و تفسیر نتایج حاصل از تحقیق پرداخته و به‌طور خاص، به ارائه‌ی تحلیل‌های آماری و مدل‌سازی‌های انجام‌شده برای بررسی عوامل مؤثر بر موفقیت شرکت‌های فعال در صنعت هوش مصنوعی و حمل‌ونقل می‌پردازیم. هدف اصلی این فصل، تجزیه و تحلیل داده‌های جمع‌آوری‌شده و ارائه‌ی تفسیر دقیق و علمی از نتایج به‌دست‌آمده است. از آنجا که در این تحقیق سعی بر مدل‌سازی و پیش‌بینی موفقیت شرکت‌ها در این صنعت بوده است، تحلیل‌های آماری انجام‌شده در این فصل می‌تواند به شفاف‌سازی ارتباط میان ویژگی‌های مختلف شرکت‌ها و موفقیت آن‌ها در بازارهای رقابتی کمک کند.

داده‌های مورد استفاده در این تحقیق از شرکت‌های فعال در زمینه‌ی هوش مصنوعی در صنعت حمل‌ونقل جمع‌آوری شده است. این شرکت‌ها در حوزه‌های گوناگونی از جمله اتوماسیون، سیستم‌های هوش مصنوعی، فناوری‌های پیشرفته حمل‌ونقل و نرم‌افزارهای مرتبط با حمل‌ونقل فعالیت دارند. ویژگی‌هایی مانند تعداد کارمندان، درآمد تخمینی، وضعیت تأسیس، تعداد اختراعات و محصولات فعال، موقعیت جغرافیایی و نوع سرمایه‌گذاری، به‌عنوان داده‌های کلیدی برای انجام تحلیل‌های این تحقیق در نظر گرفته شده است. این داده‌ها شامل اطلاعات ۴۲ ویژگی مختلف از 4132 شرکت است که در راستای تحلیل‌ها و مدل‌سازی‌های انجام‌شده استفاده شده است.

در این فصل ابتدا به توصیف ویژگی‌های داده‌ها پرداخته می‌شود. بدین منظور، ابتدا همبستگی بین متغیرها به روشهای گوناگون بررسی شده و ویژگی‌های با همبستگی بالا حذف می‌شوند. سپس ویژگی‌های مختلف داده‌ها به‌دقت توضیح داده می‌شوند و توزیع آن‌ها بررسی می‌گردد. شناخت دقیق ویژگی‌های داده‌ها پیش‌نیاز هرگونه تحلیل معتبر و علمی است.

پس از توصیف ویژگی‌های داده‌ها، در ادامه به آمار توصیفی پرداخته می‌شود. آمار توصیفی شامل محاسباتی نظیر میانگین، انحراف معیار، بیشترین و کمترین مقدار و دامنه تغییرات برای هر یک از ویژگی‌های داده‌ها است. این محاسبات به پژوهشگران امکان می‌دهد تا توزیع داده‌ها را به‌طور دقیق تحلیل کنند و بتوانند نتیجه‌گیری‌های معناداری از آن‌ها استخراج کنند. برای تحلیل این داده‌ها، از ابزارهای گرافیکی مانند نمودارهای جعبه‌ای و هیستوگرام‌ها بهره‌برداری می‌شود تا توزیع ویژگی‌هایی مانند تعداد کارمندان، درآمد تخمینی و تعداد اختراعات و محصولات فعال به‌طور بصری نمایش داده شود. نمودارها و جداول توصیفی نه‌تنها اطلاعات پایه‌ای درباره ویژگی‌های داده‌ها فراهم می‌آورند، بلکه به‌ویژه در شناسایی مقادیر پرت و ناهنجاری‌ها نیز کمک شایانی می‌کنند.

در ادامه، این فصل به تحلیل نتایج تحقیق و ارتباطات بین متغیرهای مختلف داده‌ها اختصاص دارد. هدف از این بخش، تجزیه و تحلیل روابط بین ویژگی‌های مختلف و ارزیابی تأثیر آن‌ها بر موفقیت شرکت‌ها در صنعت هوش مصنوعی و حمل‌ونقل است. برای تحلیل این روابط از آزمون‌های آماری مختلف نظیر آزمون همبستگی پیرسون، رگرسیون خطی و مدل‌های پیچیده‌تر استفاده می‌شود. این آزمون‌ها می‌توانند به‌طور دقیق تأثیرات متغیرهای مختلف همچون اندازه شرکت، موقعیت جغرافیایی، نوع سرمایه‌گذاری و تعداد اختراعات را بر موفقیت شرکت‌ها بررسی کنند

در نهایت، نتایج به‌دست‌آمده در این فصل جمع‌بندی می‌شوند. این جمع‌بندی به‌طور خلاصه روابط شناسایی‌شده میان متغیرها و تأثیر آن‌ها بر موفقیت شرکت‌ها را بازبینی کرده و نتیجه‌گیری‌های کلی از تحلیل‌های انجام‌شده ارائه می‌دهد. این بخش از فصل همچنین به بررسی پیامدهای عملی نتایج تحقیق پرداخته و پیشنهاداتی برای تحقیقات آینده ارائه می‌دهد. در این جمع‌بندی، همچنین به کاربردهای عملی نتایج تحقیق در تصمیم‌گیری‌های استراتژیک برای مدیران شرکت‌های هوش مصنوعی و حمل‌ونقل پرداخته خواهد شد.

در نهایت، نتایج این تحقیق می‌توانند به توسعه مدل‌های پیش‌بینی کمک کرده و به‌عنوان ابزارهایی برای تحلیل استراتژیک در صنعت هوش مصنوعی و حمل‌ونقل مورد استفاده قرار گیرند. این مدل‌ها می‌توانند به مدیران کمک کنند تا تصمیمات بهتری در خصوص سرمایه‌گذاری، گسترش فعالیت‌ها و شناسایی فرصت‌های رشد اتخاذ کنند. به‌علاوه، نتایج تحقیق می‌تواند برای سیاست‌گذاران این صنعت مفید باشد تا با بهره‌گیری از داده‌ها و تحلیل‌های به‌دست‌آمده، تصمیمات بهتری در جهت حمایت از شرکت‌های نوپا و ارتقاء رقابت‌پذیری در این بخش از اقتصاد اتخاذ کنند.

## توصیف داده‌ها

انتخاب ویژگی‌های مناسب یکی از مراحل کلیدی در ساخت مدل‌های پیش‌بینی است، به‌ویژه زمانی که داده‌های اولیه شامل تعداد زیادی متغیر هستند. در این پژوهش، به منظور پیش‌بینی موفقیت استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل، 42 ویژگی از پایگاه داده کرانچ‌بیس استخراج شده است. این ویژگی‌ها، طیف گسترده‌ای از جنبه‌های کسب‌وکار، مانند تأمین مالی، نوع سرمایه‌گذاران، تاریخچه شرکت و روندهای صنعت را پوشش می‌دهند. اگرچه داشتن تعداد زیادی ویژگی می‌تواند اطلاعات ارزشمندی ارائه دهد، اما به همان اندازه ممکن است چالش‌هایی ایجاد کند.

یکی از مشکلات اصلی مرتبط با استفاده از تعداد زیادی ویژگی، هم‌خطی چندگانه[[55]](#footnote-55) است. هم‌خطی چندگانه به وضعیتی اشاره دارد که در آن دو یا چند ویژگی، همبستگی بالایی با یکدیگر دارند. این وضعیت باعث می‌شود که مدل پیش‌بینی نتواند تأثیر مستقل هر ویژگی را بر متغیر هدف به درستی شناسایی کند. این امر می‌تواند منجر به افزایش خطای پیش‌بینی، کاهش دقت مدل و ناپایداری ضرایب در روش‌های رگرسیون شود (James et al., 2013).

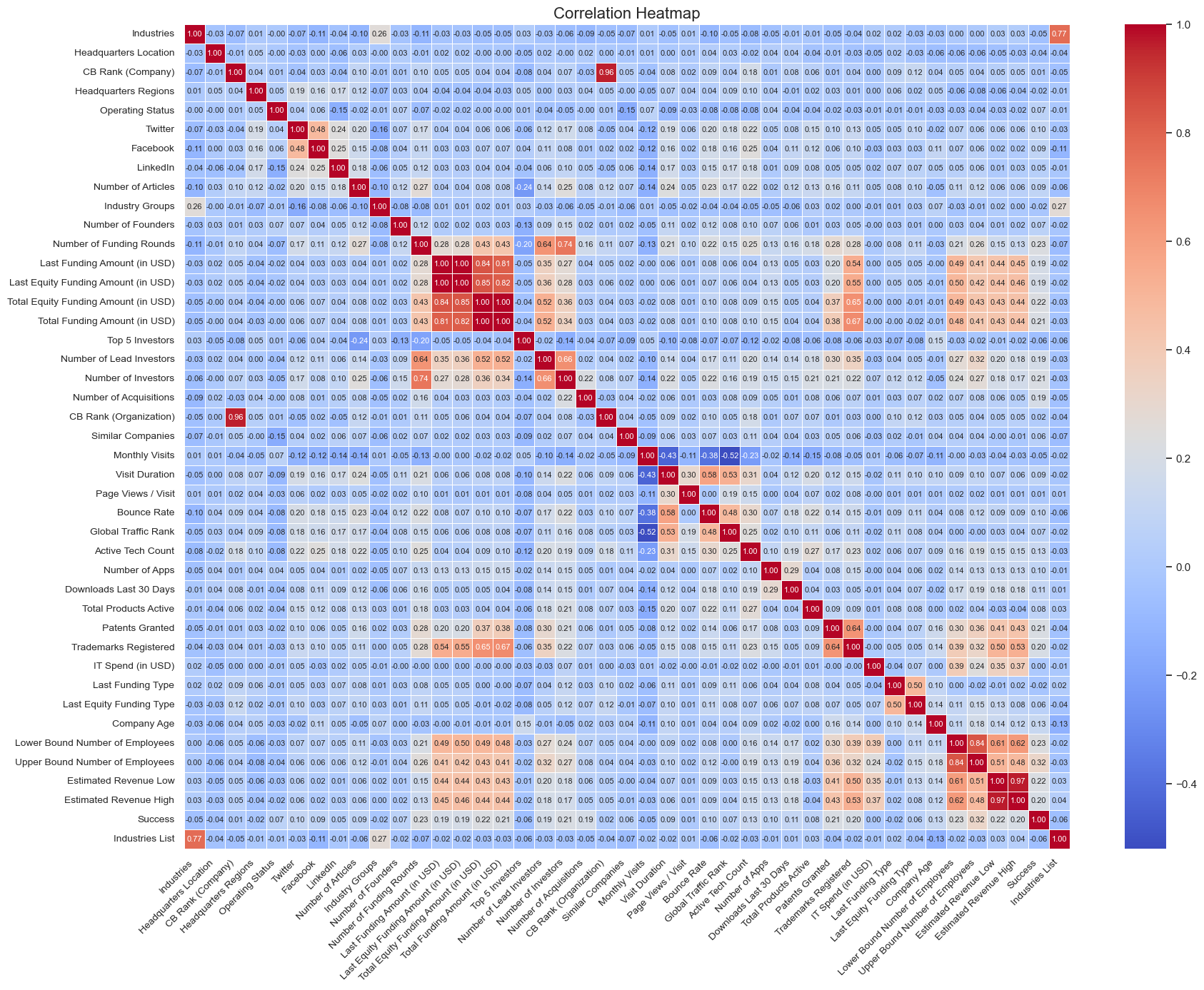
برای مقابله با این چالش، از ماتریس همبستگی استفاده می‌شود. این ماتریس روابط میان تمامی ویژگی‌ها را با استفاده از شاخص‌هایی نظیر ضریب همبستگی پیرسون[[56]](#footnote-56) و اسپیرمن[[57]](#footnote-57) (شکل 4-1 و شکل 4-2) نمایش می‌دهد. ضریب همبستگی عددی بین -1 و +1 است که میزان و جهت رابطه خطی بین دو متغیر را نشان می‌دهد:

* مقادیر نزدیک به +1 نشان‌دهنده همبستگی مثبت قوی است؛ به این معنا که افزایش یک ویژگی با افزایش دیگری همراه است.
* مقادیر نزدیک به -1 نشان‌دهنده همبستگی منفی قوی است؛ به این معنا که افزایش یک ویژگی با کاهش دیگری همراه است.
* مقادیر نزدیک به 0 نشان‌دهنده نبود همبستگی خطی بین دو ویژگی است.

در این پژوهش، ویژگی‌هایی که دارای ضریب همبستگی مطلق بیش از 6/0 هستند، به عنوان ویژگی‌های دارای همبستگی بالا در نظر گرفته می‌شوند. وجود این ویژگی‌ها ممکن است باعث شود که برخی از اطلاعات به صورت تکراری وارد مدل شوند، که این امر علاوه بر کاهش کارایی مدل، باعث افزایش پیچیدگی محاسباتی و زمان پردازش می‌شود (Kuhn, 2013).

در ادامه، برای بررسی همبستگی بین متغیرهای مختلف و موفقیت، همبستگی پیرسون و اسپیرمن با موفقیت نیز در شکل 4-1 و شکل 4-2 و همچنین جدول 4-1 و جدول 4-2 آورده شده است. این جداول برای

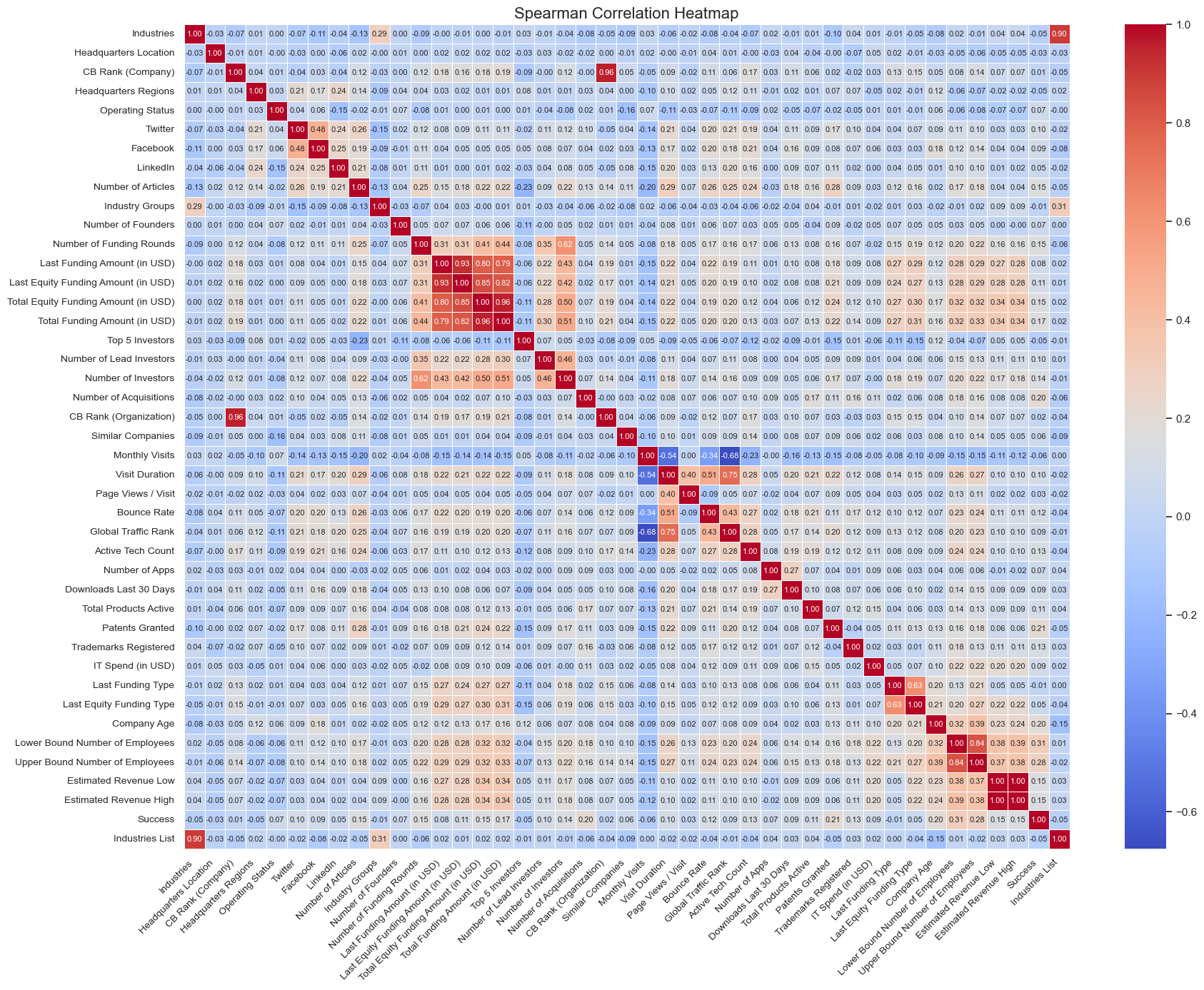
شناسایی متغیرهایی که بیشترین تأثیر را بر موفقیت دارند مفید است. متغیرهای با همبستگی بالا احتمالاً در مدل‌سازی حفظ می‌شوند، در حالی که متغیرهای با همبستگی ضعیف یا منفی می‌توانند حذف شوند.



**شکل 4- 1-** ماتریس همبستگی پیرسون

**جدول 4- 1- جدول همبستگی پیرسون ویژگی‌ها با موفقیت**

|  |  |  |
| --- | --- | --- |
| همبستگی با موفقیت | ویژگی‌ها |  |
| 1 | Success | 1 |
| 0.317362 | Upper Bound Number of Employees | 2 |
| 0.232648 | Lower Bound Number of Employees | 3 |
| 0.227699 | Number of Funding Rounds | 4 |
| 0.22481 | Total Equity Funding Amount (in USD) | 5 |
| 0.218858 | Estimated Revenue Low | 6 |
| 0.214909 | Patents Granted | 7 |
| 0.212597 | Number of Investors | 8 |
| 0.212599 | Total Funding Amount (in USD) | 9 |
| 0.200634 | Estimated Revenue High | 10 |
| 0.199859 | Trademarks Registered | 11 |
| 0.190279 | Last Equity Funding Amount (in USD) | 12 |
| 0.189638 | Last Funding Amount (in USD) | 13 |
| 0.187059 | Number of Acquisitions | 14 |
| 0.185097 | Number of Lead Investors | 15 |
| 0.127497 | Active Tech Count | 16 |
| 0.125465 | Company Age | 17 |
| 0.106691 | Downloads Last 30 Days | 18 |
| 0.104167 | Twitter | 19 |
| 0.102154 | Bounce Rate | 20 |
| 0.096722 | Number of Apps | 21 |
| 0.090391 | Visit Duration | 22 |
| 0.089847 | Number of Articles | 23 |
| 0.088634 | Facebook | 24 |
| 0.073594 | Total Products Active | 25 |
| 0.071643 | Number of Founders | 26 |
| 0.071834 | Operating Status | 27 |
| 0.069219 | Last Equity Funding Type | 28 |
| 0.068138 | Global Traffic Rank | 29 |
| 0.065111 | Similar Companies | 30 |
| 0.045078 | LinkedIn | 31 |
| 0.023353 | CB Rank (Organization) | 32 |
| 0.014799 | CB Rank (Company) | 33 |
| 0.012129 | Excluded Industries | 34 |
| 0.009301 | Page Views / Visit | 35 |
| 0.001272 | IT Spend (in USD) | 36 |
| -0.01559 | Industry Groups | 37 |
| -0.02056 | Last Funding Type | 38 |
| -0.02089 | Headquarters Regions | 39 |
| -0.02709 | Headquarters Location | 40 |
| -0.05019 | Industries | 41 |
| -0.05181 | Monthly Visits | 42 |
| -0.06131 | Top 5 Investors | 43 |
| -0.06618 | Industries List | 44 |



**شکل 4- 2- ماتریس همبستگی اسپیرمن**

**جدول 4- 2- جدول همبستگی ویژگی‌ها اسپیرمن با موفقیت**

|  |  |  |
| --- | --- | --- |
| همبستگی با موفقیت | ویژگی‌ها |  |
| 1.000000 | Success | 1 |
| 0.308868 | Lower Bound Number of Employees | 2 |
| 0.281047 | Upper Bound Number of Employees | 3 |
| 0.206278 | Patents Granted | 4 |
| 0.204950 | Number of Acquisitions | 5 |
| 0.199611 | Company Age | 6 |
| 0.166869 | Total Funding Amount (in USD) | 7 |
| 0.151594 | Number of Articles | 8 |
| 0.151299 | Number of Funding Rounds | 9 |
| 0.148502 | Estimated Revenue High | 10 |
| 0.146308 | Total Equity Funding Amount (in USD) | 11 |
| 0.145549 | Estimated Revenue Low | 12 |
| 0.142011 | Number of Investors | 13 |
| 0.125953 | Trademarks Registered | 14 |
| 0.125590 | Active Tech Count | 15 |
| 0.117071 | Bounce Rate | 16 |
| 0.108936 | Last Equity Funding Amount (in USD) | 17 |
| 0.106080 | Total Products Active | 18 |
| 0.104167 | Twitter | 19 |
| 0.102998 | Number of Lead Investors | 20 |
| 0.097270 | Visit Duration | 21 |
| 0.091874 | Downloads Last 30 Days | 22 |
| 0.088634 | Facebook | 23 |
| 0.085680 | Global Traffic Rank | 24 |
| 0.085253 | IT Spend (in USD) | 25 |
| 0.079965 | Last Funding Amount (in USD) | 26 |
| 0.072349 | Number of Apps | 27 |
| 0.071384 | Operating Status | 28 |
| 0.065144 | Number of Founders | 29 |
| 0.060580 | Similar Companies | 30 |
| 0.054018 | Last Equity Funding Type | 31 |
| 0.045078 | LinkedIn | 32 |
| 0.028015 | Page Views / Visit | 33 |
| 0.023353 | CB Rank (Organization) | 34 |
| 0.014799 | CB Rank (Company) | 35 |
| -0.00734 | Last Funding Type | 36 |
| -0.01348 | Industry Groups | 37 |
| -0.03409 | Headquarters Location | 38 |
| -0.04583 | Industries List | 39 |
| -0.04742 | Headquarters Regions | 40 |
| -0.04987 | Industries | 41 |
| -0.05463 | Top 5 Investors | 42 |
| -0.06112 | Monthly Visits | 43 |

پس از بررسی ماتریس‌های همبستگی، ویژگی‌هایی که دارای همبستگی بالایی هستند، به منظور بهبود دقت مدل، کاهش پیچیدگی‌های مدل، کاهش بیش برازش[[58]](#footnote-58) و تفسیر پذیری بهتر مدل حذف شوند و تنها یکی از این متغیرها حفظ شود. با توجه به نتایج ماتریس‌های همبستگی، Number of lead investors و Number of investors با هم همبستگی بالایی دارند. همچنین 4 ویژگی Last funding amount، Last equity amount، Total equity funding amount و Total funding amount. بنابراین از میان این ویژگی‌ها باید تنها یکی از آن‌ها باقی بماند و سایر ویژگی‌ها (5 ویژگی) حذف شوند.

معیار انتخاب متغیری که باید حذف شود یا باقی بماند، بر اساس اهمیت ویژگی‌ها[[59]](#footnote-59) تعیین می‌شود. اهمیت ویژگی‌ها به‌عنوان معیاری برای سنجش میزان تأثیر هر ویژگی بر متغیر هدف (متغیری که مدل تلاش می‌کند آن را پیش‌بینی کند) در نظر گرفته می‌شود. برای محاسبه این معیار، از 5 الگوریتم مختلف یادگیری ماشین استفاده می‌شود. هر مدل از یک روش متمایز برای تحلیل داده‌ها و شناسایی تأثیر ویژگی‌ها بر متغیر هدف استفاده می‌کند. برای مثال، رگرسیون خطی روابط خطی بین ویژگی‌ها و متغیر هدف را تحلیل می‌کند و از ضرایب ویژگی‌ها به‌عنوان شاخص تأثیر استفاده می‌کند. در مقابل، مدل‌هایی مانند جنگل تصادفی و XGBoost روابط پیچیده و غیرخطی را شناسایی کرده و از کاهش خطا در ساختارهای مبتنی بر درخت برای ارزیابی اهمیت ویژگی‌ها بهره می‌برند.

این تنوع در رویکردها کمک می‌کند تا تمامی انواع روابط، چه خطی و چه غیرخطی، بین ویژگی‌ها و متغیر هدف پوشش داده شوند. همچنین، استفاده از مدل‌های متعدد باعث می‌شود نتایج کمتر تحت تأثیر محدودیت‌ها یا سوگیری‌های خاص یک مدل قرار بگیرند. برای مثال، رگرسیون خطی ممکن است در صورت وجود هم‌خطی چندگانه نتایج دقیقی ارائه ندهد، اما مدل‌هایی مانند XGBoost در این شرایط دقیق‌تر عمل می‌کنند.

این رویکرد چندمدلی همچنین امکان شناسایی دقیق‌تر ویژگی‌های کلیدی را فراهم می‌کند. ویژگی‌هایی که در اکثر مدل‌ها اهمیت بالایی دارند، به‌عنوان عوامل اساسی در موفقیت شرکت‌ها شناسایی می‌شوند. از سوی دیگر، ویژگی‌هایی که در تمامی مدل‌ها اهمیت کمی دارند، می‌توانند در مرحله کاهش ابعاد حذف شوند. در نهایت، این رویکرد به کاهش خطا، افزایش دقت پیش‌بینی، و تفسیرپذیری بهتر مدل کمک می‌کند و نتایج حاصل را برای استفاده در تصمیم‌گیری‌های عملیاتی قابل اعتمادتر می‌سازد (Alpaydin, 2020). مدل‌های استفاده شده برای تعیین اهمیت ویژگی‌ها در ادامه آورده شده اند:

1. **رگرسیون خطی:** در رگرسیون خطی، اهمیت ویژگی‌ها با استفاده از ضرایب محاسبه می‌شود. این ضرایب نشان می‌دهند که تغییر در هر ویژگی چگونه بر متغیر هدف تأثیر می‌گذارد. ویژگی‌هایی که دارای ضرایب بزرگ‌تر (به لحاظ مقدار مطلق) هستند، اهمیت بیشتری دارند، زیرا تأثیر بیشتری بر پیش‌بینی متغیر هدف دارند. این روش برای داده‌هایی که روابط خطی بین متغیرها و هدف برقرار است، مناسب است و ضرایب آن به‌عنوان شاخصی مستقیم برای اهمیت ویژگی‌ها استفاده می‌شوند. (جدول4-3)

مزایا: ساده و قابل فهم. مناسب برای داده‌های خطی.

معایب: عدم توانایی در مدل‌سازی روابط غیرخطی.حساس به هم‌خطی چندگانه.

**جدول 4- 3- ضرایب ویژگی‌ها بر اساس مدل رگرسیون خطی**

|  |  |  |
| --- | --- | --- |
| اهمیت ویژگی‌ها | ویژگی‌ها |  |
| 0.189202 | Number of Acquisitions | 1 |
| 0.140947 | Operating Status | 2 |
| 0.027176 | Twitter | 3 |
| 0.016298 | Number of Funding Rounds | 4 |
| 0.012016 | Number of Founders | 5 |
| 0.009241 | Facebook | 6 |
| 0.002511 | Total Products Active | 7 |
| 0.002066 | Number of Apps | 8 |
| 0.002021 | Trademarks Registered | 9 |
| 0.001749 | LinkedIn | 10 |
| 0.001619 | Downloads Last 30 Days | 11 |
| 0.001511 | Company Age | 12 |
| 0.000797 | Number of Investors | 13 |
| 0.000644 | Last Equity Funding Type | 14 |
| 0.000618 | Similar Companies | 15 |
| 0.000508 | Patents Granted | 16 |
| 0.000115 | CB Rank (Organization) | 17 |
| 7.79E-05 | Upper Bound Number of Employees | 18 |
| 7.58E-05 | Global Traffic Rank | 19 |
| 5.83E-05 | Monthly Visits | 20 |
| 2.19E-05 | Industries | 21 |
| 1.35E-05 | Active Tech Count | 22 |
| 5.57E-06 | Industry Groups | 23 |
| 1.49E-09 | Total Equity Funding Amount (in USD) | 24 |
| 3.72E-10 | Last Funding Amount (in USD) | 25 |
| 3.58E-10 | Estimated Revenue Low | 26 |
| 6.88E-12 | Estimated Revenue High | 27 |
| 6.93E-10 | Last Equity Funding Amount (in USD) | 28 |
| 7.18E-10 | IT Spend (in USD) | 29 |
| -1.1E-09 | Total Funding Amount (in USD) | 30 |
| -2.3E-07 | Visit Duration | 31 |
| -4.1E-06 | Bounce Rate | 32 |
| -3.4E-05 | Headquarters Location | 33 |
| -7.2E-05 | Lower Bound Number of Employees | 34 |
| -0.0001 | Number of Articles | 35 |
| -0.00011 | CB Rank (Company) | 36 |
| -0.00013 | Top 5 Investors | 37 |
| -0.00046 | Headquarters Regions | 38 |
| -0.00268 | Industries List | 39 |
| -0.00406 | Page Views / Visit | 40 |
| -0.0041 | Last Funding Type | 41 |
| -0.00624 | Number of Lead Investors | 42 |

رگرسیون خطی به ما نشان می‌دهد که ویژگی‌هایی مانند تعداد ادغام‌ها و تملیک‌ها، وضعیت عملیاتی و تعداد دورهای تأمین مالی (Number of Funding Rounds) از مهم‌ترین عوامل تأثیرگذار مثبت بر موفقیت هستند.

1. **درخت تصمیم:** درخت تصمیم به‌عنوان یکی از روش‌های ساده و قابل فهم یادگیری ماشین، برای محاسبه اهمیت ویژگی‌ها از شاخص‌هایی نظیر کاهش آنتروپی یا شاخص جینی استفاده می‌کند. ویژگی‌هایی که در سطوح بالاتری از درخت قرار می‌گیرند، معمولاً اهمیت بیشتری دارند، زیرا تأثیر بیشتری بر طبقه‌بندی یا پیش‌بینی داده‌ها داشته‌اند (جدول 4-4).

مزایا: ساده و قابل تفسیر.توانایی مدیریت داده‌های غیرخطی.

معایب: در برابر بیش‌برازش حساس است.

**جدول 4- 4- ضرایب ویژگی‌ها بر اساس مدل درخت تصمیم**

|  |  |  |
| --- | --- | --- |
| اهمیت ویژگی‌ها | ویژگی‌ها |  |
| 0.138846 | Lower Bound Number of Employees | 1 |
| 0.074623 | Headquarters Location | 2 |
| 0.059537 | Headquarters Regions | 3 |
| 0.052612 | Total Funding Amount (in USD) | 4 |
| 0.048428 | Trademarks Registered | 5 |
| 0.043483 | Top 5 Investors | 6 |
| 0.041367 | Company Age | 7 |
| 0.040842 | Estimated Revenue Low | 8 |
| 0.035848 | Active Tech Count | 9 |
| 0.033616 | Last Funding Type | 10 |
| 0.031207 | Last Funding Amount (in USD) | 11 |
| 0.031015 | CB Rank (Organization) | 12 |
| 0.02882 | Total Products Active | 13 |
| 0.024071 | CB Rank (Company) | 14 |
| 0.023957 | Number of Founders | 15 |
| 0.022824 | IT Spend (in USD) | 16 |
| 0.021697 | Number of Articles | 17 |
| 0.02074 | Estimated Revenue High | 18 |
| 0.019868 | Last Equity Funding Type | 19 |
| 0.019817 | Visit Duration | 20 |
| 0.019245 | Upper Bound Number of Employees | 21 |
| 0.018167 | Downloads Last 30 Days | 22 |
| 0.016967 | Number of Funding Rounds | 23 |
| 0.015079 | Patents Granted | 24 |
| 0.014187 | Global Traffic Rank | 25 |
| 0.012611 | Operating Status | 26 |
| 0.01135 | Facebook | 27 |
| 0.01092 | Number of Investors | 28 |
| 0.010707 | Number of Acquisitions | 29 |
| 0.00949 | Industry Groups | 30 |
| 0.008658 | Similar Companies | 31 |
| 0.008193 | Number of Apps | 32 |
| 0.007882 | Industries List | 33 |
| 0.007455 | Bounce Rate | 34 |
| 0.006618 | Monthly Visits | 35 |
| 0.005507 | Industries | 36 |
| 0.004054 | Last Equity Funding Amount (in USD) | 37 |
| 0 | Page Views / Visit | 38 |
| 0 | Number of Lead Investors | 39 |
| 0 | Total Equity Funding Amount (in USD) | 40 |
| 0 | LinkedIn | 41 |
| 0 | Twitter | 42 |

مدل درخت تصمیم به ما نشان می‌دهد که ویژگی‌هایی مانند تعداد کارمندان، مکان دفتر مرکزی و کل سرمایه جذب شده (Total Funding Amount) از مهم‌ترین عوامل تأثیرگذار مثبت بر موفقیت هستند.

1. جنگل تصادفی: جنگل تصادفی مجموعه‌ای از درخت‌های تصمیم است که هر یک از آن‌ها با نمونه‌گیری تصادفی از داده‌ها و ویژگی‌ها ایجاد می‌شود. اهمیت ویژگی‌ها در این الگوریتم از طریق تأثیر آن‌ها بر کاهش میانگین خطای طبقه‌بندی یا پیش‌بینی در تمامی درخت‌ها محاسبه می‌شود. این روش به دلیل استفاده از میانگین نتایج چندین درخت، مقاوم‌تر در برابر نویز است (جدول 4-5).

مزایا: مقاوم در برابر بیش‌برازش. توانایی کار با داده‌های دارای نویز.

معایب: کاهش تفسیرپذیری مدل. نیاز به منابع محاسباتی بالا.

جدول 4- 5- ضرایب ویژگی‌ها بر اساس مدل جنگل تصادفی

|  |  |  |
| --- | --- | --- |
| اهمیت ویژگی‌ها | ویژگی‌ها |  |
| 0.045446 | Upper Bound Number of Employees | 1 |
| 0.044102 | Headquarters Location | 2 |
| 0.041432 | Total Equity Funding Amount (in USD) | 3 |
| 0.041342 | Industries | 4 |
| 0.040743 | CB Rank (Company) | 5 |
| 0.040376 | Lower Bound Number of Employees | 6 |
| 0.039134 | CB Rank (Organization) | 7 |
| 0.038783 | Estimated Revenue Low | 8 |
| 0.037631 | Company Age | 9 |
| 0.035916 | Estimated Revenue High | 10 |
| 0.035370 | Total Funding Amount (in USD) | 11 |
| 0.029248 | Active Tech Count | 12 |
| 0.029956 | Patents Granted | 13 |
| 0.029087 | Industry Groups | 14 |
| 0.028911 | Top 5 Investors | 15 |
| 0.027917 | Headquarters Regions | 16 |
| 0.027869 | Last Equity Funding Type | 17 |
| 0.025333 | Similar Companies | 18 |
| 0.025186 | Last Funding Type | 19 |
| 0.024867 | IT Spend (in USD) | 20 |
| 0.023495 | Last Funding Amount (in USD) | 21 |
| 0.023956 | Last Equity Funding Amount (in USD) | 22 |
| 0.022123 | Number of Articles | 23 |
| 0.020881 | Total Products Active | 24 |
| 0.019781 | Monthly Visits | 25 |
| 0.018466 | Number of Investors | 26 |
| 0.017744 | Trademarks Registered | 27 |
| 0.017498 | Page Views / Visit | 28 |
| 0.017231 | Number of Funding Rounds | 29 |
| 0.015784 | Global Traffic Rank | 30 |
| 0.014576 | Number of Founders | 31 |
| 0.013956 | Bounce Rate | 32 |
| 0.012961 | Number of Apps | 33 |
| 0.012834 | Number of Acquisitions | 34 |
| 0.012820 | Visit Duration | 35 |
| 0.011316 | Industries List | 36 |
| 0.009309 | Downloads Last 30 Days | 37 |
| 0.007847 | Number of Lead Investors | 38 |
| 0.006626 | Operating Status | 39 |
| 0.006072 | Twitter | 40 |
| 0.005959 | Facebook | 41 |
| 0.003858 | LinkedIn | 42 |

مدل جنگل تصادفی به ما نشان می‌دهد که ویژگی‌هایی مانند تعداد کارمندان، مکان دفتر مرکزی و کل سرمایه جذب‌شده (Total Equity Funding Amount) از مهم‌ترین عوامل تأثیرگذار مثبت بر موفقیت هستند.

1. مدل XGboost: الگوریتم XGBoost نسخه‌ای بهینه‌شده و قدرتمند از الگوریتم تقویت گرادیان است که با استفاده از تکنیک‌های منظم‌سازی و پردازش موازی، عملکرد بهتری در شناسایی ویژگی‌های مهم ارائه می‌دهد. در این الگوریتم، اهمیت ویژگی‌ها با محاسبه میانگین تأثیر هر ویژگی بر کاهش تابع زیان در تمام درخت‌ها اندازه‌گیری می‌شود. علاوه بر این، استفاده از منظم‌سازی به جلوگیری از تأثیر بیش‌ازحد ویژگی‌های کم‌اهمیت کمک می‌کند و ویژگی‌های کلیدی با دقت بالاتری شناسایی می‌شوند. (جدول 4-6)

مزایا: سرعت و دقت بالا. قابلیت مدیریت روابط غیرخطی پیچیده.

معایب: پیچیدگی تنظیمات. نیازمند منابع محاسباتی بالا.

**جدول 4- 6- ضرایب ویژگی‌ها بر اساس مدل XGboost**

|  |  |  |
| --- | --- | --- |
| اهمیت ویژگی‌ها | ویژگی‌ها |  |
| 0.113511 | Lower Bound Number of Employees | 1 |
| 0.083072 | Total Equity Funding Amount (in USD) | 2 |
| 0.081033 | Number of Acquisitions | 3 |
| 0.040226 | Downloads Last 30 Days | 4 |
| 0.035397 | Operating Status | 5 |
| 0.034875 | Last Equity Funding Amount (in USD) | 6 |
| 0.032761 | Patents Granted | 7 |
| 0.027536 | Bounce Rate | 8 |
| 0.026852 | Last Funding Amount (in USD) | 9 |
| 0.026155 | Total Products Active | 10 |
| 0.025769 | Number of Investors | 11 |
| 0.023930 | Estimated Revenue High | 12 |
| 0.022765 | Monthly Visits | 13 |
| 0.022442 | Total Funding Amount (in USD) | 14 |
| 0.022077 | Company Age | 15 |
| 0.021517 | Last Funding Type | 16 |
| 0.020520 | Number of Founders | 17 |
| 0.019512 | CB Rank (Organization) | 18 |
| 0.019481 | CB Rank (Company) | 19 |
| 0.019402 | IT Spend (in USD) | 20 |
| 0.019279 | Facebook | 21 |
| 0.018334 | Top 5 Investors | 22 |
| 0.016838 | Headquarters Location | 23 |
| 0.016733 | Last Equity Funding Type | 24 |
| 0.016107 | Page Views / Visit | 25 |
| 0.016022 | Number of Funding Rounds | 26 |
| 0.014758 | Estimated Revenue Low | 27 |
| 0.014720 | Twitter | 28 |
| 0.014406 | Industries | 29 |
| 0.013109 | Global Traffic Rank | 30 |
| 0.012663 | Active Tech Count | 31 |
| 0.012607 | LinkedIn | 32 |
| 0.012538 | Number of Articles | 33 |
| 0.012277 | Number of Apps | 34 |
| 0.012835 | Upper Bound Number of Employees | 35 |
| 0.011861 | Headquarters Regions | 36 |
| 0.011681 | Similar Companies | 37 |
| 0.010702 | Industry Groups | 38 |
| 0.010369 | Visit Duration | 39 |
| 0.010106 | Trademarks Registered | 40 |
| 0.004920 | Number of Lead Investors | 41 |
| 0.000000 | Industries List | 42 |

الگوریتم XGboost به ما نشان می‌دهد که ویژگی‌هایی مانند تعداد کارمندان، کل سرمایه جذب‌شده (Total Equity Funding Amount) و تعداد ادغام و تملیک‌ها از مهم‌ترین عوامل تأثیرگذار مثبت بر موفقیت هستند.

1. **مدل تقویت گرادیان:** الگوریتم تقویت گرادیان مجموعه‌ای از مدل‌های ضعیف[[60]](#footnote-60) را، که معمولاً درخت‌های تصمیم کوچک هستند، به‌صورت ترتیبی آموزش می‌دهد. در هر مرحله، مدل جدید تلاش می‌کند تا خطاهای پیش‌بینی مدل قبلی را کاهش دهد. اهمیت ویژگی‌ها در این الگوریتم بر اساس تأثیر تجمعی هر ویژگی در کاهش تابع زیان[[61]](#footnote-61) طی تمامی مراحل محاسبه می‌شود. ویژگی‌هایی که سهم بیشتری در کاهش خطا دارند، اهمیت بیشتری دریافت می‌کنند. این رویکرد به دلیل استفاده از چندین مدل ترتیبی و تمرکز بر خطاها، برای شناسایی ویژگی‌های کلیدی بسیار مؤثر است. (جدول 4-7)

**مزایا:**

* دقت بالا در پیش‌بینی.
* توانایی یادگیری روابط پیچیده میان متغیرها.

**معایب:**

* حساس به مقادیر پرت
* نیازمند تنظیمات دقیق (مانند نرخ یادگیری و تعداد درخت‌ها).

**جدول 4- 7- ضرایب ویژگی‌ها بر اساس مدل تقویت گرادیان**

|  |  |  |
| --- | --- | --- |
| اهمیت ویژگی‌ها | ویژگی‌ها |  |
| 0.085439 | Upper Bound Number of Employees | 1 |
| 0.071773 | Lower Bound Number of Employees | 2 |
| 0.070031 | Total Equity Funding Amount (in USD) | 3 |
| 0.047516 | Company Age | 4 |
| 0.04043 | Active Tech Count | 5 |
| 0.039008 | CB Rank (Organization) | 6 |
| 0.038395 | Total Funding Amount (in USD) | 7 |
| 0.036423 | CB Rank (Company) | 8 |
| 0.035009 | Patents Granted | 9 |
| 0.034946 | Headquarters Location | 10 |
| 0.03402 | Last Funding Type | 11 |
| 0.033416 | IT Spend (in USD) | 12 |
| 0.032514 | Headquarters Regions | 13 |
| 0.030804 | Last Funding Amount (in USD) | 14 |
| 0.026986 | Number of Acquisitions | 15 |
| 0.026763 | Last Equity Funding Type | 16 |
| 0.025985 | Number of Articles | 17 |
| 0.025596 | Trademarks Registered | 18 |
| 0.025233 | Industry Groups | 19 |
| 0.023646 | Industries | 20 |
| 0.020925 | Estimated Revenue Low | 21 |
| 0.019725 | Monthly Visits | 22 |
| 0.019653 | Total Products Active | 23 |
| 0.019452 | Top 5 Investors | 24 |
| 0.018392 | Global Traffic Rank | 25 |
| 0.016494 | Number of Founders | 26 |
| 0.015763 | Estimated Revenue High | 27 |
| 0.015483 | Visit Duration | 28 |
| 0.013677 | Number of Funding Rounds | 29 |
| 0.012158 | Downloads Last 30 Days | 30 |
| 0.010817 | Page Views / Visit | 31 |
| 0.01015 | Last Equity Funding Amount (in USD) | 32 |
| 0.005825 | Similar Companies | 33 |
| 0.005731 | Operating Status | 34 |
| 0.004079 | Number of Investors | 35 |
| 0.00241 | LinkedIn | 36 |
| 0.002242 | Bounce Rate | 37 |
| 0.001499 | Number of Apps | 38 |
| 0.001313 | Industries List | 39 |
| 0.00028 | Facebook | 40 |
| 0 | Number of Lead Investors | 41 |
| 0 | Twitter | 42 |

مدل تقویت گرادیان به ما نشان می‌دهد که ویژگی‌هایی مانند تعداد کارمندان،کل سرمایه جذب‌شده (Total Equity Funding Amount) و سن شرکت از مهم‌ترین عوامل تأثیرگذار مثبت بر موفقیت هستند.

با توجه به الگوریتم‌ها استفاده شده و همچنین میران همبستگی هریک از متغیرها با موفقیت، در نهایت دو ویژگی تعداد سرمایه گذاران (Number of investors ) و مقدار کل سرمایه جذب شده (Total equity funding amount) به عنوان ویژگی‌های با اثر بالاتر انتخاب شده و 5 ویژگی دیگر حذف شدند.

## آمار توصیفی

در این بخش، به ارائه آمار توصیفی داده‌های جمع‌آوری‌شده برای تحلیل و مدل‌سازی موفقیت شرکت‌های هوش مصنوعی فعال در صنعت حمل و نقل می‌پردازیم. آمار توصیفی ابزارهایی ارزشمند برای شناخت بهتر ویژگی‌ها و الگوهای موجود در داده‌ها فراهم می‌آورد و پایه‌ای مناسب برای تحلیل‌های پیشرفته‌تر و مدل‌سازی‌های آماری فراهم می‌کند. با استفاده از آمار توصیفی، می‌توان به بررسی ویژگی‌های مرکزی، پراکندگی و توزیع متغیرهای مختلف پرداخت و درک بهتری از ساختار داده‌ها به دست آورد. داده‌های مورد استفاده در این تحقیق شامل اطلاعات مربوط به ۴۱۳۲ شرکت فعال در زمینه هوش مصنوعی در صنعت حمل و نقل می‌باشد. این داده‌ها از وبسایت کرانچ بیس استخراج شده و شامل 37 ویژگی مختلف هستند که در جدول 4-8 آورده شده‌اند.

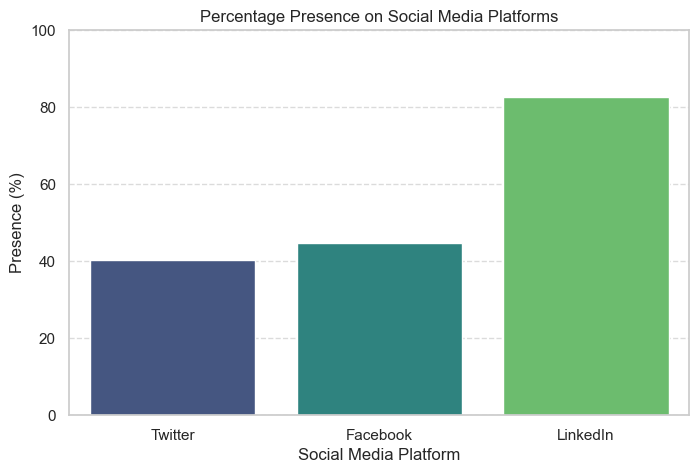
برای تحلیل توصیفی، ابتدا به بررسی شاخص‌های مرکزی و پراکندگی برای هر یک از ویژگی‌های عددی پرداخته می‌شود. این شاخص‌ها شامل میانگین، میانه، مد، انحراف معیار، کمینه، بیشینه و دامنه تغییرات هستند، که این ویژگی‌ها در جدول 4-8 آورده شده‌اند. همچنین، با استفاده از نمودارهای گرافیکی نظیر هیستوگرام‌ها و نمودارهای جعبه‌ای، توزیع هر یک از متغیرها به‌صورت بصری نمایش داده می‌شود تا بتوان نقاط پرت و ناهنجاری‌های موجود در داده‌ها را شناسایی کرد.

جدول 4-8- آمار توصیفی

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Feature | Count | Mean | Std | Min | 25% | 50% | 75% | Max |
| Industries | 1067 | 454.71 | 260.05 | 0 | 224.5 | 463 | 681.5 | 885 |
| Headquarters Location | 1067 | 284.56 | 156.41 | 0 | 149.5 | 299 | 422 | 547 |
| CB Rank (Company) | 1067 | 533 | 308.16 | 0 | 266.5 | 533 | 799.5 | 1066 |
| Headquarters Regions | 1067 | 14.74 | 14.54 | 0 | 0 | 12 | 24 | 49 |
| Operating Status | 1067 | 0.043 | 0.203 | 0 | 0 | 0 | 0 | 1 |
| Twitter | 1067 | 0.402 | 0.491 | 0 | 0 | 0 | 1 | 1 |
| Facebook | 1067 | 0.447 | 0.497 | 0 | 0 | 0 | 1 | 1 |
| LinkedIn | 1067 | 0.826 | 0.38 | 0 | 1 | 1 | 1 | 1 |
| Number of Articles | 1067 | 11 | 20.65 | 0 | 0 | 0 | 11.5 | 72 |
| Industry Groups | 1067 | 305.22 | 148.49 | 0 | 162.5 | 338 | 452 | 476 |
| Trademarks Registered | 1067 | 1.49 | 4.14 | 0 | 1 | 1 | 1 | 95 |
| IT Spend (in USD) | 1067 | 1102926 | 20049207 | 8333 | 237426 | 237426 | 237426 | 6.5E+08 |
| Last Funding Type | 1067 | 10.94 | 4.52 | 0 | 10 | 12 | 12 | 19 |
| Last Equity Funding Type | 1067 | 7.33 | 3.42 | 0 | 5 | 7 | 8 | 15 |
| Company Age | 1067 | 13.39 | 16.24 | 0 | 5.5 | 8 | 14 | 131 |
| Lower Bound Number of Employees | 1067 | 132.96 | 777.77 | 1 | 1 | 1 | 51 | 10001 |
| Upper Bound Number of Employees | 1067 | 324.08 | 1237.58 | 10 | 10 | 50 | 100 | 10001 |
| Estimated Revenue Low | 1067 | 14708716 | 73776000 | 0 | 800000 | 1000000 | 2800000 | 1E+09 |
| Estimated Revenue High | 1067 | 95388379 | 6.92E+08 | 1000000 | 8200000 | 10000000 | 18000000 | 1E+10 |
| Excluded Industries | 1067 | 428.03 | 218.59 | 0 | 239.5 | 466 | 641.5 | 694 |

### حضور فعال در شبکه‌های اجتماعی

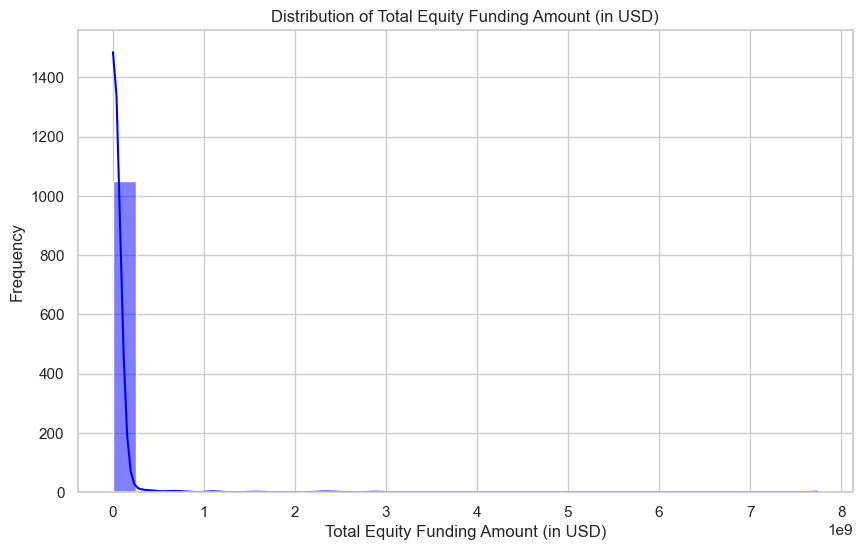
حدود 40 درصد از شرکت‌ها حساب فعال در توییتر و 45 درصد در فیس‌بوک دارند. لینکدین، با نرخ حضور 83 درصدی، بیشترین استفاده را در میان شرکت‌ها دارد. این موضوع می‌تواند نشان‌دهنده اهمیت شبکه‌های حرفه‌ای برای شرکت‌های فعال در این حوزه باشد. لینکدین به‌ویژه برای شرکت‌هایی که در مراحل اولیه رشد هستند یا به دنبال فرصت‌های شراکت و همکاری هستند، اهمیت زیادی دارد. این تفاوت در نرخ حضور می‌تواند سرنخی برای تحلیل ارتباط میان استفاده از پلتفرم‌های اجتماعی و موفقیت شرکت‌ها باشد. به‌عنوان مثال، ممکن است شرکت‌هایی که در لینکدین حضور فعال دارند، به دلیل تعاملات بهتر با سرمایه‌گذاران و جامعه تجاری، عملکرد موفق‌تری داشته باشند. همچنین می‌توان بررسی کرد که آیا حضور در فیس‌بوک و توییتر منجر به افزایش آگاهی عمومی از برند و جذب مشتریان می‌شود یا خیر. دلیل اینکه در این صنعت لینکدین از اهمیت بالاتری نسبت به سایر پلتفرم‌ها قرار دارد می‌تواند این باشد که بیشتر شرکتهای این حوزه بیزینس به بیزینس هستند.



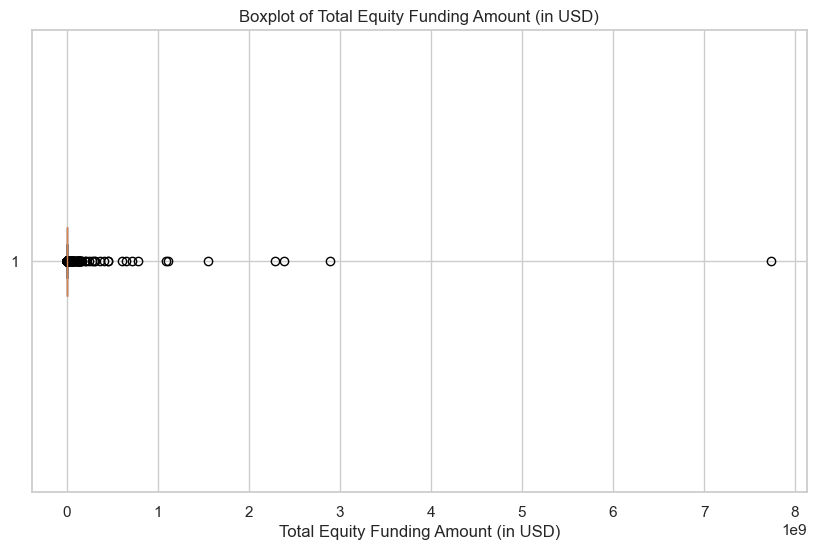
**شکل 4- 3- میزان حضور هریک از شرکت‌ها در شبکه‌های اجتماعی**

### کل سرمایه‌گذاری سهام

از آنجایی که صنعت تأمین مالی سهام معمولاً شامل شرکت‌هایی با مقیاس بسیار متفاوت (از استارتاپ‌های کوچک تا غول‌های فناوری) است، داده‌های پرت در این ویژگی وجود دارد (شکل 4-4 و 4-5).



**شکل 4- 4- نمودار هیستوگرام کل سرمایه‌گذاری سهام**

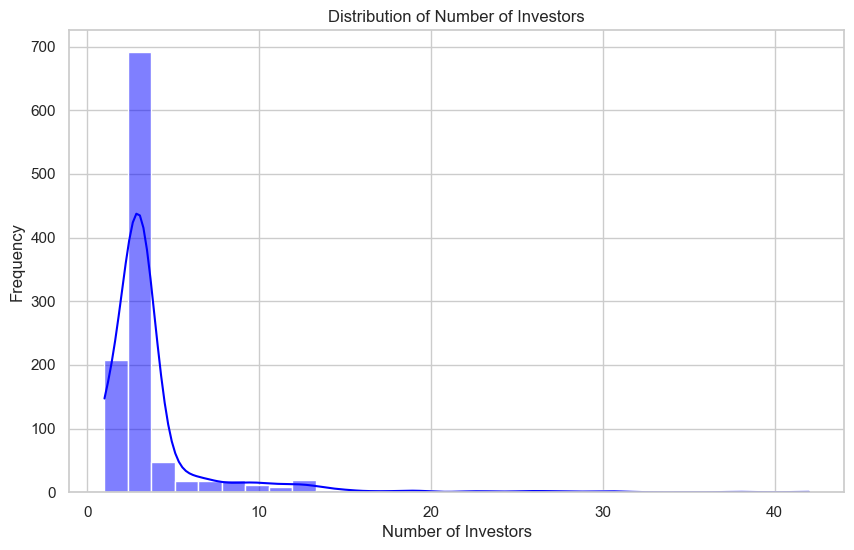


**شکل 4- 5- نمودارجعبه‌ای کل سرمایه‌گذاری سهام**

### تعداد سرمایه‌گذاران

نمودار هیستوگرام (شکل4-6 )نشان می‌دهد که توزیع تعداد سرمایه‌گذاران به شدت به سمت چپ منحرف شده است، به این معنا که بیشتر شرکت‌ها تعداد کمی سرمایه‌گذار دارند. محور افقی این نمودار نشان می‌دهد که بخش عمده‌ای از شرکت‌ها کمتر از پنج سرمایه‌گذار دارند و تنها بخش کوچکی از داده‌ها شامل تعداد بیشتری سرمایه‌گذار است. این مسئله به‌صورت یک دم بلند در سمت راست نمودار قابل مشاهده است. از این نمودار می‌توان نتیجه گرفت که سرمایه‌گذاری‌ها معمولاً محدود و متمرکز هستند و تنها تعداد کمی از شرکت‌ها از سرمایه‌گذاران بیشتری برخوردارند.

نمودار جعبه‌ای (شکل 4-7) تعداد سرمایه‌گذاران نیز این موضوع را تأیید می‌کند. میانه داده‌ها نشان می‌دهد که نیمی از شرکت‌ها کمتر از دو سرمایه‌گذار دارند. بازه‌ای که ۵۰ درصد از داده‌ها را شامل می‌شود (بین چارک اول و سوم) در نزدیکی صفر تا پنج سرمایه‌گذار قرار دارد. وجود نقاط پرت در این نمودار نشان‌دهنده شرکت‌هایی است که تعداد غیرمعمولی زیادی از سرمایه‌گذاران دارند. این نقاط پرت می‌توانند اطلاعاتی درباره ویژگی‌های خاص این شرکت‌ها ارائه دهند؛ مثلاً اینکه این شرکت‌ها ممکن است در مراحل رشد پیشرفته‌تر یا موفق‌تر باشند. به‌طور کلی، این دو نمودار نشان می‌دهند که بیشتر شرکت‌ها با تعداد سرمایه‌گذاران محدود عمل می‌کنند، درحالی‌که برخی شرکت‌ها تعداد بیشتری سرمایه‌گذار دارند که به‌عنوان موارد استثنا محسوب می‌شوند.



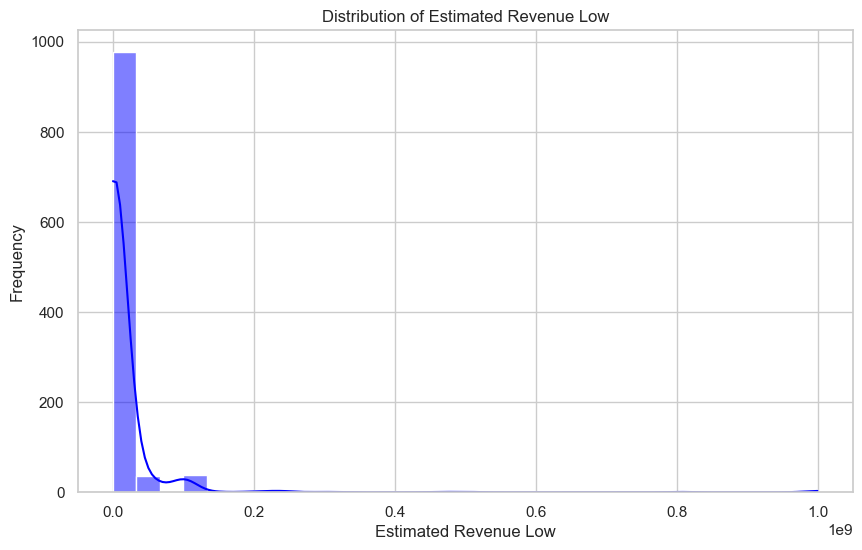
**شکل 4- 6- نمودار هیستوگرام تعداد سرمایه‌گذاران**



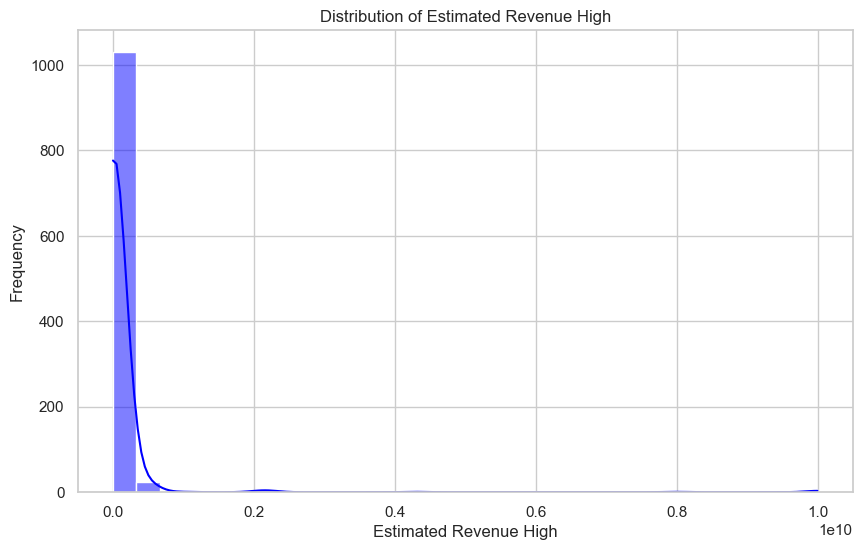
**شکل 4- 7-نمودار جعبه‌ای تعداد سرمایه‌گذاران**

### درآمد تخمینی

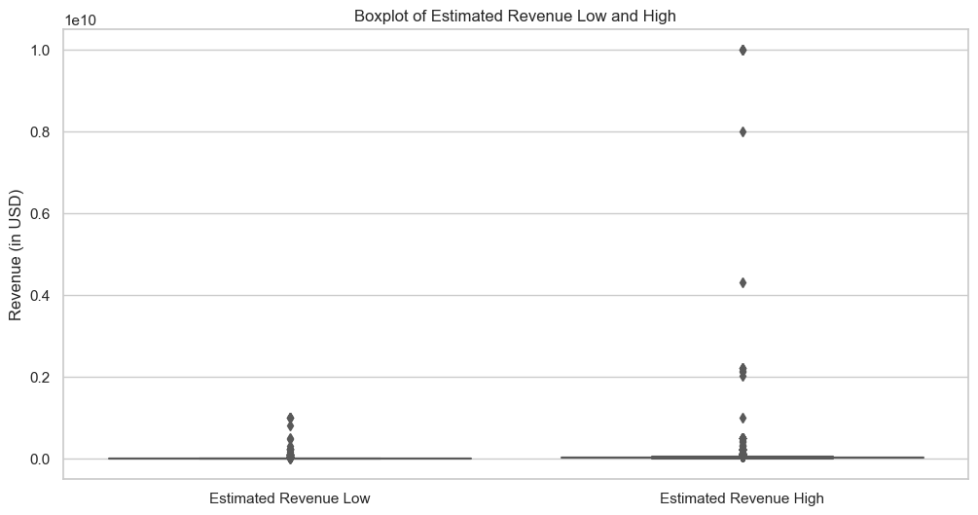
درآمد تخمینی شرکت‌ها در وبسایت کرانچ بیس به شکل بازه‌ای از درآمد داده شده است. برای تحلیل بهتر درآمد هرکدام از شرکت‌ها به صورت حداقل و حداکثر بررسی شده است.



**شکل 4- 8- نمودار هیستوگرام درآمد تخمینی حد پایین**



**شکل 4- 9- نمودار هیستوگرام درآمد تخمینی حد بالا**

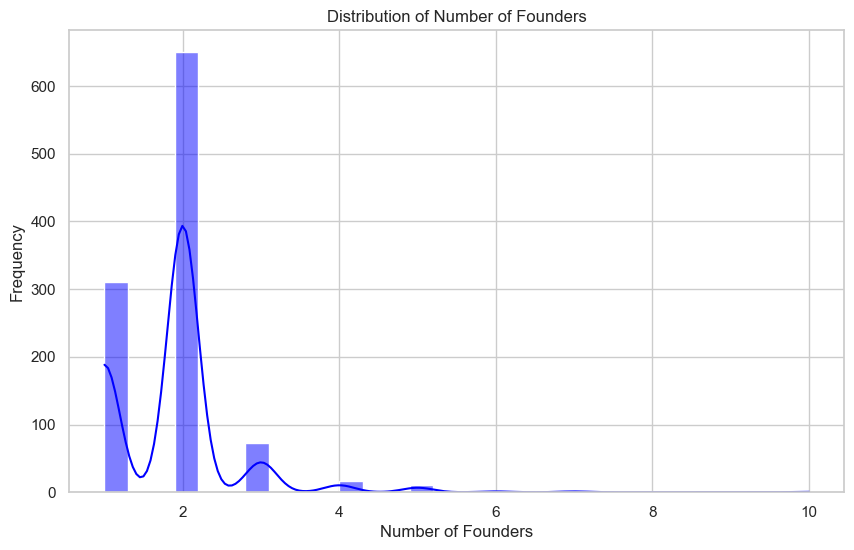


**شکل 4- 10- نمودار جعبه‎ای درآمد تخمینی**

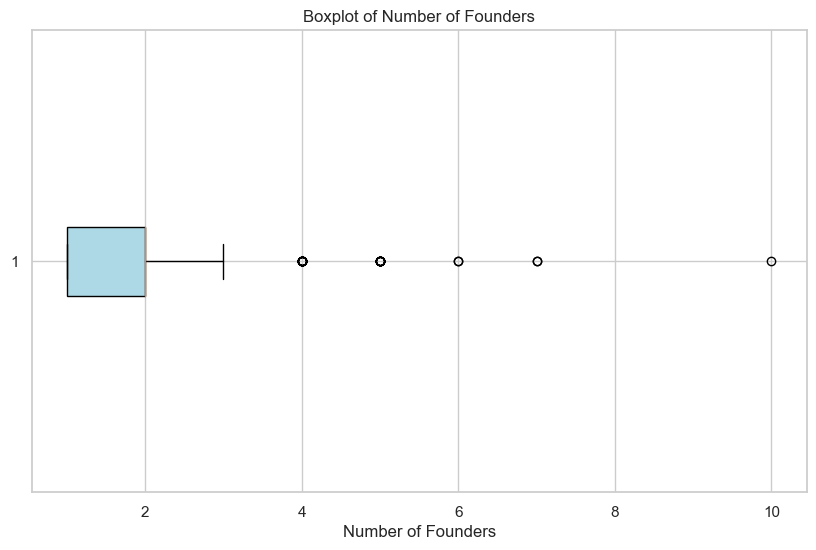
این نمودارها (شکل 4-8 تا 4-10) می‌توانند نشان‌دهنده‌ی وجود نابرابری زیاد در درآمد تخمینی شرکت‌ها باشند، که می‌تواند ناشی از اندازه، صنعت، یا سایر عوامل مرتبط با شرکت‌ها باشد.

### تعداد بنیان‌گذاران

توزیع داده‌ها نشان‌دهنده این است (شکل 4-11 و 4-12) که ساختار معمول شرکت‌ها شامل تعداد کمی مؤسس است (اغلب 2 نفر). شرکت‌هایی که تعداد مؤسسان بیشتری دارند، نسبت به سایرین غیرعادی‌تر هستند و در دسته‌بندی نقاط پرت قرار می‌گیرند. این الگو می‌تواند نشان‌دهنده‌ی این باشد که تشکیل شرکت‌ها اغلب نیازمند هماهنگی کمتر و تیم‌های کوچک‌تر است.



**شکل 4- 11- نمودار هیستوگرام تعداد بنیان‌گذاران**



**شکل 4- 12- نمودار جعبه‌ای تعداد بنیان‌گذاران**

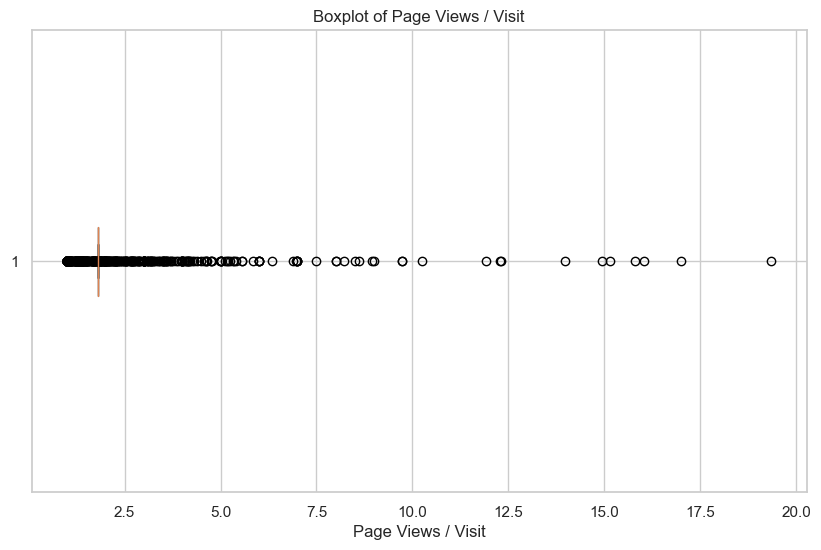
### بازدید از صفحه

نمودار جعبه‌ای (شکل 4-14) نشان می‌دهد که بیشتر داده‌ها در نزدیکی بازه 2 تا 3 قرار دارند، و مقادیر پرت به طور قابل توجهی در سمت راست (بالاتر از 5) وجود دارند. این به معنای آن است که در بیشتر بازدیدها تعداد صفحات مشاهده شده به طور متوسط پایین است، اما تعدادی از بازدیدها به طور استثنایی صفحات بیشتری را شامل می‌شوند.

در نمودار هیستوگرام (شکل 4-13)، توزیع به وضوح چولگی به راست دارد و اکثر داده‌ها در بازه 2 تا 3 تجمع یافته‌اند و با افزایش مقدار بازدید از صفحه، تعداد داده‌ها به شدت کاهش می‌یابد. این نشان می‌دهد که بیشتر کاربران تعداد صفحات کمی را بازدید می‌کنند و تعداد کمی از کاربران رفتار استثنایی از خود نشان می‌دهند (بازدید از تعداد زیادی صفحه).



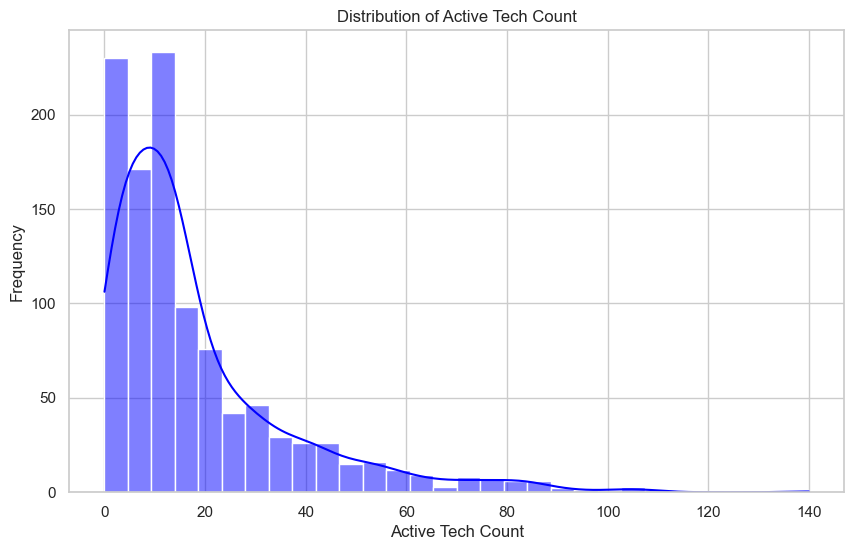
**شکل 4- 13- نمودار هیستوگرام بازدید از صفحه**



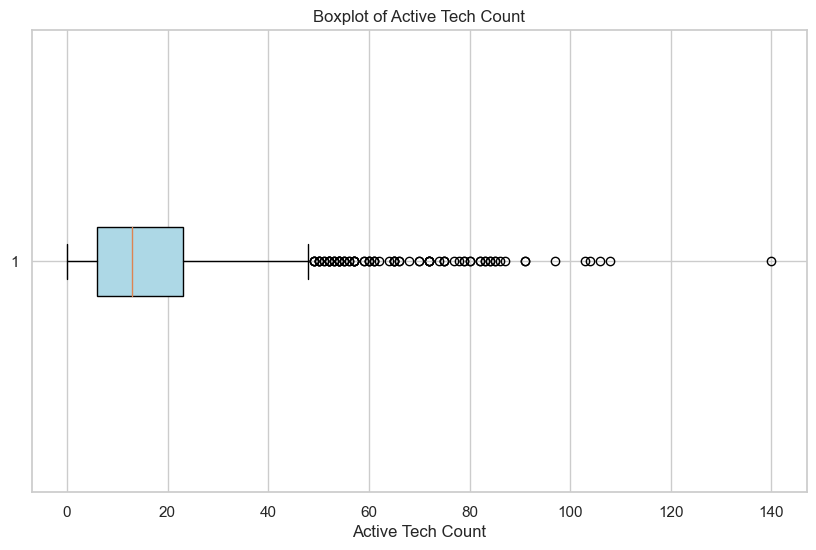
**شکل 4- 14- نمودار جعبه‌ای بازدید از صفحه**

### تعداد فناوری‌های فعال

نمودار جعبه‌ای و هیستوگرام (شکل 4-15 و 4-16) تعداد تکنولوژی‌های فعال نشان‌دهنده توزیع غیرمتقارن داده‌ها با وجود تعداد قابل توجهی نقاط پرت است. در نمودار جعبه‌ای، محدوده بیشتر داده‌ها (حدود 50 درصد) بین مقادیر نزدیک به صفر و حدود 20 قرار دارد و مقادیر بالاتر از این محدوده به عنوان نقاط پرت مشخص شده‌اند. این موضوع در هیستوگرام نیز مشهود است، جایی که بیشتر شرکت‌ها تعداد تکنولوژی‌های فعال کمی دارند و تعداد کمتری از شرکت‌ها دارای تکنولوژی‌های فعال بالایی هستند. شکل هیستوگرام که یک توزیع دم‌دار به سمت راست را نشان می‌دهد، بیانگر این است که تعداد معدودی از شرکت‌ها دارای تکنولوژی‌های فعال بسیار زیاد هستند. به‌طور کلی، این داده‌ها نشان می‌دهند که اکثر شرکت‌ها به تعداد محدودی از تکنولوژی‌های فعال متکی هستند، در حالی که تنها تعداد اندکی از آن‌ها طیف گسترده‌ای از تکنولوژی‌ها را به‌کار گرفته‌اند.



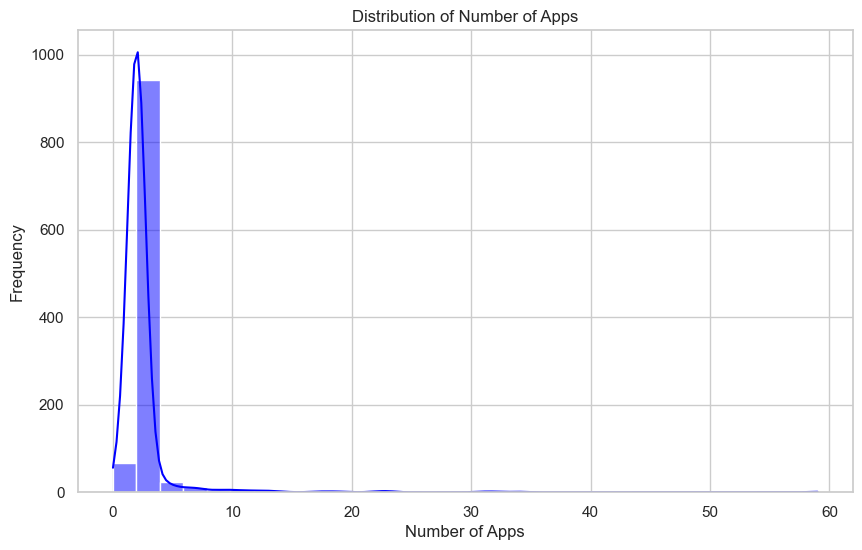
**شکل 4- 15- نمودار هیستوگرام تعداد فناوری‌های فعال**



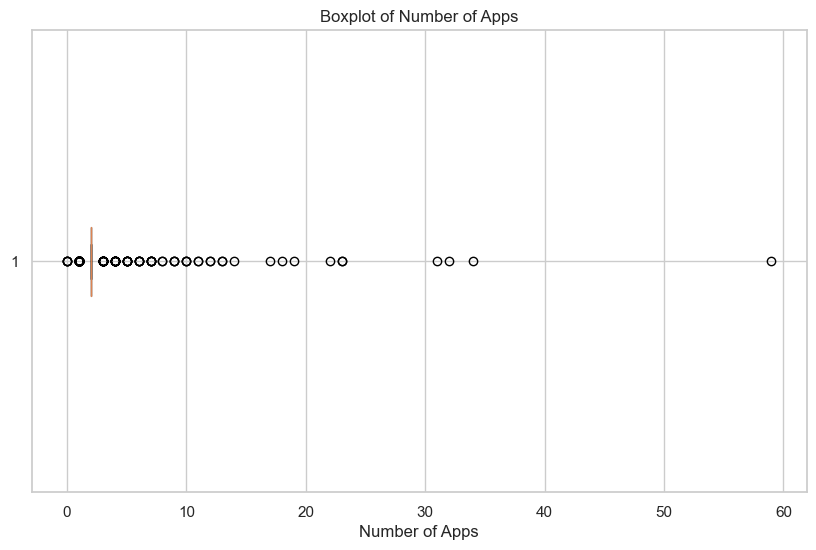
**شکل 4- 16- نمودار جعبه‌ای تعداد فناوری‌های فعال**

### تعداد اپلیکیشن‌ها

با توجه به نمودارهای جعبه‌ای و هیستوگرام (شکل 4-17 و 4-18) بیشتر شرکت‌ها تعداد برنامه‌های کمی (1تا2) دارند و فقط تعداد معدودی از شرکت‌ها تعداد برنامه‌های بسیار زیادی دارند که به‌عنوان نقاط پرت قابل‌شناسایی هستند. این توزیع می‌تواند ناشی از تفاوت در مقیاس یا حوزه فعالیت شرکت‌ها باشد.



**شکل 4- 17- نمودار هیستوگرام تعداد اپلیکیشن‌ها**



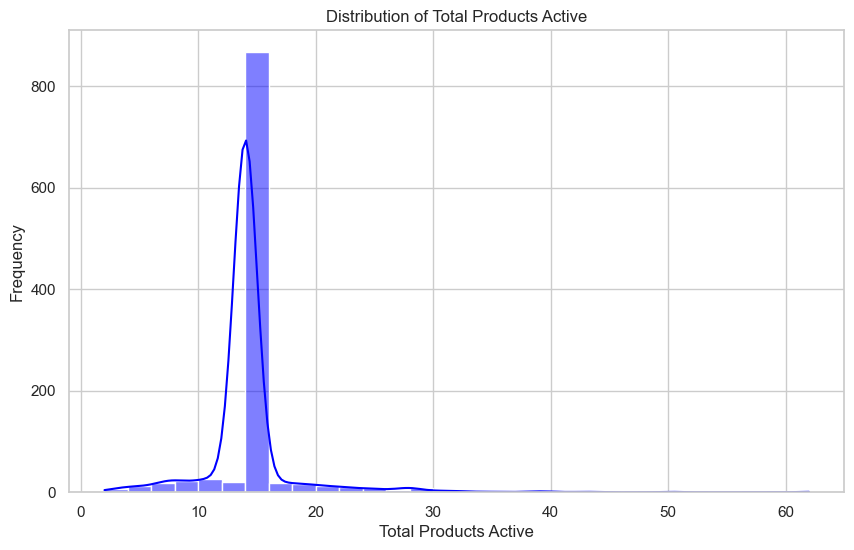
**شکل 4- 18- نمودار جعبه‌ای تعداد اپلیکیشن‌ها**

### تعداد محصولات فعال

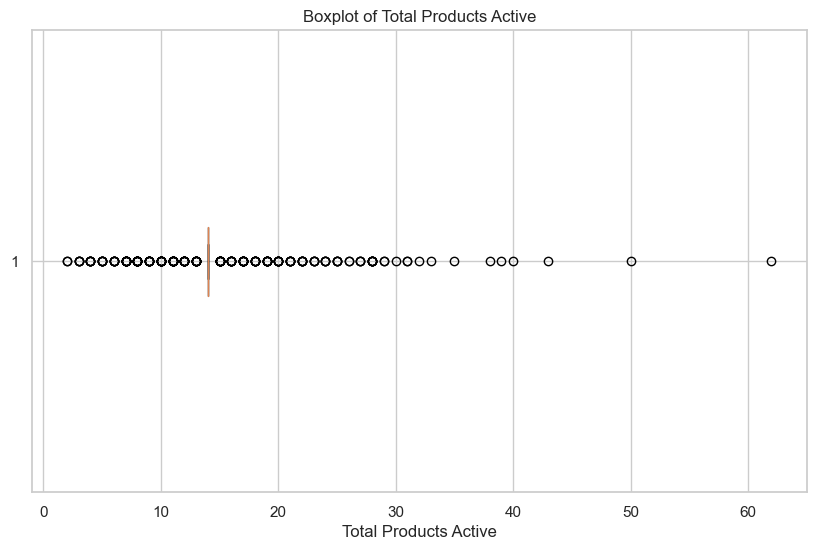
نمودار هیستوگرام (شکل 4-19) نشان می‌دهد که بیشتر شرکت‌ها تعداد محصولات فعالی در حدود بازه 10 تا 15 دارند. با توجه به شکل منحنی و تراکم داده‌ها، توزیع به شدت به سمت چپ تمایل دارد و توزیعی راست‌چوله است، که بیانگر تعداد زیادی از شرکت‌ها با تعداد محصولات فعال کم و تعداد کمی شرکت با محصولات فعال بسیار زیاد است.

نمودار جعبه‌ای (شکل 4-20) این موضوع را تأیید می‌کند که بخش زیادی از داده‌ها در بازه 10 تا 15 قرار دارند و مقادیر پرت نیز به وضوح در سمت راست مشاهده می‌شوند. این مقادیر پرت نشان‌دهنده شرکت‌هایی با تعداد محصولات فعال بسیار بیشتر از میانگین هستند. این موضوع ممکن است به دلیل تفاوت در نوع فعالیت شرکت‌ها یا استراتژی‌های مختلف آن‌ها برای گسترش محصولات باشد.

در مجموع، این نمودارها نشان می‌دهند که شرکت‌ها عموماً تعداد محدودی محصول فعال دارند و تنها تعداد کمی از شرکت‌ها از نظر تعداد محصولات فعال برجسته هستند.



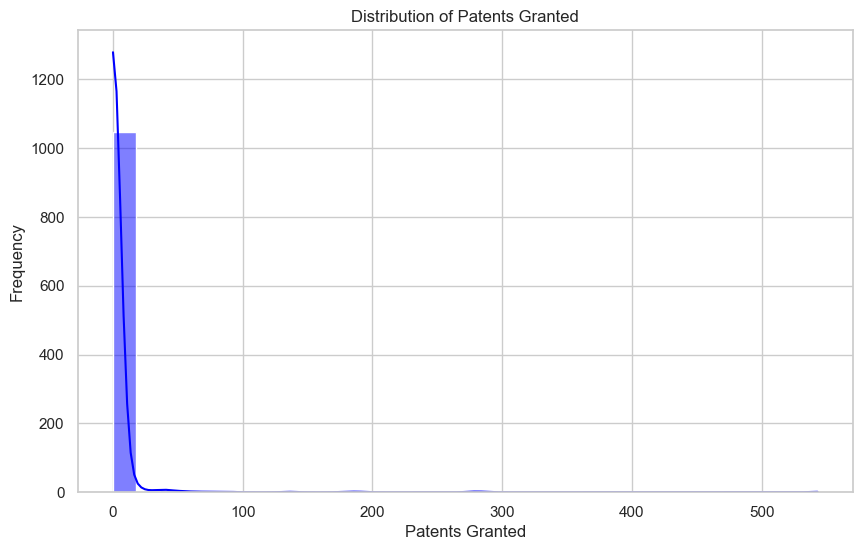
**شکل 4- 19- نمودار هیستوگرام محصولات فعال**



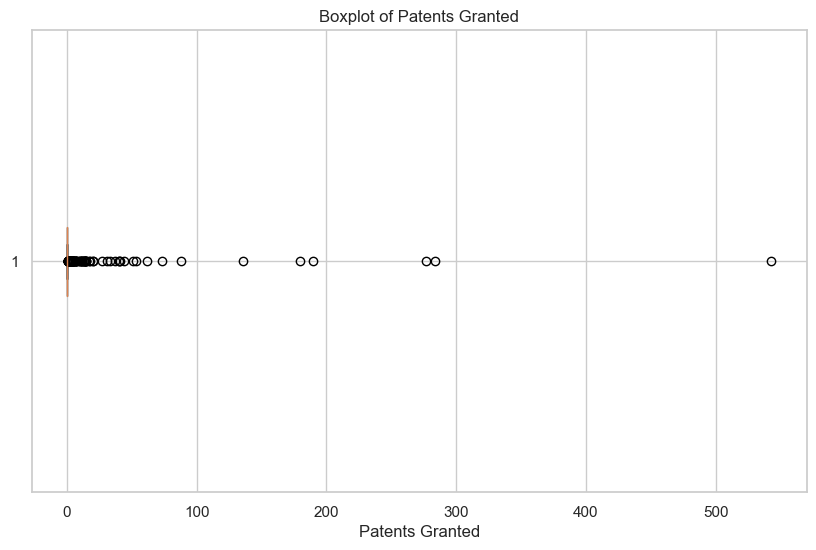
**شکل 4- 20- نمودار جعبه‌ای محصولات فعال**

### تعداد پتنت‌های ثبت شده

نمودار‌های زیر (شکل 4-21 و 4-22) نشان می‌دهند که پتنت‌های ثبت‌شده در میان شرکت‌ها به صورت نامتقارن توزیع شده‌اند و تعداد زیادی از شرکت‌ها یا هیچ پتنتی ندارند یا تعداد کمی پتنت ثبت کرده‌اند، در حالی که تعداد معدودی از شرکت‌ها دارای تعداد بسیار زیادی پتنت هستند. این موضوع می‌تواند بیانگر تمرکز نوآوری و تحقیق و توسعه در تعداد کمی از شرکت‌ها باشد.



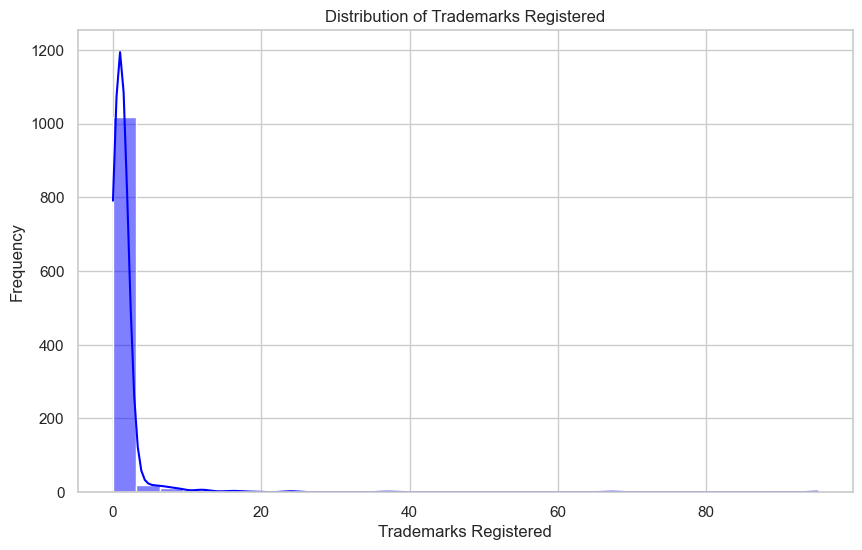
**شکل 4- 21- نمودار هیستوگرام پتنت‌های ثبت شده**



**شکل 4- 22- نمودار جعبه‌ای پتنت‌های ثبت شده**

### تعداد علائم تجاری ثبت شده

نمودارهای جعبه‌ای و هیستوگرام (شکل 4-23 و شکل 4-24) برای تعداد علائم تجاری ثبت‌شده نشان می‌دهند که بیشتر شرکت‌ها تعداد کمی علامت تجاری ثبت کرده‌اند، در حالی که تعداد محدودی از شرکت‌ها تعداد بسیار بالایی دارند که به‌عنوان مقادیر پرت در نمودار جعبه‌ای مشخص شده است. توزیع داده‌ها در هیستوگرام چوله به سمت راست است، به این معنا که تراکم بالایی از داده‌ها در مقادیر پایین وجود دارد و مقادیر بالا به‌صورت پراکنده مشاهده می‌شوند. فاصله کوچک بین چارک اول و سوم در نمودار جعبه‌ای نشان‌دهنده تراکم داده‌ها در محدوده مقادیر پایین است، اما کشیدگی توزیع به دلیل وجود مقادیر پرت در داده‌ها مشاهده می‌شود. این وضعیت نشان می‌دهد که اکثریت شرکت‌ها علائم تجاری محدودی دارند، اما تعداد کمی از آن‌ها با تعداد بسیار بالا، توزیع کلی داده‌ها را تحت تأثیر قرار داده‌اند.



**شکل 4- 23- نمودار هیستوگرام تعداد علائم تجاری ثبت شده**



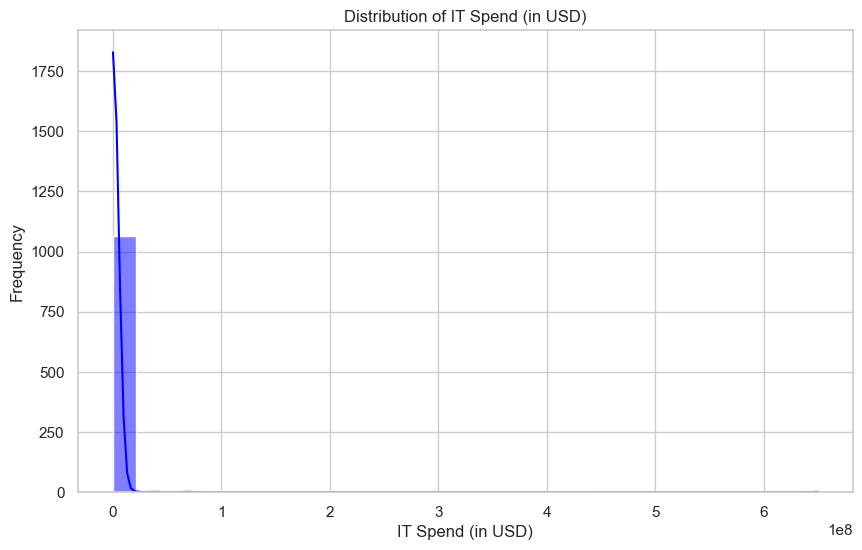
**شکل 4- 24- نمودار جعبه‌ای تعداد علائم تجاری ثبت شده**

### هزینه‌های فناوری اطلاعات

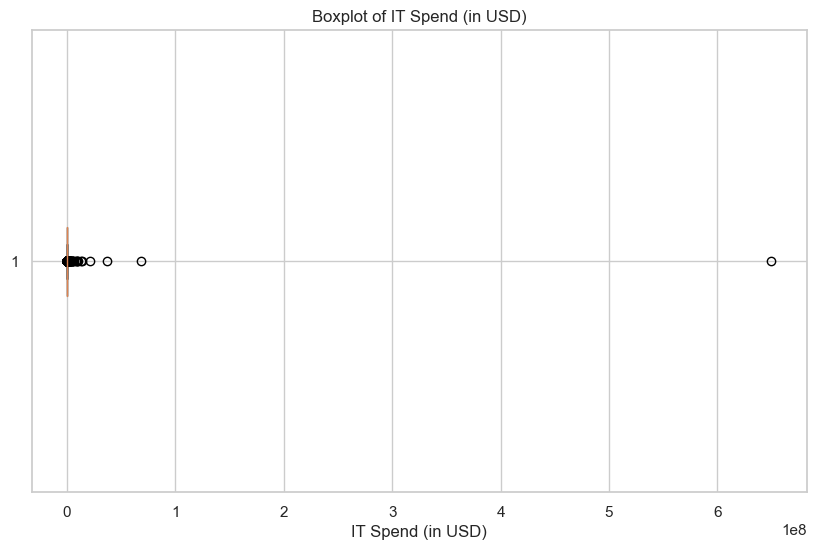
در نمودار هیستوگرام (شکل 4-25)، توزیع مقادیر به شدت چولگی به سمت راست دارد. اکثر داده‌ها در مقادیر پایین متمرکز هستند و نشان می‌دهد که اکثر شرکت‌ها هزینه‌ای نسبتاً کم برای فناوری اطلاعات دارند. این موضوع نشان می‌دهد که میانگین هزینه‌ها ممکن است توسط مقادیر بسیار بالایی که به عنوان داده‌های پرت در نظر گرفته می‌شوند، به شدت تحت تأثیر قرار گیرد.

در نمودار جعبه‌ای (شکل4-26)، بخش اعظم داده‌ها در نزدیکی صفر قرار دارند و مقادیر پرت مشخصی وجود دارد که به طور قابل توجهی از محدوده میان چارک‌ها فاصله دارند. این داده‌های پرت نشان‌دهنده شرکت‌هایی هستند که هزینه‌های بسیار بالایی در زمینه فناوری اطلاعات دارند و می‌توانند جزو شرکت‌های بزرگ یا دارای پروژه‌های فناوری اطلاعات پیشرفته باشند.

به طور کلی، این توزیع نشان‌دهنده عدم توازن شدید در هزینه‌های IT در بین شرکت‌ها است، به طوری که تعداد محدودی از شرکت‌ها هزینه‌های بسیار بالایی دارند، در حالی که اکثریت هزینه‌های کم یا متوسطی دارند.



**شکل 4- 25- نمودار هیستوگرام هزینه فناوری اطلاعات**

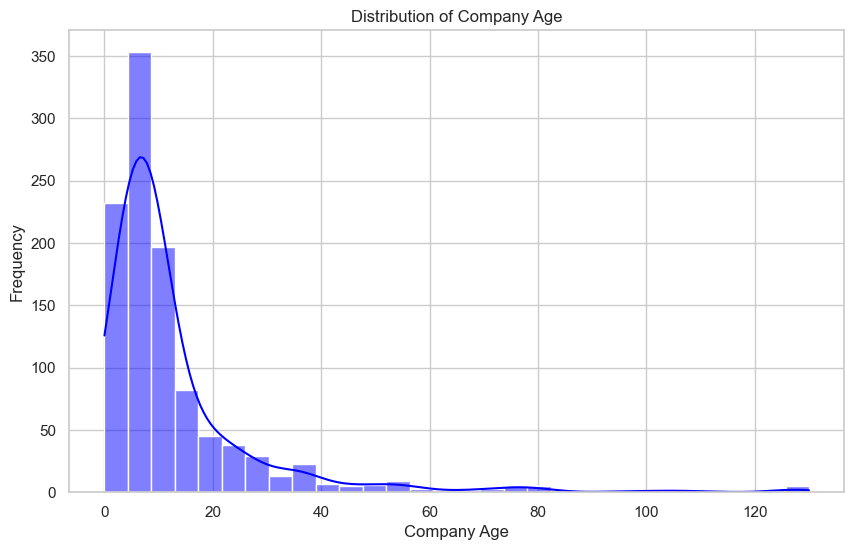


**شکل 4- 26- نمودار جعبه‌ای هزینه فناوری اطلاعات**

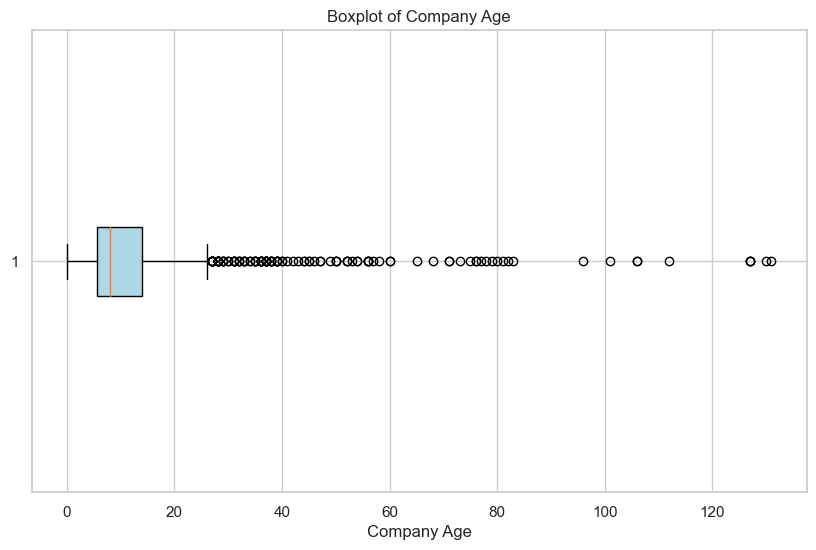
### سن شرکت

نمودار هیستوگرام (شکل4-27) نشان می‌دهد که بخش عمده‌ای از شرکت‌ها در گروه‌های سنی پایین قرار دارند، به‌ویژه در بازه‌های 0 تا 10 سال، و با افزایش سن، تعداد شرکت‌ها به‌شدت کاهش می‌یابد. این نشان‌دهنده این است که بیشتر شرکت‌های موجود در مجموعه داده جدید هستند یا سن کمتری دارند.

نمودار جعبه‌ای (شکل4-28) نشان می‌دهد که میانه سن شرکت‌ها کمتر از 10 سال است. طول جعبه (بین چارک اول و سوم) نشان‌دهنده تمرکز داده‌ها در بازه سنی کم است. همچنین، مقادیر پرت که به‌صورت نقاطی خارج از خطوط افقی جعبه نشان داده شده‌اند، به شرکت‌هایی با سن بالا (بیش از 40 سال) اشاره دارند که تعداد آن‌ها کم است اما وجود دارند. برخی از این شرکت‌ها حتی بیش از 100 سال سن دارند که در نقاط پرت قابل مشاهده هستند.



**شکل 4- 27- نمودار هیستوگرام سن شرکت**



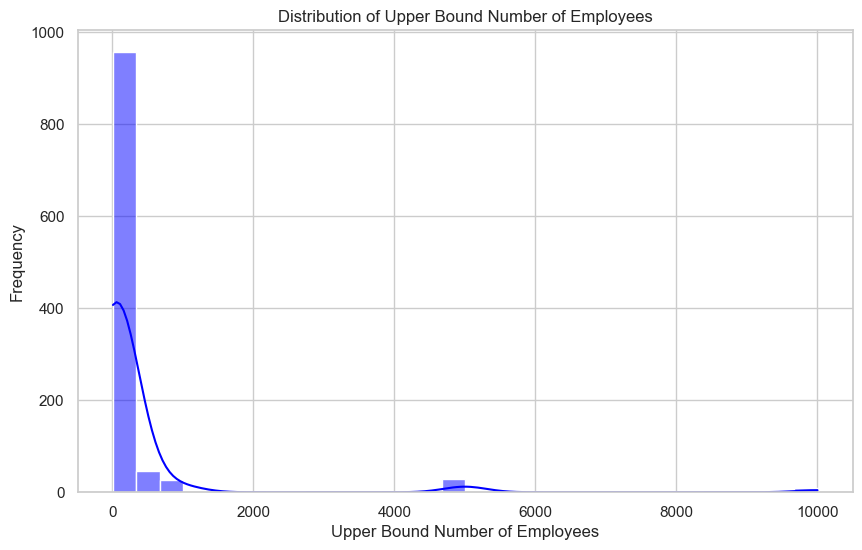
**شکل 4- 28- نمودار جعبه‌ای سن شرکت**

### تعداد کارکنان

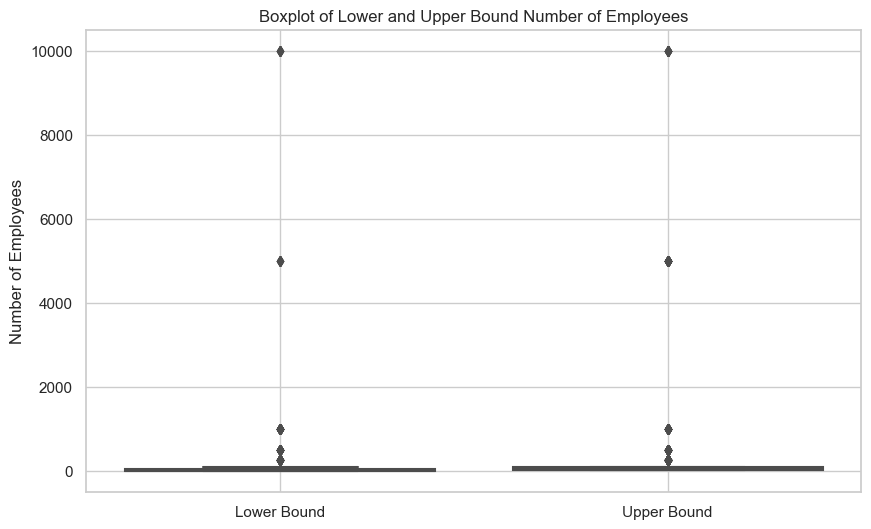
تحلیل توزیع ویژگی‌های «حد پایین تعداد کارکنان» و «حد بالای تعداد کارکنان» (شکل 4-29 تا شکل 4-31 ) نشان می‌دهد که هر دو ویژگی توزیعی با چولگی به سمت راست دارند، به طوری که بیشتر داده‌ها در مقادیر پایین‌تر متمرکز شده‌اند و تنها تعداد کمی از داده‌ها در مقادیر بالاتر قرار گرفته‌اند. این الگو بیانگر آن است که اکثر شرکت‌ها تعداد کارکنان کمتری دارند و فقط تعداد محدودی از آن‌ها در بازه‌های بالاتر جای می‌گیرند. نمودار جعبه‌ای نیز وجود مقادیر پرت قابل توجهی را تأیید می‌کند، به ویژه در مقادیر بیشتر از دو هزار، که نشان‌دهنده حضور شرکت‌هایی با تعداد کارکنان بسیار زیاد است. همچنین گستردگی بیشتر در «حد بالای تعداد کارکنان» نشان‌دهنده تخمین بازه وسیع‌تری در تعداد کارکنان شرکت‌ها است. در مجموع، این نتایج بر تمرکز داده‌ها بر شرکت‌های کوچک‌تر و تعداد اندکی شرکت‌های بزرگ‌تر دلالت دارد.



**شکل 4- 29- نمودار هیستوگرام تعداد کارکنان حد پایین**



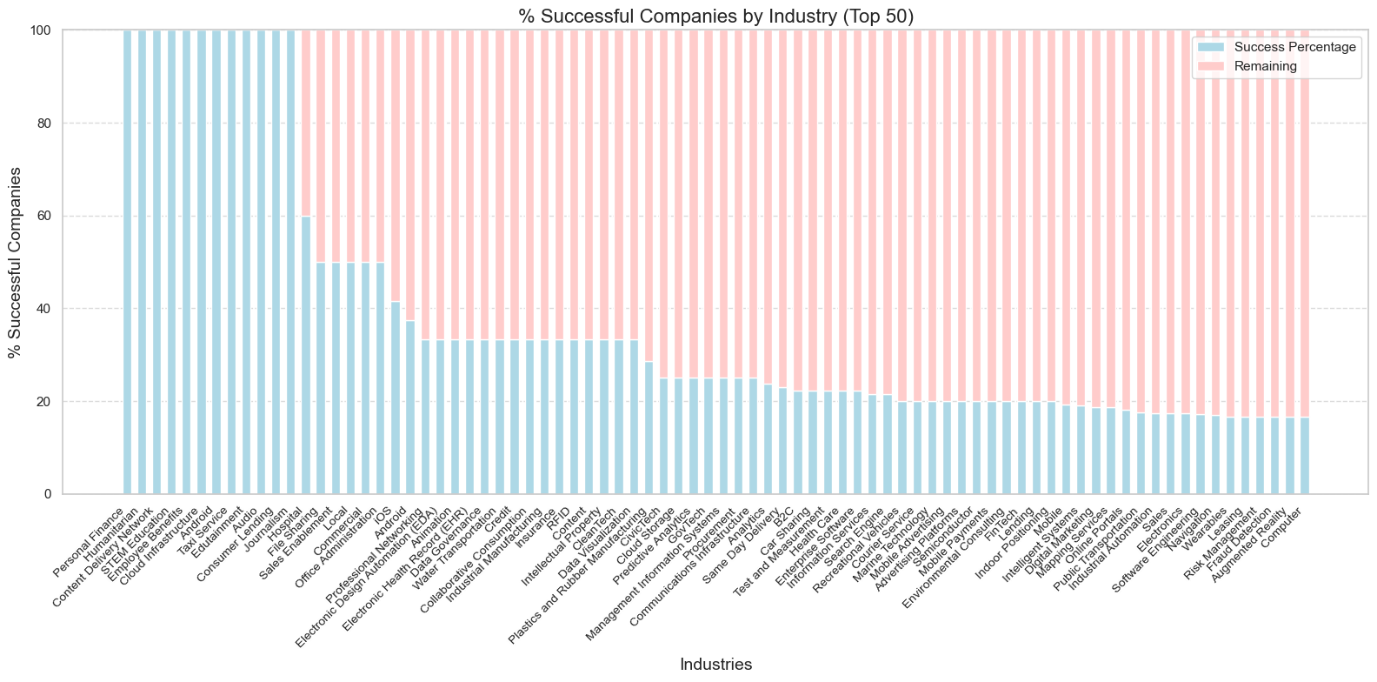
**شکل 4- 30- نمودار هیستوگرام تعداد کارکنان حد بالا**



**شکل 4- 31- نمودار جعبه‌ای تعداد کارکنان**

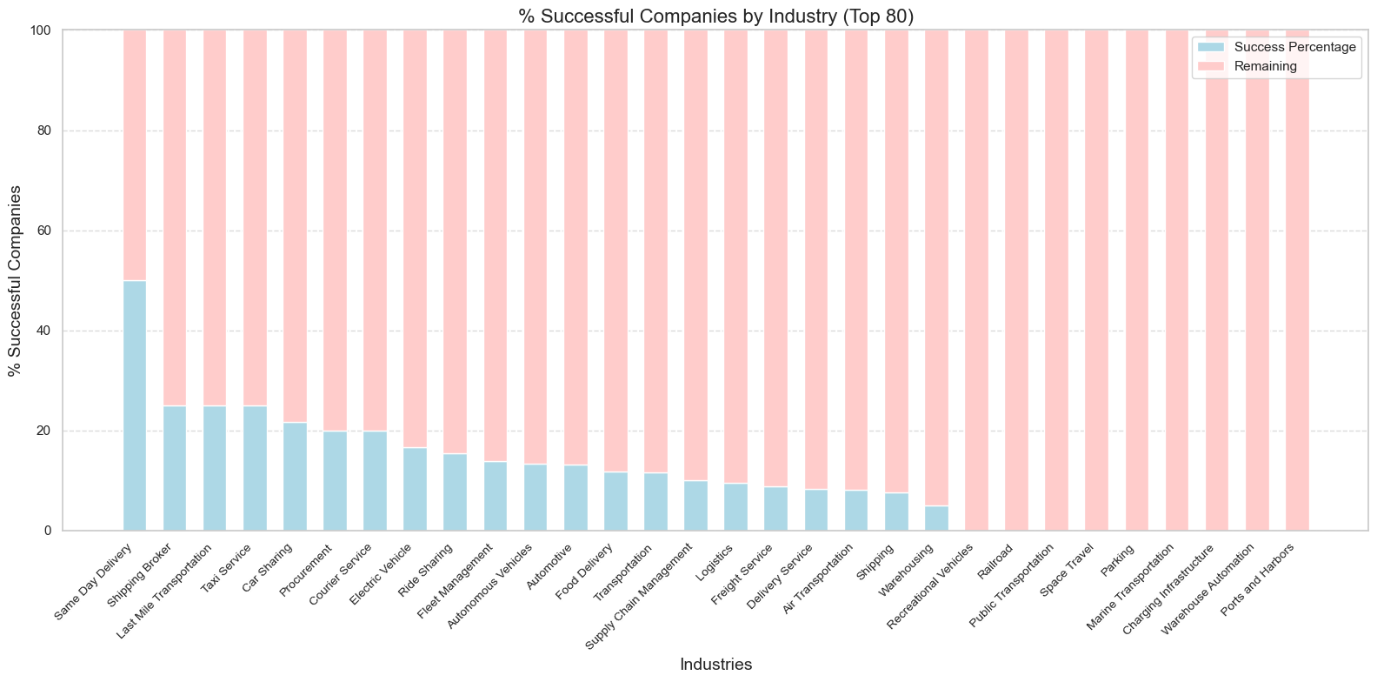
### موفقیت صنایع مختلف

در شکل 4-32 نشان داده شده است که کدامیک از صنایع درصد موفقیت بالاتری داشته‌اند.



**شکل 4- 32- میزان موفقیت صنایع مختلف**

در ادامه در شکل4-33 نشان داده شده است که از بین صنایع زیر مجموعه‌ی صنعت حمل و نقل (جدول 2-1) کدام صنایع موفقیت بیشتری داشته اند.



**شکل 4- 33-میزان موفقیت صنایع مختلف صنعت حمل و نقل**

## نتایج داده‌ها

در فصل قبل، ویژگی‌های صنایع را به 5 گروه مجزا تقسیم بندی کردیم تا اثر هر یک از این گروه‌ها بر موفقیت صنایع سنجیده شود. با توجه به این تقسیم‌بندی می‌توان اثر هر یک از ویژگی‌ها و گروه‌ها را بر موفقیت شرکت‌ها سنجید.

### عوامل موثر بر موفقیت شرکت‌های هوش مصنوعی حوزه حمل و نقل

با بررسی جدول 4-3 تا 4-7 می‌توان اهمیت هر یک از ویژگی‌ها بر موفقیت شرکت‌ها را سنجید. در ادامه هر الگوریتم به صورت مجزا بررسی می‌شود.

**الگوریتم درخت تصمیم** یکی از ساده‌ترین و قابل‌فهم‌ترین مدل‌ها در یادگیری ماشین است. این مدل به‌طور خودکار به دسته‌بندی داده‌ها یا پیش‌بینی خروجی با استفاده از تقسیم‌بندی ویژگی‌ها به صورت دوجانبه (بله/خیر) می‌پردازد. درخت‌های تصمیم با ساختار درختی نمایش داده می‌شوند که هر گره آن به یک ویژگی خاص اشاره دارد و هر برگ نمایانگر پیش‌بینی نهایی است.

* ویژگی‌های مهم در الگوریتم درخت تصمیم:

مجموع مبلغ تأمین مالی: این ویژگی در بالاترین اولویت قرار دارد، که نشان‌دهنده اهمیت بالای منابع مالی در تعیین موفقیت شرکت‌ها است.

تعداد کارکنان: اندازه شرکت یکی از عوامل کلیدی در درخت تصمیم است که به وضوح ارتباط مستقیم با موفقیت دارد.

موقعیت جغرافیایی دفتر مرکزی: دفاتر مرکزی شرکت‌ها که در مناطق توسعه‌یافته قرار دارند، معمولاً موفق‌تر هستند زیرا به منابع و بازارهای بزرگتری دسترسی دارند.

نوع آخرین تأمین مالی: درخت تصمیم بر این ویژگی تأکید دارد که نوع سرمایه‌گذاری (مثلاً سرمایه‌گذاری خطرپذیر یا خصوصی) می‌تواند تأثیر زیادی بر استراتژی‌های رشد و تصمیمات کلیدی شرکت‌ها داشته باشد.

درخت‌های تصمیم در مدل‌های پیچیده‌تر و متنوع‌تری مانند جنگل تصادفی و تقویت گرادیان به‌کار می‌روند، اما درخت تصمیم به‌تنهایی همچنان می‌تواند شواهد مهمی از ویژگی‌های تأثیرگذار ارائه دهد.

**الگوریتم تقویت گرادیان** یک روش پیشرفته است که به‌طور تدریجی مدل‌ها را تقویت می‌کند تا خطاهای مدل‌های قبلی را کاهش دهد. در این الگوریتم، چندین درخت تصمیم ساده ساخته می‌شود که هر کدام به‌طور متوالی تلاش می‌کنند تا خطای مدل‌های قبلی را تصحیح کنند.

* ویژگی‌های مهم در تقویت گرادیان:

مجموع مبلغ تأمین مالی: همان‌طور که در درخت تصمیم نیز دیده شد، این ویژگی در تقویت گرادیان نیز به‌عنوان عامل کلیدی در نظر گرفته شده است. تأمین مالی بالا به‌طور معمول نشان‌دهنده توان مالی بیشتر و امکان رشد سریع‌تر است.

تعداد اکتساب‌ها: یکی دیگر از ویژگی‌های مهم در مدل تقویت گرادیان است. خرید و ادغام با دیگر شرکت‌ها می‌تواند به شرکت‌ها مزیت رقابتی بدهد و ظرفیت‌های فنی و بازار آنها را گسترش دهد.

تعداد مقالات و اختراعات ثبت‌شده: این ویژگی‌ها در مدل تقویت گرادیان نیز به‌طور برجسته قرار دارند، که نشان‌دهنده نوآوری و تحقیق و توسعه در شرکت است.

تعداد سرمایه‌گذاران: در این مدل، تعداد سرمایه‌گذاران به‌عنوان فاکتور مهمی برای ارزیابی موفقیت شرکت‌ها مطرح شده است. سرمایه‌گذاران متعدد می‌توانند دسترسی به منابع گسترده‌تر و شبکه‌های تأثیرگذاری بیشتری را برای شرکت‌ها فراهم کنند.

**الگوریتم رگرسیون خطی** یکی از ساده‌ترین و معمول‌ترین الگوریتم‌ها برای پیش‌بینی خروجی پیوسته است. این الگوریتم به‌دنبال شناسایی روابط خطی میان ویژگی‌های ورودی و خروجی است.

* ویژگی‌های مهم در الگوریتم رگرسیون خطی:

تعداد اکتساب‌ها: در رگرسیون خطی، تعداد اکتساب‌ها با اهمیت بالایی مورد توجه قرار گرفته است. این امر نشان‌دهنده این است که اکتساب‌ها می‌توانند رشد و گسترش شرکت‌ها را تسریع کنند.

تعداد مقالات و اختراعات ثبت‌شده: همانطور که در سایر مدل‌ها نیز مشاهده شد، این ویژگی به‌ویژه در رگرسیون خطی نیز به‌عنوان یکی از عوامل مؤثر در موفقیت شرکت‌ها شناخته شده است. به دلیل اهمیت نوآوری در صنعت هوش مصنوعی، این ویژگی برای مدل‌های مختلف از اهمیت بالایی برخوردار است.

تعداد سرمایه‌گذاران و نوع تأمین مالی: در رگرسیون خطی، این دو ویژگی نشان‌دهنده تأثیر منابع مالی و استراتژی‌های تأمین مالی بر موفقیت شرکت‌ها هستند.

رگرسیون خطی ممکن است نتایج به‌دست‌آمده در مقایسه با مدل‌های پیچیده‌تر نظیر تقویت گرادیان و جنگل تصادفی کمی ساده و محدود به نظر برسد، اما همچنان در برخی موارد می‌تواند در پیش‌بینی نتایج مفید باشد.

**جنگل تصادفی** یک الگوریتم مبتنی بر یادگیری جمعی است که از تعداد زیادی درخت تصمیم استفاده می‌کند. برخلاف درخت تصمیم که به‌صورت تک‌درخت عمل می‌کند، جنگل تصادفی از مجموعه‌ای از درخت‌های تصمیم به‌طور هم‌زمان برای پیش‌بینی استفاده می‌کند و از آنجا که هر درخت به‌صورت مستقل عمل می‌کند، مدل به میزان قابل توجهی نسبت به خطاهای ناشی از داده‌های نادرست یا نویز ها مقاوم است.

* ویژگی‌های مهم در الگوریتم جنگل تصادفی:

مجموع مبلغ تأمین مالی: همانند سایر مدل‌ها، این ویژگی در جنگل تصادفی نیز به‌عنوان ویژگی مهمی شناخته شده است که نشان‌دهنده تأثیر منابع مالی بر توانایی‌های اجرایی و توسعه‌ای یک شرکت است.

تعداد اکتساب‌ها و مقالات ثبت‌شده: در جنگل تصادفی، این ویژگی‌ها نیز اهمیت دارند که به‌ویژه در زمینه نوآوری و گسترش شرکت‌ها از طریق خریداری شرکت‌های دیگر تأثیرگذار است.

تعداد کارکنان: اندازه نیروی انسانی به‌عنوان فاکتوری مهم در موفقیت و مقیاس‌پذیری یک شرکت شناخته شده است.

**الگوریتم XGBoost** یکی از الگوریتم‌های پیشرفته و بسیار مؤثر در یادگیری ماشین است که در رقابت‌های مختلف یادگیری ماشین نتایج چشمگیری کسب کرده است. این الگوریتم همان الگوریتم تقویت گرادیان است که بهینه‌سازی شده است تا سرعت و دقت بالاتری داشته باشد. XGBoost به‌ویژه برای داده‌های بزرگ و پیچیده مناسب است.

* ویژگی‌های مهم در XGBoost:

مجموع مبلغ تأمین مالی (در USD): همانند تقویت گرادیان، این ویژگی در XGBoost نیز بیشترین تأثیر را در موفقیت یک شرکت دارد.

تعداد اکتساب‌ها و مقالات ثبت‌شده: این ویژگی‌ها به‌ویژه در XGBoost به‌عنوان نشان‌دهنده قدرت نوآوری و گسترش شرکت‌ها از طریق خرید و اکتساب اهمیت دارند.

تعداد کارکنان و موقعیت جغرافیایی دفتر مرکزی: در XGBoost، ویژگی‌هایی مانند تعداد کارکنان و موقعیت جغرافیایی نیز به‌عنوان ویژگی‌های مهمی برای ارزیابی موفقیت یک شرکت شناخته شده است.

با بررسی پنج الگوریتم مختلف یادگیری ماشین (درخت تصمیم، تقویت گرادیان، رگرسیون خطی، جنگل تصادفی، و XGBoost)، مشخص شد که برخی ویژگی‌ها در پیش‌بینی موفقیت شرکت‌ها در این صنعت نقش برجسته‌ای دارند.

برای جمع بندی مهم‌ترین ویژگی‌هایی که در مدل‌های مختلف به‌عنوان عوامل تأثیرگذار شناسایی شده‌اند عبارتند از:

مجموع مبلغ تأمین مالی: این ویژگی در تمامی مدل‌ها به‌عنوان مؤثرترین عامل شناسایی شده است. منابع مالی بالا به شرکت‌ها این امکان را می‌دهد که از ظرفیت‌های مالی برای رشد، تحقیق و توسعه، و گسترش بازار استفاده کنند.

تعداد کارکنان: اندازه شرکت‌ها، به‌ویژه تعداد کارکنان، نشان‌دهنده ظرفیت عملیاتی و توانایی در مدیریت پروژه‌ها و مقیاس‌پذیری است. این عامل در بیشتر مدل‌ها به‌ویژه در درخت تصمیم و تقویت گرادیان برجسته است.

موقعیت جغرافیایی دفتر مرکزی: محل قرارگیری شرکت‌ها در مناطقی با اکوسیستم‌های فناوری پیشرفته و دسترسی به منابع و بازارهای بزرگ، از دیگر عوامل مهم است که در جنگل تصادفی و XGBoost اهمیت دارد.

نوع آخرین تأمین مالی: نوع تأمین مالی می‌تواند بر استراتژی‌های رشد و توسعه شرکت تأثیرگذار باشد. سرمایه‌گذاری‌های خطرپذیر و خصوصی معمولاً به شرکت‌ها امکان می‌دهند که استراتژی‌های تهاجمی و رشد سریع‌تری داشته باشند.

تعداد اکتساب‌ها و مقالات ثبت‌شده: این ویژگی‌ها به‌ویژه در مدل‌های درخت تصمیم و تقویت گرادیان اهمیت دارند و نشان‌دهنده توانایی شرکت در نوآوری و رشد از طریق خرید شرکت‌های دیگر هستند.

در نهایت، الگوریتم‌های مختلف به‌طور متفاوتی ویژگی‌ها را رتبه‌بندی کرده‌اند، اما در مجموع، عوامل مالی، اندازه شرکت، نوآوری و موقعیت جغرافیایی، مهم‌ترین عوامل در تعیین موفقیت شرکت‌های هوش مصنوعی در صنعت حمل و نقل به شمار می‌آیند. این تحلیل می‌تواند به سرمایه‌گذاران و مدیران کمک کند تا به‌طور مؤثرتر پتانسیل موفقیت شرکت‌ها را ارزیابی کرده و در تصمیم‌گیری‌های استراتژیک خود از این اطلاعات بهره‌برداری کنند.

برخی ویژگی‌ها میتواند تاثیر ناچیزی بر پیش بینی موفقیت شرکت‎ها داشته باشد، که با حذف این ویژگی‌ها میتوان ا ز پیچیدگی مدل کاست. در ادامه این ویژگی ها نام برده شده‌اند.

توییتر:

این ویژگی در تمام مدل‌ها اهمیت بسیار کمی دارد. در واقع، اهمیت آن در بیشتر مدل‌ها برابر با صفر یا نزدیک به صفر است، که نشان می‌دهد استفاده از داده‌های مربوط به توییتر برای پیش‌بینی موفقیت شرکت‌ها به‌نظر مفید نمی‌آید.

لینکدین:

مشابه با توییتر، این ویژگی نیز در بیشتر مدل‌ها اهمیت بسیار کمی دارد و می‌توان آن را از مجموعه ویژگی‌ها حذف کرد.

فیسبوک:

مشابه با توییتر و لینکدین، این ویژگی در مدل‌ها تأثیر کمی دارد و می‌توان آن را حذف کرد.

تعداد سرمایه‌گذاران اصلی (Number of Lead Investors):

این ویژگی در برخی مدل‌ها اهمیت کمی داشته است و بنابراین می‌توان آن را در برخی موارد حذف کرد.

نرخ پرش:

این ویژگی در بیشتر مدل‌ها تأثیر کمی دارد و در صورتی که هدف ساده‌تر کردن مدل و کاهش پیچیدگی باشد، می‌توان آن را حذف کرد.

برای ارزیابی اثر هر یک از این عوامل بر موفقیت شرکت‌ها از دو معیار خطای میانگین مربعی[[62]](#footnote-62) و ضریب تعیین[[63]](#footnote-63) استفاده می‌شود. برای این منظور ابتدا مدل جنگل تصادفی به عنوان مدل پیش‌بینی متغیر هدف (موفقیت) استفاده شده است. نتایج آن در جدول-12 آورده شده است.

**جدول 3- 2- تاثیر ویژگی‌های مختلف هر گروه از ویژگی‌ها بر موفقیت مدل جنگل تصادفی**

|  |  |  |  |
| --- | --- | --- | --- |
| ویژگی‌ها | MSE | R-squared | مهم‌ترین ویژگی‌ها در این گروه |
| ویژگی‌های مرتبط با شرکت | 0.0762 | 0.0868 | تعداد کارمندان (حد پایین و بالا)، سن شرکت، درآمد تخمینی، وضعیت فعالیت، تعداد بنیان‌گذاران |
| ویژگی‌های مرتبط با جذب سرمایه | 0.0006 | 0.0000 | وضعیت تأمین مالی، تعداد مراحل تأمین مالی، نوع تأمین مالی اخیر |
| ویژگی‌های مرتبط با سرمایه‌گذاران | 0.0880 | -0.0540 | تعداد سرمایه‌گذاران، تعداد سرمایه‌گذاران اصلی، گروه‌های سرمایه‌گذاران اصلی |
| ویژگی‌های مرتبط با فناوری | 0.0891 | 0.0345 | هزینه IT، تعداد فناوری‌های فعال، تعداد محصولات فعال، اختراعات ثبت‌شده، تعداد اپلیکیشن‌ها |
| ویژگی‌های مرتبط با صنعت | 0.0992 | -0.1886 | صنایع خودرو، نرم‌افزار، یادگیری ماشین، خدمات حمل‌ونقل، صنایع هوش مصنوعی |

1. ویژگی‌های مرتبط با شرکت

در این گروه، MSE نسبتا پایین است که نشان‌دهنده دقت قابل قبول مدل است. با این حال، R-squared (که نشان‌دهنده درصد واریانس توضیح داده‌شده توسط مدل است) مقدار نسبتا پایین 0.0868 دارد. این مقدار نشان می‌دهد که مدل تنها توانسته است درصد کمی از واریانس داده‌ها را توضیح دهد، که ممکن است به معنای پیچیدگی بالای روابط بین ویژگی‌ها و نتایج باشد.

ویژگی‌هایی مانند تعداد کارمندان بالاترین تأثیر را در مدل دارند که نشان‌دهنده اهمیت اطلاعات مرتبط با اندازه شرکت‌ها در پیش‌بینی موفقیت در صنعت حمل‌ونقل است.

سایر ویژگی‌ها مانند سن شرکت و درآمد تخمینی نیز تأثیرگذار هستند، اما اهمیت‌شان کمتر از اندازه شرکت‌ها است.

1. ویژگیهای مرتبط با جذب سرمایه

در این گروه، مدل عملکرد بسیار خوبی از نظر MSE داشته است که به مقدار بسیار کم 0.0006 می‌رسد. با این حال، R-squared برابر با صفر است که نشان‌دهنده این است که مدل نتوانسته است واریانس داده‌ها را توضیح دهد. ممکن است دلیل این موضوع این باشد که داده‌های موجود در این گروه از ویژگی‌ها برای پیش‌بینی موفقیت شرکت‌ها در صنعت حمل‌ونقل به‌اندازه کافی اطلاعات معنادار نداشته باشند.

تعداد راندهای سرمایه‌گذاری و سایر ویژگی‌های مرتبط با نوع تأمین مالی تأثیر زیادی در مدل دارند.

با این حال، از آنجایی که R-squared نزدیک به صفر است، می‌توان نتیجه گرفت که مدل در استفاده از ویژگی‌های تأمین مالی برای پیش‌بینی موفقیت شرکت‌ها در این صنعت موفق نبوده است.

1. ویژگی‌های مرتبط با سرمایه‌گذاران

در این گروه، MSE نسبتاً بالا است و به مقدار 0.0880 می‌رسد، که نشان می‌دهد مدل قادر به پیش‌بینی دقیق نبوده است. همچنین، R-squared منفی است که بدین معناست که مدل حتی از یک مدل ساده‌تر (مثلاً میانگین ساده) هم عملکرد ضعیف‌تری داشته است. این امر می‌تواند نشان‌دهنده نبود ارتباط معنی‌دار بین ویژگی‌های سرمایه‌گذاران و موفقیت شرکت‌ها در صنعت حمل‌ونقل باشد.

تعداد سرمایه‌گذاران بیشترین تأثیر را دارند. اما با توجه به عملکرد ضعیف مدل، این ویژگی‌ها نمی‌توانند به‌طور مؤثر موفقیت شرکت‌ها را در این صنعت پیش‌بینی کنند.

1. ویژگی‌های مرتبط با فناوری

در این گروه، MSE نسبتا بالا است، و R-squared نیز مقدار کمی دارد که نشان می‌دهد مدل در توضیح و پیش‌بینی موفقیت شرکت‌ها با استفاده از ویژگی‌های فناوری عملکرد قابل قبولی نداشته است. اگرچه بعضی از ویژگی‌ها مانند هزینه‌های فناوری اطلاعات و تعداد فناوری‌های فعال تأثیر نسبتا زیادی دارند، مدل نمی‌تواند ارتباط معنایی بین این ویژگی‌ها و موفقیت واقعی شرکت‌ها در صنعت حمل‌ونقل برقرار کند.

1. ویژگی‌های مرتبط با صنعت

در این گروه، MSE بزرگ‌ترین مقدار را دارد و R-squared منفی است که نشان‌دهنده این است که ویژگی‌های صنعت نتوانسته‌اند هیچ‌گونه واریانس مفیدی را توضیح دهند. این نشان می‌دهد که ویژگی‌های صنعتی که شرکت‌ها در آن‌ها فعال هستند، تأثیر زیادی بر موفقیت آن‌ها در صنعت حمل‌ونقل ندارند.

**نتیجه‌گیری کلی:**

گروه ویژگی‌های تأمین مالی کمترین MSE را دارد، اما مدل نتواسته است واریانس داده‌ها را توضیح دهد. این نشان‌دهنده بی‌تأثیری یا ناتوانی در استفاده از ویژگی‌های تأمین مالی برای پیش‌بینی موفقیت شرکت‌ها در این صنعت است.

گروه ویژگی‌های شرکت‌ها با اینکه مدل نسبتا دقیق عمل کرده است، اما با توجه به مقدار پایین R-squared، نشان می‌دهد که تنها بخش کوچکی از موفقیت شرکت‌ها قابل توضیح با این ویژگی‌ها است.

گروه ویژگی‌های فناوری و گروه ویژگی‌های صنعت نشان‌دهنده این هستند که ویژگی‌های مختلف در حوزه‌های فناوری و صنعت به نظر نمی‌رسد که بتوانند به‌طور مؤثر موفقیت شرکت‌ها را در صنعت حمل‌ونقل پیش‌بینی کنند.

در ادامه از مدل گرادیان تقویت شده به عنوان مدل پیش‌بینی متغیر هدف (موفقیت) استفاده شده است. نتایج آن در جدول-13 آورده شده است.

|  |  |  |
| --- | --- | --- |
| ویژگی‌ها | MSE | R-squared |
| ویژگی‌های مرتبط با شرکت | 0.0725 | 0.1308 |
| ویژگی‌های مرتبط با جذب سرمایه | 0.0000 | 0.0000 |
| ویژگی‌های مرتبط با سرمایه‌گذاران | 0.0819 | 0.0191 |
| ویژگی‌های مرتبط با فناوری | 0.0818 | 0.1145 |
| ویژگی‌های مرتبط با صنعت | 0.0842 | -0.0084 |

جدول 1- تاثیر ویژگی‌های مختلف بر موفقیت مدل گرادیان تقویت شده

???

### الگوریتم‌های داده کاوی

پس از تمیز کردن داده میتوان الگوریتم‌های مختلف داده‌کاوی را روی ویژگی‌های بدست آمده اجرا کرد. در این پژوهش برای دقت بالاتر و ارزیابی کاملتر عوامل موفقیت 7 الگوریتم داده کاوی برای بررسی ویژگی‌های شرکت‌های هوش مصنوعی در صنعت حمل و نقل استفاده شده است که در ادامه به بررسی نتایج هریک از این مدل‌ها می‌پردازیم .

در جدول 4-9 خلاصه‌ای از نتایج هر یک از مدل‌ها آورده شده است.

#### الگوریتم درخت تصمیم:

بر اساس اجرای 10 بار اعتبارسنجی متقابل، میانگین دقت مدل بر روی داده‌های تست 86.98% محاسبه شد که نشان‌دهنده عملکرد نسبتاً مناسب مدل در پیش‌بینی کلی نمونه‌ها است. با این حال، سایر معیارهای ارزیابی عملکرد مدل نشان می‌دهند که مدل در شناسایی کلاس‌های مثبت با چالش‌هایی مواجه است.

میانگین مقادیر معیارهای ارزیابی بر روی داده‌های تست:

دقت: 86.98 %

صحت: 50.27%

فراخوانی 40.44%

امتیاز F1: 43.23%

* ماتریس سردرگمی:

|  |  |  |
| --- | --- | --- |
| Negative | Positive |  |
| 89 | 6 | Positive |
| 8 | 4 | Negative |

این ماتریس نشان می‌دهد که مدل عملکرد خوبی در شناسایی کلاس منفی دارد اما در شناسایی کلاس مثبت دقت کمتری نشان می‌دهد.

#### الگوریتم جنگل تصادفی

بر اساس اجرای 10 بار اعتبارسنجی متقابل، میانگین دقت مدل بر روی داده‌های تست 90.72% محاسبه شد که نشان‌دهنده عملکرد نسبتاً مناسب مدل در پیش‌بینی کلی نمونه‌ها است. با این حال، سایر معیارهای ارزیابی عملکرد مدل نشان می‌دهند که مدل در شناسایی کلاس‌های مثبت با چالش‌هایی مواجه است.

میانگین مقادیر معیارهای ارزیابی بر روی داده‌های تست:

دقت: 90.72%

صحت: 86.67%

فراخوانی: 19%

امتیاز F1: 30.34%

* ماتریس سردرگمی:

|  |  |  |
| --- | --- | --- |
| Negative | Positive |  |
| 94.5 | 0.3 | Positive |
| 9.6 | 2.3 | Negative |

این ماتریس نشان می‌دهد که مدل عملکرد خوبی در شناسایی کلاس منفی دارد اما در شناسایی کلاس مثبت دقت کمتری نشان می‌دهد.

#### الگوریتم ماشین بردار پیشتیبان

بر اساس اجرای 10 بار اعتبارسنجی متقابل، میانگین دقت مدل بر روی داده‌های تست 89.78% محاسبه شد که نشان‌دهنده عملکرد نسبتاً مناسب مدل در پیش‌بینی کلی نمونه‌ها است. با این حال، سایر معیارهای ارزیابی عملکرد مدل نشان می‌دهند که مدل در شناسایی کلاس‌های مثبت با چالش‌هایی مواجه است.

میانگین مقادیر معیارهای ارزیابی بر روی داده‌های تست:

دقت: 86.31%

صحت: 36.52%

فراخوانی: 32.61%

امتیاز F1: 33.87%

### ماتریس سردرگمی:

|  |  |  |
| --- | --- | --- |
| Negative | Postitive |  |
| 94.3 | 0.5 | Positive |
| 10.4 | 1.5 | Negative |

نشان می‌دهد که مدل توانسته 94.3 نمونه منفی را به‌درستی پیش‌بینی کند و تنها 0.5 نمونه منفی را به اشتباه مثبت طبقه‌بندی کرده است. اما مدل در تشخیص کلاس مثبت عملکرد نسبتا ضعیفی دارد و نمونه‌های این کلاس را به اشتباه در کلاس منفی دسته‌بندی کرده است.

#### الگوریتم رگرسیون لجستیک

بر اساس اجرای 10 بار اعتبارسنجی متقابل، میانگین دقت مدل بر روی داده‌های تست 87.73% محاسبه شد که نشان‌دهنده عملکرد نسبتاً مناسب مدل در پیش‌بینی کلی نمونه‌ها است. با این حال، سایر معیارهای ارزیابی عملکرد مدل نشان می‌دهند که مدل در شناسایی کلاس‌های مثبت با چالش‌هایی مواجه است.

میانگین مقادیر معیارهای ارزیابی بر روی داده‌های تست:

دقت: 87.73%

صحت: 40.04%

فراخوانی: 17.17%

امتیاز F1: 23.64%

* ماتریس سردرگمی:

|  |  |  |
| --- | --- | --- |
| Negative | Positive |  |
| 91.4 | 3.4 | Positive |
| 9.7 | 2.2 | Negative |

نشان می‌دهد که مدل توانسته 91 نمونه منفی را به‌درستی پیش‌بینی کند و تنها 3.4 نمونه منفی را به اشتباه مثبت طبقه‌بندی کرده است. اما مدل در تشخیص کلاس مثبت عملکرد خیلی قوی ندارد و نمونه‌های این کلاس را به اشتباه در کلاس منفی دسته‌بندی کرده است.

#### الگوریتم تقویت گرادیان:

بر اساس اجرای 10 بار اعتبارسنجی متقابل، میانگین دقت مدل بر روی داده‌های تست 87.44% محاسبه شد که نشان‌دهنده عملکرد نسبتاً مناسب مدل در پیش‌بینی کلی نمونه‌ها است. با این حال، سایر معیارهای ارزیابی عملکرد مدل نشان می‌دهند که مدل در شناسایی کلاس‌های مثبت با چالش‌هایی مواجه است.

میانگین مقادیر معیارهای ارزیابی بر روی داده‌های تست:

دقت: 91.44%

صحت: 80.86%

فراخوانی: 39.33%

امتیاز F1: 45.19%

* ماتریس سردرگمی

|  |  |  |
| --- | --- | --- |
| Negative | Positive |  |
| 90 | 4 | Positive |
| 5.5 | 4.5 | Negtive |

نشان می‌دهد که مدل توانسته 90 نمونه منفی را به‌درستی پیش‌بینی کند و 4 نمونه منفی را به اشتباه مثبت طبقه‌بندی کرده است. اما مدل همچنان در شناسایی نمونه‌های مثبت کمی مشکل دارد.

#### الگوریتم بیز ساده گوسی:

بر اساس اجرای 10 بار اعتبارسنجی متقابل، میانگین دقت مدل بر روی داده‌های تست 88.76% محاسبه شد که نشان‌دهنده عملکرد نسبتاً مناسب مدل در پیش‌بینی کلی نمونه‌ها است. با این حال، سایر معیارهای ارزیابی عملکرد مدل نشان می‌دهند که مدل در شناسایی کلاس‌های مثبت با چالش‌هایی مواجه است.

میانگین مقادیر معیارهای ارزیابی بر روی داده‌های تست:

دقت: 88.76%

صحت: 48.58%

فراخوانی: 20.51%

امتیاز F1: 28.15%

* ماتریس سردرگمی

|  |  |  |
| --- | --- | --- |
| Negative | Positive |  |
| 92.2 | 2.6 | Positive |
| 9.4 | 2.5 | Negative |

#### الگوریتم شبکه عصبی:

* شبکه عصبی اول :

عملکرد ضعیف با دقت ۶۱.۶٪ و امتیاز F1 پایین (۱۸٪). این مدل به دلیل عدم پیچیدگی کافی نتوانسته داده‌ها را به‌خوبی مدل کند.

* شبکه عصبی دوم

بهبود جزئی در فراخوانی (۱.۳٪) اما همچنان این مدل به دلیل پیچیدگی بیش از حد دچار مشکل شده است.

**جدول 4- 8- نتیجه الگوریتم‌های داده کاوی**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| مدل | دقت | صحت | فراخوانی | امتیاز F1 |
| درخت تصمیم | 86.98% | 50.27% | 40.44% | 43.23% |
| جنگل تصادفی | 90.72% | 86.67% | 19% | 30.34% |
| ماشین بردار پشتیبان | 86.31% | 36.52% | 32.61% | 33.87% |
| رگرسیون لجستیک | 87.73% | 40.04% | 17.17% | 23.64% |
| تقویت گرادیان | 91.44% | 80.86% | 39.33% | 45.18% |
| بیز ساده گوسی | 88.76% | 48.58% | 20.51% | 28.15% |
| شبکه عصبی اول | 61.6% | 11.2% | 46% | 18% |
| شبکه عصبی دوم | 90.9% | 100% | 1.3% | 2.6% |

نتایج نشان می‌دهد که الگوریتم تقویت گرادیان دارای بالاترین دقت معادل 91.44% بوده است که حاکی از توانایی بالای این مدل در تفکیک استارتاپ‌های موفق و ناموفق است. علاوه بر این، این مدل صحتی برابر با 80.86% را ثبت کرده است که نشان‌دهنده عملکرد مطلوب آن در کاهش پیش‌بینی‌های مثبت کاذب (False Positives) است. یکی از مهم‌ترین شاخص‌های ارزیابی مدل، امتیاز F1 است که ترکیبی از صحت و فراخوانی بوده و در این الگوریتم مقدار 45.18% را نشان می‌دهد. این مقدار در مقایسه با سایر مدل‌ها بالاترین سطح تعادل را بین دو معیار صحت و فراخوانی ارائه می‌دهد.

یکی از دلایل برتری تقویت گرادیان در این پژوهش، قابلیت آن در مدیریت پیچیدگی‌های داده‌های آموزشی است. این الگوریتم با استفاده از روش تقویت و ترکیب چندین مدل ضعیف‌تر، ساختاری مقاوم در برابر مشکلاتی همچون بیش‌برازش و نویز داده‌ها ایجاد می‌کند. همچنین، تقویت گرادیان قادر به شناسایی الگوهای غیرخطی پیچیده است که در سایر روش‌های کلاسیک همچون رگرسیون لجستیک و ماشین بردار پشتیبان کمتر مشاهده شد.

در کنار تقویت گرادیان، جنگل تصادفی نیز عملکرد نسبتاً خوبی داشت و دقتی معادل 90.72% را ثبت کرد. با این حال، میزان فراخوانی این الگوریتم تنها 19% بود که نشان می‌دهد مدل در تشخیص موارد موفق ضعف دارد و بسیاری از استارتاپ‌های موفق را به‌درستی شناسایی نکرده است. در مقابل، درخت تصمیم با دقت 86.98% و فراخوانی 40.44% تعادل بهتری میان معیارهای صحت و یادآوری ایجاد کرده است، اما همچنان در مقایسه با تقویت گرادیان عملکرد ضعیف‌تری داشته است.

برخی دیگر از الگوریتم‌های مورد بررسی، به‌ویژه شبکه‌های عصبی مصنوعی، عملکرد ضعیفی از خود نشان دادند. در حالی که مدل اول شبکه عصبی دارای دقت 61.6% و امتیاز F1 برابر با 18% بود، مدل دوم علی‌رغم دقت 90.9%، دارای فراخوانی 1.3% و امتیاز F1 معادل 2.6% بود. این نتایج نشان می‌دهد که شبکه‌های عصبی در این پژوهش به دلیل پیچیدگی بیش‌ازحد و نیاز به داده‌های گسترده‌تر برای آموزش، دچار مشکل بیش‌برازش شده‌اند و نتوانسته‌اند عملکرد قابل اعتمادی ارائه دهند. دلیل این امر می‌تواند کمبود داده‌های در دسترس در این حوزه باشد.

همچنین، ماشین بردار پشتیبان و رگرسیون لجستیک نیز عملکرد چندان مطلوبی نداشتند. ماشین بردار پشتیبان با دقت 86.31% و صحت 36.52% در تشخیص موارد مثبت ضعف نشان داد. از سوی دیگر، رگرسیون لجستیک با دقت 87.73% و فراخوانی 17.17% نشان داد که در مقایسه با مدل‌های غیربرخط مانند جنگل تصادفی و تقویت گرادیان، توانایی کمتری در استخراج روابط پیچیده میان متغیرها دارد.

نتایج این پژوهش نشان می‌دهد که الگوریتم تقویت گرادیان در مقایسه با سایر روش‌های داده‌کاوی، بالاترین عملکرد را در پیش‌بینی موفقیت استارتاپ‌های حوزه حمل‌ونقل داشته است. این برتری ناشی از توانایی بالای این مدل در کاهش خطاهای مثبت کاذب و کشف الگوهای پیچیده در داده‌ها است. در کنار آن، جنگل تصادفی و درخت تصمیم نیز عملکرد نسبتاً خوبی داشته‌اند، اما ضعف آن‌ها در میزان فراخوانی مانع از برتری کامل این روش‌ها شده است. از سوی دیگر، شبکه‌های عصبی و ماشین بردار پشتیبان به دلیل نیاز به داده‌های بیشتر و احتمال بیش‌برازش، عملکرد قابل قبولی ارائه ندادند.

این یافته‌ها نشان می‌دهد که برای مسائل مرتبط با پیش‌بینی موفقیت استارتاپ‌ها، الگوریتم‌های مبتنی بر Boosting گزینه‌های مناسبی هستند، زیرا قادرند ترکیبی بهینه از دقت و صحت را ارائه دهند و نسبت به سایر روش‌های کلاسیک و پیچیده‌تر، انعطاف‌پذیری بالاتری در تحلیل داده‌های پرتلاطم بازار داشته باشند.

## جمع‌بندی

در این فصل، ابتدا به بررسی ویژگی‌های مورد مطالعه و یافتن همبستگی میان آن‌ها برای انتخاب بهینه ویژگی‌ها پرداخته شد. سپس با توجه به این ویژگی‌ها چندین مدلسازی با الگوریتم‌های داده کاوی صورت گرفت و سپس به تحلیل نتایج حاصل از مدل‌سازی و ارزیابی داده‌های مربوط به شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل پرداخته شد. این تحلیل‌ها بر اساس الگوریتم‌های یادگیری ماشین و تکنیک‌های پیشرفته داده‌کاوی صورت گرفت تا عواملی که تأثیر مستقیمی بر موفقیت یا شکست شرکت‌ها دارند شناسایی شوند. داده‌های مورد استفاده در این پژوهش از منابع معتبر و جامع مانند پایگاه داده کرانچ‌بیس جمع‌آوری شدند، که اطلاعاتی همچون تعداد دورهای تأمین مالی، تعداد کارکنان، میزان سرمایه‌گذاری، و موقعیت جغرافیایی شرکت‌ها را شامل می‌شود. از ترکیب داده‌های ساختاریافته مانند تعداد اختراعات، سرمایه‌گذاری‌ها و سایر ویژگی‌های ذکر شده، مدل‌های مختلفی برای پیش‌بینی موفقیت استفاده شد. تحلیل ویژگی‌ها با توجه به خروجی مدل‌های مختلف نشان می‌دهد که: Number of Patents (تعداد اختراعات)، Total Funding Amount (مقدار کل سرمایه‌گذاری) و Estimated Revenue (درآمد تخمینی) به‌عنوان عوامل تأثیرگذار کلیدی بر موفقیت شرکت‌ها شناسایی شده‌اند.

ویژگی‌های مرتبط با موقعیت جغرافیایی و تعداد کارمندان نیز تأثیر معناداری دارند اما شدت آن در مقایسه با ویژگی‌های مالی و نوآوری کمتر است.

بهترین عملکرد مربوط به مدل‌هایی است که توانسته‌اند تعادل مناسبی بین دقت، فراخوانی و امتیاز F1 برقرار کنند.

بر اساس دقت، فراخوانی و امتیاز F1، مدل تقویت گرادیان بهترین عملکرد کلی را ارائه داده است. در رتبه بعد، مدل جنگل تصادفی قرار می‌گیرد که با وجود فراخوانی پایین‌تر نسبت به الگوریتم تقویت گرادیان، دقت بالایی داشته است.

این پژوهش می‌تواند بینشی جدید در پیش‌بینی موفقیت شرکت‌های هوش مصنوعی در صنعت حمل‌ونقل ارائه دهد. با استفاده از مدل‌های پیش‌بینی‌کننده توسعه‌یافته، می‌توان نه‌تنها تصمیمات بهتری در زمینه سرمایه‌گذاری اتخاذ کرد، بلکه مسیرهایی برای بهبود و توسعه پایدار این صنعت فراهم آورد. این نتایج به‌ویژه در محیط‌های پویا و پیچیده امروزی، ابزاری قدرتمند برای کاهش ریسک و افزایش بازدهی در اختیار ذی‌نفعان قرار می‌دهند.

1. جمع‌بندي و پيشنهادها

## مقدمه

صنعت حمل‌ونقل، به‌عنوان یکی از ارکان اساسی اقتصاد جهانی، از دیرباز نقش مهمی در تسهیل جابه‌جایی کالاها و افراد و بهبود کیفیت زندگی ایفا کرده است. با افزایش جمعیت، رشد اقتصادی، و نیاز به تحرک بیشتر، این صنعت با چالش‌های پیچیده‌ای مانند ترافیک سنگین، آلودگی‌های محیطی، ایمنی، و افزایش هزینه‌های حمل‌ونقل روبه‌رو است. در این راستا، استفاده از فناوری‌های نوین و به‌ویژه هوش مصنوعی و یادگیری ماشین به عنوان ابزاری نوآورانه، به سرعت به‌عنوان راه‌حل‌های مؤثر برای این مشکلات شناخته شده‌اند. هوش مصنوعی، با قابلیت پردازش داده‌های عظیم و استخراج الگوهای پنهان، می‌تواند به بهبود عملکرد و کارایی این صنعت کمک شایانی کند.

هوش مصنوعی و یادگیری ماشین نه تنها در مدیریت ترافیک، خودروهای خودران، و بهینه‌سازی مسیرها، بلکه در تحلیل داده‌های بازار، پیش‌بینی تقاضا، و بهبود خدمات حمل‌ونقل عمومی نیز کاربرد دارند. در همین راستا، استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل، به‌ویژه در بخش‌هایی چون خودروهای خودران، حمل‌ونقل هوشمند، و لجستیک، به‌سرعت در حال رشد و توسعه هستند. این استارتاپ‌ها با ارائه فناوری‌های نوین، در تلاش‌اند تا مشکلات موجود در صنعت حمل‌ونقل را حل کنند و به‌نوبه خود موجب تحول در این صنعت شوند.

با توجه به اهمیت روزافزون، پیش‌بینی موفقیت استارتاپ‌ها و شرکت‌های نوپا در این صنعت، یکی از چالش‌های اساسی پیش روی کارآفرینان، سرمایه‌گذاران، و سیاست‌گذاران است. موفقیت یک استارتاپ، به‌ویژه در حوزه‌ای پیچیده و نوظهور مانند هوش مصنوعی، تحت تأثیر عواملی متعدد قرار دارد که برخی از آن‌ها می‌توانند به‌راحتی نادیده گرفته شوند. از جمله این عوامل می‌توان به تأمین مالی، کیفیت تیم مدیریتی، تعداد سرمایه‌گذاران، ویژگی‌های بازار، و تأثیرات تکنولوژیکی اشاره کرد.

در این تحقیق، هدف اصلی ارائه یک مدل پیش‌بینی موفقیت برای استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل بود. برای رسیدن به این هدف، از داده‌های موجود در وب‌سایت‌های معتبر مانند کرانچ بیس استفاده شد که اطلاعات جامع و دقیقی در مورد استارتاپ‌ها، میزان سرمایه جذب‌شده، تعداد دورهای تأمین مالی، ویژگی‌های بنیان‌گذاران، و روندهای بازار در اختیار قرار می‌دهد. از این داده‌ها به‌عنوان مبنای اصلی برای تحلیل و پیش‌بینی موفقیت استفاده شد.

در این راستا، از الگوریتم‌های داده‌کاوی و یادگیری ماشین مانند درخت تصمیم، شبکه‌های عصبی مصنوعی، جنگل تصادفی، و ماشین بردار پشتیبان بهره گرفته شد. این الگوریتم‌ها به‌ویژه در شناسایی الگوهای پنهان و روابط پیچیده میان متغیرها، عملکرد بسیار خوبی از خود نشان داده‌اند. به‌طور خاص، نتایج نشان داد که برخی از متغیرها مانند میزان تأمین مالی، تعداد دورهای سرمایه‌گذاری، حضور در رسانه‌ها، و نوع صنعت، تأثیر زیادی در پیش‌بینی موفقیت استارتاپ‌ها دارند.

در این فصل، به جمع‌بندی کلی نتایج تحقیق پرداخته می‌شود و تلاش خواهد شد که یافته‌های بدست‌آمده در قالب پیشنهادات کاربردی و علمی ارائه شوند. همچنین، محدودیت‌های تحقیق مورد بررسی قرار خواهند گرفت و پیشنهاداتی برای تحقیقات آتی ارائه خواهد شد. این تحقیق، از جنبه‌های مختلفی به توسعه دانش و نوآوری در حوزه پیش‌بینی موفقیت استارتاپ‌ها، به‌ویژه در زمینه‌های هوش مصنوعی و حمل‌ونقل کمک کرده و می‌تواند به کارآفرینان و سرمایه‌گذاران ابزارهای ارزشمندی برای تصمیم‌گیری‌های هوشمندانه در فرآیندهای سرمایه‌گذاری ارائه دهد.

در ادامه، به بررسی دستاوردهای این تحقیق و تأثیرات آن بر توسعه پایدار استارتاپ‌های هوش مصنوعی در صنعت حمل‌ونقل پرداخته می‌شود. نتایج این تحقیق نه تنها به شناسایی عواملی که موجب موفقیت استارتاپ‌ها در این حوزه می‌شوند، کمک کرده، بلکه نشان داده است که استفاده از مدل‌های پیش‌بینی می‌تواند در شناسایی فرصت‌های سرمایه‌گذاری و کاهش ریسک‌های مرتبط با آن نقش کلیدی ایفا کند.

## پاسخ به سوالات پژوهش

**سوال اول: چه عواملی بیشترین تاثیر را بر موفقیت شرکت‌های هوش مصنوعی در صنعت حمل و نقل دارند؟**

در این پژوهش از 5 مدل مختلف جنگل تصادفی، درخت تصمیم، تقویت گرادیان، رگرسیون خطی و XGBoost برای بررسی اهمیت ویژگی‌های مختلف بر موفقیت به کار گرفته شده است. بررسی مدل‌های مختلف نشان می‌دهد که برخی متغیرها در تمام مدل‌ها اهمیت زیادی دارند و می‌توان آن‌ها را به‌عنوان عوامل کلیدی موفقیت استارتاپ‌ها در نظر گرفت. این متغیرها عبارتند از:

**تعداد کارمندان:** تقریباً در تمامی مدل‌ها، تعداد کارمندان استارتاپ یک عامل مهم در موفقیت شرکت شناخته شده است. شرکت‌هایی که دارای تیم‌های بزرگ‌تر و با مهارت بالا هستند، احتمال موفقیت بیشتری دارند.

**میزان کل سرمایه جذب شده:** سرمایه‌گذاری‌های انجام شده در استارتاپ‌ها، نقشی کلیدی در رشد و توسعه آن‌ها دارد. شرکت‌هایی که توانسته‌اند جذب سرمایه قوی‌تری داشته باشند، در مسیر موفقیت قرار گرفته‌اند.

**وضعیت عملیاتی شرکت**: اینکه آیا استارتاپ همچنان فعال است یا فعالیت‌های آن متوقف شده، در موفقیت آن بسیار مؤثر است.

**تعداد خریدهای انجام شده**~~:~~ شرکت‌هایی که توانسته‌اند استارتاپ‌های دیگر را خریداری کنند، شانس بیشتری برای رشد و بقای بلندمدت دارند. این متغیر در مدل رگرسیون خطی و XGBoost اهمیت بسیار بالایی داشت.

**موقعیت جغرافیایی و دفتر مرکزی شرکت:** برخی مناطق جهان به دلیل داشتن اکوسیستم استارتاپی قوی‌تر، نرخ موفقیت بالاتری دارند. به همین دلیل، موقعیت مکانی استارتاپ تأثیر بسزایی در موفقیت آن دارد.

عمر شرکت: شرکت‌های با سابقه بیشتر، شانس بالاتری برای دریافت حمایت مالی و موفقیت در بازار دارند.

**تعداد پتنت‌های ثبت‌شده:** شرکت‌هایی که نوآوری بیشتری دارند و موفق به ثبت اختراعات شده‌اند، به‌طور کلی عملکرد بهتری در بازار داشته‌اند.

بررسی مدل‌های مختلف یادگیری ماشین نشان داد که مهم‌ترین گروه از ویژگی‌ها در پیش‌بینی موفقیت استارتاپ‌ها، ویژگی‌های مرتبط با شرکت است. درواقع، متغیرهایی مانند تعداد کارکنان، عمر شرکت و موقعیت جغرافیایی بیشترین نقش را در تعیین موفقیت استارتاپ دارند.

پس از آن، ویژگی‌های مرتبط با جذب سرمایه قرار دارند، که شامل میزان سرمایه جذب‌شده و نوع سرمایه‌گذاری دریافت شده هستند. این ویژگی‌ها به‌ویژه در مدل‌های Gradient Boosting و XGBoost اهمیت بالایی دارند.

در رتبه سوم، ویژگی‌های مرتبط با سرمایه‌گذاران قرار دارند. اگرچه حضور سرمایه‌گذاران مطرح می‌تواند به رشد استارتاپ کمک کند، اما در مقایسه با ویژگی‌های شرکت و جذب سرمایه تأثیر کمتری دارد.

ویژگی‌های مرتبط با فناوری و صنعت نیز اهمیت دارند، اما تأثیر آن‌ها در مقایسه با سایر گروه‌ها کمتر است. به‌ویژه، نوع صنعتی که استارتاپ در آن فعالیت می‌کند، در پیش‌بینی موفقیت آن نقش قابل‌توجهی ندارد.

در نهایت، ترکیب این متغیرها با یکدیگر نشان می‌دهد که موفقیت استارتاپ‌ها وابسته به ترکیبی از نیروی انسانی، سرمایه، نوآوری، موقعیت جغرافیایی و تعامل در رسانه‌ها است. بسته به نوع صنعت و بازار، ممکن است وزن هر یک از این متغیرها تغییر کند، اما نتایج این پژوهش تأکید دارد که سرمایه، تعداد کارکنان و تعامل در بازار از مهم‌ترین شاخص‌های موفقیت استارتاپ‌های حوزه حمل‌ونقل و فناوری‌های هوش مصنوعی هستند.

**سوال دوم: چگونه می‌توان با استفاده از داده‌های موجود، موفقیت شرکت‌های هوش مصنوعی را پیش‌بینی کرد؟**

پیش‌بینی موفقیت شرکت‌های هوش مصنوعی نیازمند تجزیه و تحلیل دقیق داده‌های مالی، ساختاری، فناوری، سرمایه‌گذاری و سایر عوامل کلیدی است. با استفاده از یادگیری ماشین و تحلیل داده‌ها، می‌توان الگوهای مشخصی را شناسایی کرد که نشان می‌دهند کدام شرکت‌ها احتمال بیشتری برای رشد، جذب سرمایه‌گذاری و موفقیت در بازار دارند. در این راستا، ابتدا باید معیارهای موفقیت شرکت‌های هوش مصنوعی تعریف شوند. موفقیت می‌تواند بر اساس میزان سرمایه جذب‌شده، افزایش تعداد کاربران یا مشتریان، افزایش ارزش بازار، خروجی‌های فناورانه مانند پتنت‌های ثبت‌شده، ادغام یا خریداری شدن توسط شرکت‌های بزرگ‌تر و رسیدن به مراحل پیشرفته تأمین مالی مانند سری A، B و عرضه اولیه سهام سنجیده شود. این معیارها به‌عنوان متغیرهای خروجی مدل یادگیری ماشین برای پیش‌بینی موفقیت شرکت‌ها استفاده می‌شوند. در این پژوهش عرضه اولیه سهام و ادغام و تملیک به عنوان معیارهای موفقیت در نظر گرفته شده اند.

برای رسیدن به پیش‌بینی‌های دقیق، باید داده‌های متنوعی جمع‌آوری شود. در این پژوهش این داده‌ها به پنج گروه کلی دسته‌بندی شده‌اند: ویژگی‌های مرتبط با شرکت، ویژگی‌های مرتبط با جذب سرمایه، ویژگی‌های مرتبط با سرمایه‌گذاران، ویژگی‌های مرتبط با فناوری و ویژگی‌های مرتبط با صنعت. ویژگی‌های مرتبط با شرکت شامل عواملی مانند تعداد کارمندان، عمر شرکت، موقعیت مکانی دفتر مرکزی، وضعیت عملیاتی و رتبه شرکت در پایگاه داده‌های کسب‌وکار است که نشان‌دهنده ساختار و عملکرد کلی شرکت هستند. در کنار آن، ویژگی‌های مرتبط با جذب سرمایه همچون میزان کل سرمایه جذب‌شده، نوع آخرین جذب سرمایه، تعداد دورهای تأمین مالی و ارزش کل سرمایه‌گذاری سهام، تصویری روشن از وضعیت مالی استارتاپ‌ها ارائه می‌دهند.

نقش سرمایه‌گذاران نیز در موفقیت استارتاپ‌های هوش مصنوعی قابل توجه است. تعداد سرمایه‌گذاران، تعداد سرمایه‌گذاران اصلی و حضور سرمایه‌گذاران معتبر از جمله عواملی هستند که بر رشد و توسعه این شرکت‌ها تأثیر دارند. در کنار این عوامل، نوآوری‌های فناورانه نیز نقش مهمی در موفقیت ایفا می‌کنند. تعداد پتنت‌های ثبت‌شده، میزان هزینه‌های فناوری اطلاعات، تعداد محصولات و خدمات هوش مصنوعی فعال و میزان دانلود محصولات از جمله شاخص‌هایی هستند که نشان‌دهنده میزان نوآوری و جایگاه فناوری شرکت در بازار است. علاوه بر این، ویژگی‌های صنعتی مانند دسته‌بندی صنعتی، میزان حضور رقبا در حوزه مشابه، تعداد خریدهای انجام‌شده توسط شرکت‌های هوش مصنوعی و رتبه جهانی ترافیک اینترنتی شرکت به درک بهتر از فضای رقابتی کمک می‌کنند.

با جمع‌آوری و پردازش این داده‌ها، از روش‌های یادگیری ماشین برای پیش‌بینی موفقیت شرکت‌های هوش مصنوعی استفاده شده است. الگوریتم‌های داده‌کاوی مانند درخت تصمیم، جنگل تصادفی، تقویت گرادیان، رگرسیون لجستیک، بیز ساده، ماشین بردار پشتیبان و شبکه‌های عصبی برای این کار مناسب هستند. الگوریتم جنگل تصادفی می‌تواند به‌خوبی تأثیر متغیرهای مختلف را بررسی کند، در حالی که الگوریتم تقویت گرادیان دقت پیش‌بینی را افزایش می‌دهند.

با اجرای این مدل‌ها، مشخص شد که برخی متغیرها تأثیر بیشتری بر موفقیت شرکت‌های هوش مصنوعی دارند. مهم‌ترین عواملی که در تمام مدل‌ها تکرار شده‌اند شامل تعداد کارمندان، میزان سرمایه جذب‌شده، وضعیت عملیاتی شرکت، تعداد خریدهای انجام‌شده، تعداد پتنت‌های ثبت‌شده و میزان تعاملات شبکه‌ای بنیان‌گذاران هستند. بررسی مدل‌های مختلف نشان داده است که شرکت‌هایی با تیم‌های تخصصی بزرگ‌تر، سرمایه بیشتر، حضور فعال در بازار و نوآوری‌های ثبت‌شده، احتمال بیشتری برای موفقیت دارند.

پس از توسعه مدل‌های پیش‌بینی، یک روش تحلیلی برای نمایش نتایج و ارائه ابزار تصمیم‌گیری برای سرمایه‌گذاران و استارتاپ‌ها ایجاد شده است. چنین ابزاری به سرمایه‌گذاران کمک می‌کند تا شرکت‌های هوش مصنوعی دارای پتانسیل را شناسایی کرده و ریسک سرمایه‌گذاری را کاهش دهند. همچنین، کارآفرینان می‌توانند با بررسی ویژگی‌های موفق‌ترین شرکت‌ها، استراتژی‌های خود را برای بهبود عملکرد و جذب سرمایه بهینه کنند.

در نهایت، این پژوهش نشان می‌دهد که ترکیب داده‌های مالی، ساختاری، فناوری و صنعتی با استفاده از مدل‌های یادگیری ماشین می‌تواند دقت پیش‌بینی موفقیت شرکت‌های هوش مصنوعی را به میزان قابل توجهی افزایش دهد. دقت به دست آمده از طریق الگوریتم تقویت گرادیان در این پژوهش عدد 91.44% را گزارش می‌دهد. عوامل کلیدی که بر موفقیت این شرکت‌ها تأثیرگذارند، شامل تعداد کارکنان، میزان سرمایه جذب‌شده، وضعیت عملیاتی شرکت، تعداد خریدهای انجام‌شده، تعداد پتنت‌های ثبت‌شده و سطح تعاملات شبکه‌ای بنیان‌گذاران است. این روش‌ها می‌توانند به سرمایه‌گذاران در اتخاذ تصمیم‌های آگاهانه و به استارتاپ‌ها در بهبود استراتژی‌های رشد و توسعه کمک کنند.

**سوال سوم: کدام الگوریتم‌های داده‌کاوی بهترین عملکرد را در پیش‌بینی موفقیت این شرکت‌ها دارند؟**

در تحلیل عملکرد الگوریتم‌های داده‌کاوی برای پیش‌بینی موفقیت استارتاپ‌های حوزه حمل‌ونقل، معیارهایی مانند دقت، صحت، فراخوانی و امتیاز F1 بررسی شدند. این مقایسه نشان داد که برخی از الگوریتم‌ها توانایی بهتری در شناسایی استارتاپ‌های موفق دارند. در این میان، الگوریتم تقویت گرادیان برترین عملکرد را ارائه داده است. نتایج به‌دست‌آمده نشان می‌دهد که تقویت گرادیان با دقت 91.44% بهترین تفکیک را بین استارتاپ‌های موفق و ناموفق ارائه می‌دهد. علاوه بر این، میزان صحت آن 80.86% است، که نشان می‌دهد این مدل کمترین میزان خطا در پیش‌بینی استارتاپ‌های موفق را دارد. معیار مهم امتیاز F1 در این الگوریتم 45.18% ثبت شده است که بالاترین تعادل بین صحت و فراخوانی را نشان می‌دهد. دلیل موفقیت تقویت گرادیان در این پژوهش به توانایی آن در ترکیب چندین مدل ضعیف و ایجاد یک مدل قوی‌تر مربوط می‌شود. این الگوریتم قابلیت شناسایی الگوهای غیرخطی پیچیده را دارد و در مقابل مشکلاتی مانند بیش‌برازش و نویز در داده‌ها مقاوم است. به همین دلیل، این روش نسبت به سایر مدل‌های متداول، عملکرد بهتری داشته است. در کنار تقویت گرادیان، جنگل تصادفی نیز دقت نسبتاً بالایی برابر با 90.72% داشت، اما میزان فراخوانی آن 19% است که نشان می‌دهد این مدل توانایی محدودی در شناسایی استارتاپ‌های موفق دارد.

در مجموع، برای پیش‌بینی موفقیت استارتاپ‌ها، الگوریتم‌های مبتنی بر Boosting بهترین گزینه محسوب می‌شوند، زیرا قادرند تعادلی بهینه میان دقت، صحت و فراخوانی ایجاد کنند و نسبت به سایر روش‌های کلاسیک، تحلیل جامع‌تری از داده‌های پیچیده ارائه دهند.

## مقایسه با تحقیقات گذشته

در تحقیق‌های پیشین مختلف، پژوهشگران به بررسی موفقیت استارتاپ‌ها با استفاده از داده‌های مختلف و الگوریتم‌های متعدد پرداخته‌اند. بسیاری از این تحقیق‌ها از داده‌های وبسایت کرانچ بیس به‌عنوان منبع اصلی استفاده کرده‌اند، که در این پژوهش نیز از داد‌های همین پایگاه داده برای پیش‌بینی موفقیت شرکت‌ها بهره گرفته شده است. با این حال، این پژوهش به‌دلیل گنجاندن ویژگی‌های صنعتی (37 ویژگی)، بررسی یک صنعت خاص و استفاده از الگوریتم‌های مختلف داده‌کاوی، تفاوت‌های عمده‌ای با دیگر پژوهش‌ها دارد. در جدول 5-1 الگوریتم‌های داده کاوی که در تحقیقات مختلف مورد بررسی قرار گرفته بود و نتایج آن‌ها آورده شده و با این پژوهش مورد مقایسه قرار گرفته است.

در پژوهش کیم و همکاران (Kim et al., 2023) نیز از داده‌های وبسایت کرانچ بیس و ویژگی‌های صنعتی برای پیش‌بینی موفقیت استارتاپ‌ها استفاده شده است. این تحقیق نشان می‌دهد که جنگل تصادفی بهترین عملکرد را در پیش‌بینی موفقیت استارتاپ‌ها داشته است و توانسته است دقت 84% را بدست آورد. یکی از نقاط قوت این تحقیق استفاده از ویژگی‌های صنعتی است که به‌طور مستقیم بر موفقیت استارتاپ‌ها تأثیر دارد و این مورد در این پژوهش نیز مدنظر قرار گرفته است.

در تحقیق تی و همکاران (Te et al., 2023)، استفاده از داده‌های لینکدین به‌عنوان منبع مکمل در کنار وبسایت کرانچ بیس باعث بهبود عملکرد مدل‌ها شده است. نتایج این تحقیق نشان می‌دهند که ترکیب این داده‌ها به‌طور قابل توجهی در دقت مدل‌ها تأثیرگذار بوده است. این تحقیق به‌خوبی نشان می‌دهد که داده‌های بیشتر و متنوع‌تر می‌توانند دقت مدل‌ها را افزایش دهند. با این حال، در پژوهش تی و همکاران، فقط به داده‌های رسانه‌ای و ویژگی‌های صنعت پرداخته شده است و تحلیل‌های جامع‌تری از تأثیرات دیگر ویژگی‌ها نظیر نوع سرمایه‌گذاران، ویژگی‌های جغرافیایی و سایر ویژگی‌ها در مدل نادیده گرفته شده است. در حالی که در این تحقیق، این ویژگی‌ها به‌طور دقیق‌تری بررسی شده و در مدل گنجانده شده‌اند.

در تحقیق دایاس و همکاران (Deias & Magrini, 2023) از رگرسیون لجستیک چندگانه برای پیش‌بینی موفقیت استارتاپ‌ها استفاده شده است و این تحقیق به‌ویژه بر تأثیر پویایی‌های تأمین مالی و تعداد دوره‌های سرمایه‌گذاری بر موفقیت استارتاپ‌ها تمرکز کرده است. این تحقیق به‌طور واضح نشان می‌دهد که تأمین مالی و مراحل مختلف آن تأثیر زیادی بر موفقیت استارتاپ‌ها دارند، که این ویژگی‌ها در این پژوهش نیز گنجانده شده است. با این حال، یکی از ضعف‌های این تحقیق عدم توجه به داده‌های دیگر مانند ویژگی‌های مدیریتی و صنعت است. در حالی که در این پژوهش، تمامی این ویژگی‌ها مورد توجه قرار گرفته و مدل پیش‌بینی موفقیت استارتاپ‌ها به‌طور کامل‌تری طراحی شده است. همچنین در این تحقیق از الگوریتم‌های مختلف داده‌کاوی برای پیش بینی موفقیت استفاده شده است.

در پژوهش راس و همکاران (Ross et al., 2021)، برای پیش‌بینی موفقیت استارتاپ‌ها از داده‌های پایگاه داده کرانچ بیس و USPTO و الگوریتم‌های داده کاوی جنگل تصادفی، یادگیری عمیق و XGBoost استفاده شده است. این مدل‌ها تا دقت 89% را در پیش‌بینی موفقیت استارتاپ‌ها داشته و ویژگی‌هایی مانند تعداد پتنت‌ها و تعداد دورهای تأمین مالی را به‌عنوان عوامل کلیدی شناسایی کرده است. در مقایسه با این تحقیق، در این پژوهش تنها از داده های پایگاه داده کرانچ بیس استفاده شده است، اما الگوریتم‌های داده‌کاوی متنوع تری برای بررسی داده‌ها استفاده شده است. همچنین داده‌های مربوط به شبکه‌های اجتماعی و ویژگی‌های صنعتی را نیز مدنظر قرار داده‌ایم. در نظر گرفتن این ویژگی‌های باعث افزایش دقت پیش‌بینی در صنعت حمل‌ونقل شده است. همچنین، این تحقیق به بررسی عوامل بیرونی مانند تأثیرات رسانه‌ای و وضعیت صنعتی پرداخته است که در مدل راس و همکاران به آن‌ها توجه نشده است.

در تحقیق ژیبوفسکی و همکاران (Żbikowski & Antosiuk, 2021) نیز از داده‌های پایگاه داده کرانچ بیس و مدل‌های یادگیری ماشین مانند رگرسیون لجستیک، ماشین بردار پشتیبان و تقویت گرادیان برای پیش‌بینی موفقیت کسب‌وکارها استفاده شد. این مدل‌ها با هدف جلوگیری از look-ahead bias طراحی شدند و مدل تقویت گرادیان بهترین عملکرد را داشته است. در مقایسه با این تحقیق، در این پژوهش علاوه بر ویژگی‌های مالی، ویژگی‌های مرتبط با شرکت مانند تعداد کارکنان، نوع صنعت و موقعیت جغرافیایی نیز در مدل وارد شده است. همچنین در این پژوهش ویژگی‌های صنعتی و رسانه‌ای را در نظر گرفته است، در حالی که تحقیق ژیبوفسکی و همکاران بیشتر بر روی ویژگی‌های مالی و موقعیت شرکت‌ها تمرکز کرده بود.

در تحقیق بنتو (Bento, 2018)، از داده‌های وبسایت کرانچ بیس استفاده شده است و الگوریتم‌های مختلفی مانند جنگل تصادفی و ماشین بردار پشتیبان برای پیش‌بینی موفقیت استارتاپ‌ها به‌کار رفته‌اند. این تحقیق نشان داد که جنگل تصادفی بهترین نتایج را در پیش‌بینی موفقیت استارتاپ‌ها دارد. با این حال، در این تحقیق به ویژگی‌های صنعت توجه نشده است و فقط به ویژگی‌های مالی و مدیریتی پرداخته شده است. این در حالی است که در این پژوهش، علاوه بر داده‌های مشابه، ویژگی‌های صنعت و تأثیرات فناوری نیز در مدل گنجانده شده است، که این باعث افزایش دقت پیش‌بینی‌ها می‌شود.

در مجموع، تحقیق حاضر در مقایسه با تحقیقات پیشین به دلیل گنجاندن ویژگی‌های متنوع‌تر، استفاده از الگوریتم‌های داده‌کاوی گوناگون و تمرکز بر یک صنعت خاص، دقت بالاتری در پیش‌بینی موفقیت استارتاپ‌ها دارد. مدل‌هایی که فقط به ویژگی‌های مالی و مدیریتی توجه دارند، ممکن است قادر به پیش‌بینی دقیق موفقیت استارتاپ‌ها نباشند چرا که موفقیت استارتاپ‌ها تحت تأثیر عوامل مختلف دیگری نیز قرار دارد. به‌طور کلی، تحقیق حاضر به دلیل توجه به تمامی این جنبه‌ها و استفاده از الگوریتم‌های پیشرفته‌تر، می‌تواند الگوریتم‌های پیش‌بینی موفقیت استارتاپ‌ها را به سطح بالاتری برساند.

در جدول 5-2 عملکرد الگوریتم‌های داده‌کاوی در کارا ترین مدل آورده شده است.

**جدول 5- 1- بررسی عملکرد الگوریتم‌های داده کاوی در پژوهش‌های مختلف**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| پژوهش | روش | دقت | صحت | بازخوانی | امتیاز F1 |
| (Kim et al., 2023) | رگرسیون لجستیک | 62% | 53.1% | 48.4% | 53.1% |
| درخت تصمیم | 82.5% | 80.2% | 80.2% | 80.2% |
| جنگل تصادفی | 84% | 80.4% | 73.9% | 80.4% |
| تقویت گرادیان | 80.8% | 75.6% | 67% | 75.6% |
| ماشین بردار پشتیبان | 82.8% | 77.6% | 67.1% | 77.6% |
| بیز ساده | 58.9% | 19.9% | 11.5% | 19.9% |
| (Potanin et al., 2023) | یادگیری عمیق | 92% | - | 64% | - |
| (Te et al., 2023) | رگرسیون لجستیک | 79% | 78% | 79% | - |
| (Deias & Magrini, 2023) | رگرسیون لجستیک  چندگانه | 64% | 74% | 64% | - |
| (Dellermann et al., 2021) | XGBoost | 92.6% | 80% | 39.33% | 45.71% |
| (Żbikowski & Antosiuk, 2021) | ماشین بردار پشتیبان | 87% | 86% | 20% | 32% |
| (Ross et al., 2021) | XGBoost | 89.4% | 84.7% | 75.5% | 78.3% |
| (Arroyo et al., 2019) | تقویت گرادیان | 83.6% | 44% | 30% | 37% |
| (Bento, 2018) | جنگل تصادفی | 93.1% | 93.3% | 93.2% | 93.2% |
| (Krishna et al., 2016) | جنگل تصادفی | 88.7% | 88.4% | 88.8% | 88.6% |
| پژوهش حاضر | تقویت گرادیان | 91.44% | 80.86% | 39.33% | 45.18% |

## نوآوری پژوهش

در این پژوهش، با بهره‌گیری از روش‌های یادگیری ماشین و تحلیل داده‌های گسترده، مدلی جامع برای پیش‌بینی موفقیت استارتاپ‌های فعال در صنعت حمل‌ونقل ارائه شده است. این مدل با بررسی متغیرهای کلیدی، مانند میزان سرمایه‌گذاری، تعداد بنیان‌گذاران، نوع سرمایه‌گذاران، حضور در شبکه‌های اجتماعی، ثبت اختراعات و غیره، تلاش دارد تا تصویری دقیق‌تر و جامع‌تر از عوامل مؤثر بر موفقیت استارتاپ‌ها ارائه دهد. در مقایسه با تحقیقات پیشین، این پژوهش دارای چندین جنبه نوآورانه است که آن را از سایر مطالعات مشابه متمایز می‌سازد.

یکی از مهم‌ترین نوآوری‌های این پژوهش، توسعه یک مدل چندبعدی و جامع برای پیش‌بینی موفقیت استارتاپ‌ها است. برخلاف تحقیقات گذشته که عمدتاً بر ویژگی‌های مالی و تعداد سرمایه‌گذاری‌ها تمرکز داشته‌اند، این پژوهش با در نظر گرفتن مجموعه‌ای گسترده از متغیرهای مرتبط با صنعت حمل‌ونقل، بنیان‌گذاران، ساختار سرمایه‌گذاری و تحلیل داده‌های اجتماعی، رویکردی نوین را در پیش گرفته است. همچنین، تأثیر متغیرهایی نظیر حوزه تخصصی شرکت‌ها، میزان حضور در شبکه‌های اجتماعی، و روندهای سرمایه‌گذاری در مراحل مختلف رشد استارتاپ به طور ویژه در این مدل لحاظ شده است.

همچنین، این پژوهش به طور خاص به تأثیر ویژگی‌های صنعت حمل‌ونقل و روندهای سرمایه‌گذاری در موفقیت استارتاپ‌ها پرداخته است. تحلیل روندهای سرمایه‌گذاری در این صنعت نشان می‌دهد که برخی از حوزه‌های فناوری، مانند خودروهای خودران، لجستیک هوشمند، و بهینه‌سازی ناوگان، دارای نرخ موفقیت بالاتری هستند. این یافته‌ها نشان می‌دهند که سرمایه‌گذاران باید هنگام انتخاب استارتاپ‌های مناسب، نه‌تنها به معیارهای مالی، بلکه به ویژگی‌های صنعتی و سطح بلوغ فناوری نیز توجه داشته باشند.

همچنین، این پژوهش توانسته است یک ابزار تصمیم‌گیری کاربردی برای سرمایه‌گذاران خطرپذیر ارائه دهد. مدل توسعه‌یافته در این تحقیق قابلیت این را دارد که به عنوان یک ابزار کمک‌کننده برای انتخاب استارتاپ‌های با پتانسیل بالا و کاهش ریسک سرمایه‌گذاری مورد استفاده قرار گیرد. برخلاف مطالعات قبلی که تنها به تحلیل آماری داده‌ها پرداخته‌اند، این پژوهش مدل‌هایی را توسعه داده است که می‌توانند به صورت عملی در تصمیم‌گیری‌های سرمایه‌گذاران و سیاست‌گذاران استفاده شوند.

در مجموع، این پژوهش با ارائه مدلی جامع و چندبعدی، ترکیب داده‌های مختلف، استفاده از تکنیک‌های پیشرفته یادگیری ماشین، تحلیل حساسیت و تفسیر مدل‌ها، و ارائه یک ابزار کاربردی برای سرمایه‌گذاران، نوآوری‌های متعددی را در زمینه پیش‌بینی موفقیت استارتاپ‌های حوزه حمل‌ونقل معرفی کرده است. این تحقیق نه‌تنها در سطح تئوریک، بلکه در سطح عملی نیز می‌تواند تأثیر بسزایی در بهبود روندهای سرمایه‌گذاری در این صنعت داشته باشد و به‌عنوان یک مدل مرجع برای مطالعات آینده مورد استفاده قرار گیرد.

## بحث

بر اساس نتایج این تحقیق، چندین ویژگی مختلف از جمله تعداد کارکنان، میزان سرمایه جذب‌شده، وضعیت عملیاتی و ویژگی‌های جغرافیایی بیشترین تأثیر را بر موفقیت استارتاپ‌ها در صنعت حمل‌ونقل داشته‌اند. به‌طور خاص، تعداد کارکنان به‌عنوان یکی از ویژگی‌های کلیدی در تمامی مدل‌ها شناخته شد. این یافته نشان می‌دهد که استارتاپ‌هایی که تیم‌های بزرگ‌تر و تخصصی‌تری دارند، معمولاً شانس بیشتری برای موفقیت دارند. این نتیجه با بسیاری از پژوهش‌های پیشین همسو است که نشان داده‌اند تیم‌های با تجربه و توانمند می‌توانند چالش‌های بیشتری را در بازارهای رقابتی به خوبی مدیریت کنند.

در کنار تعداد کارکنان، میزان سرمایه جذب‌شده نیز به‌طور مداوم در تمامی مدل‌ها به‌عنوان یک ویژگی برجسته در پیش‌بینی موفقیت استارتاپ‌ها شناخته شد. استارتاپ‌هایی که قادر به جذب سرمایه بیشتر و در مراحل اولیه تأسیس هستند، معمولاً شانس بالاتری برای رشد و بقا دارند. این یافته نیز با نتایج (Kim et al., 2023) و (Bento, 2018) تطابق دارد که نشان داده‌اند سرمایه‌گذاری‌های اولیه یکی از مهم‌ترین عوامل موفقیت استارتاپ‌ها است، همخوانی دارد.

در تحقیق حاضر، ویژگی‌های صنعتی و فناوری نیز تأثیر زیادی بر موفقیت استارتاپ‌ها داشتند. از جمله ویژگی‌های صنعتی که در مدل‌ها تأثیرگذار بودند، موقعیت جغرافیایی و نوع صنعت بود. استارتاپ‌هایی که در مناطق جغرافیایی با اکوسیستم‌های قوی‌تر و نیاز بیشتر به فناوری‌های نوین فعالیت می‌کنند، شانس بالاتری برای موفقیت دارند. این یافته تأکید دارد که موقعیت مکانی و نوع صنعت می‌تواند به‌طور مستقیم بر موفقیت استارتاپ‌ها تأثیر بگذارد. در صنایع خاصی مانند حمل‌ونقل، که نیاز به فناوری‌های پیشرفته و سرمایه‌گذاری‌های بزرگ دارند، انتخاب صحیح موقعیت جغرافیایی و توجه به روندهای صنعتی می‌تواند تأثیر زیادی در موفقیت شرکت داشته باشد.

ویژگی‌های فناوری مانند تعداد پتنت‌ها و تعداد فناوری‌های فعال نیز در این تحقیق نقش مهمی ایفا کردند. استارتاپ‌هایی که در زمینه نوآوری و ثبت اختراعات فعال‌تر بودند، شانس بیشتری برای جذب سرمایه و موفقیت در بازار داشتند. این نتیجه نشان می‌دهد که نوآوری و توانایی در توسعه فناوری‌های جدید می‌تواند مزیت رقابتی بزرگی برای استارتاپ‌ها ایجاد کند و به آن‌ها کمک کند تا در بازارهای پیچیده و رقابتی صنعت حمل‌ونقل موفق شوند.

یکی از نکات جالب توجه این تحقیق، تأثیر کم شبکه‌های اجتماعی بر موفقیت استارتاپ‌ها است. برخلاف بسیاری از پژوهش‌های پیشین که نشان داده‌اند فعالیت در شبکه‌های اجتماعی و تعاملات رسانه‌ای می‌توانند تأثیر زیادی در جذب سرمایه‌گذاران و مشتریان داشته باشند، در این تحقیق نتایج نشان دادند که شبکه‌های اجتماعی تأثیر کمی بر موفقیت استارتاپ‌ها در صنعت حمل‌ونقل داشته‌اند. این یافته ممکن است به دلیل ماهیت B2B استارتاپ‌های حوزه حمل‌ونقل باشد. در صنایع B2B، تصمیم‌گیری‌ها معمولاً بر اساس نیازهای تخصصی و قراردادهای بزرگ‌تر با شرکت‌های دیگر انجام می‌شود و کمتر تحت تأثیر تبلیغات و فعالیت‌های رسانه‌ای قرار می‌گیرند. در مقابل، در صنایع B2C، فعالیت‌های رسانه‌ای و اجتماعی معمولاً نقش بیشتری در جذب مشتری و موفقیت شرکت دارند. به‌طور خلاصه، در صنعت حمل‌ونقل که بیشتر استارتاپ‌ها در فضای B2B فعالیت دارند، ویژگی‌های مالی، فنی و شبکه‌های سرمایه‌گذاری اهمیت بیشتری دارند تا فعالیت در شبکه‌های اجتماعی.

در این تحقیق، از الگوریتم‌های مختلف یادگیری ماشین برای پیش‌بینی موفقیت استارتاپ‌ها استفاده شد. پس از ارزیابی دقت، صحت، بازخوانی و امتیاز F1 هر مدل، نتایج نشان داد که الگوریتم تقویت گرادیان بهترین عملکرد را در میان سایر الگوریتم‌ها داشته است (جدول 4-9). این مدل توانست از تقویت تدریجی[[64]](#footnote-64) درخت‌ها استفاده کرده و مدل‌هایی با دقت بالا بسازد. یکی از دلایل موفقیت این الگوریتم، توانایی آن در مدیریت ویژگی‌های پیچیده و غیرخطی است. این ویژگی باعث می‌شود که الگوریتم تقویت گرادیان قادر به شناسایی روابط پیچیده‌تری بین ویژگی‌ها و پیش‌بینی دقیق‌تری در زمینه‌های پیچیده‌ای مانند پیش بینی موفقیت کسب و کارها باشد.

## محدودیت‌های پژوهش

در این تحقیق، محدودیت‌های مختلفی وجود دارد که می‌تواند بر نتایج حاصل از این تحقیق نتیجه گذارد. یکی از بزرگ‌ترین محدودیت‌ها کمبود داده‌های دقیق و جامع در حوزه هوش مصنوعی در صنعت حمل‌ونقل است. داده‌های موجود در پایگاه داده کرانچ‌بیس برای استارتاپ‌های فعال در این صنعت محدود است، و این کمبود داده‌ها باعث شده که مدل پیشنهادی در برخی از موارد با دقت کمتری روبه‌رو شود. برای مثال، مدل‌های شبکه‌های عصبی برای تعداد کم داده‌ها نمی‌تواند جواب دقیقی بدهد. برای تحلیل دقیق‌تر و پیش‌بینی بهتر نتایج، به داده‌های بیشتری نیاز است که این مسئله یکی از چالش‌های اصلی در این تحقیق بوده است.

یکی دیگر از محدودیت‌ها، عدم تعادل در داده‌ها است. بسیاری از ویژگی‌های داده‌ای مورد استفاده در مدل، به‌ویژه ویژگی‌های مربوط به شرکت‌ها، تعداد ناکافی نمونه‌ها را دارند. این مشکل می‌تواند باعث کاهش دقت مدل‌های یادگیری ماشین شود، چرا که الگوریتم‌ها با داده‌های نامتوازن مواجه می‌شوند. این چالش در پژوهش‌های پیشین نیز مشاهده شده و معمولاً با استفاده از تکنیک‌هایی مانند وزن‌دهی به داده‌ها یا بالانس داده‌ها سعی در حل آن شده است.

محدودیت دیگری که در این تحقیق وجود دارد، عدم توجه به متغیرهای خارجی است که می‌توانند تأثیر زیادی بر موفقیت استارتاپ‌ها داشته باشند. این تحقیق بیشتر بر داده‌های داخلی استارتاپ‌ها و ویژگی‌های مربوط به آن‌ها تمرکز کرده است، اما عواملی مانند روندهای جهانی یا تأثیرات سیاسی و اقتصادی می‌توانند به‌شدت بر موفقیت یا شکست استارتاپ‌ها تأثیر بگذارند. این جنبه‌ها در مدل مورد نظر لحاظ نشده‌اند که می‌تواند منجر به نادیده‌گرفتن برخی از عوامل کلیدی برای پیش‌بینی موفقیت استارتاپ‌ها شود.

در نهایت، این تحقیق به طور خاص به پیش‌بینی موفقیت استارتاپ‌ها در صنعت حمل‌ونقل پرداخته است و ممکن است نتواند به طور دقیق برای تمام صنایع یا بخش‌های دیگر به کار گرفته شود. هر صنعت ویژگی‌های خاص خود را دارد که باید در مدل‌های پیش‌بینی موفقیت در نظر گرفته شوند و این مسئله یکی از محدودیت‌های این تحقیق است.

## پیشنهادات برای تحقیقات آتی

با توجه به محدودیت‌های موجود در این تحقیق، پیشنهاد می‌شود که پژوهش‌های بعدی به جمع‌آوری داده‌های بیشتر و گسترش دامنه تحقیق بپردازند. یکی از مشکلات اصلی در این تحقیق کمبود داده‌های دقیق و جامع در زمینه هوش مصنوعی در صنعت حمل‌ونقل است. برای تحقیق‌های آینده، پیشنهاد می‌شود که از مجموعه داده‌های متنوع‌تری از جمله داده‌های جهانی و پلتفرم‌های دیگر مانند لینکدین و TechCrunch استفاده شود تا دقت پیش‌بینی‌ها افزایش یابد.

یکی دیگر از پیشنهادات، استفاده از مدل‌های هیبریدی است که می‌تواند ترکیبی از روش‌های مختلف یادگیری ماشین را به کار گیرد. مدل‌های XGBoost و جنگل تصادفی در این تحقیق عملکرد خوبی داشتند، اما ترکیب این مدل‌ها با الگوریتم‌های شبکه‌های عصبی یا الگوریتم‌های تقویتی ممکن است دقت مدل را بیشتر کند. همچنین، توجه به الگوهای پیچیده‌تر و مدل‌های پیشرفته‌تر مانند یادگیری عمیق در تحلیل داده‌ها و پیش‌بینی موفقیت استارتاپ‌ها می‌تواند باعث بهبود قابل توجهی در نتایج شود.

یکی از مسائل دیگر که در این تحقیق نادیده گرفته شد، تأثیر متغیرهای خارجی مانند شرایط اقتصادی و روندهای جهانی بر موفقیت استارتاپ‌ها بود. به‌طور خاص، بررسی تأثیر بحران‌های اقتصادی، تغییرات سیاست‌های دولتی و تحولات بازار جهانی بر موفقیت استارتاپ‌ها می‌تواند به توسعه مدل‌های دقیق‌تر کمک کند. همچنین، پیشنهاد می‌شود که در تحقیقات آینده، مقیاس زمانی گسترده‌تری مورد بررسی قرار گیرد تا نتایج دقیق‌تری در خصوص موفقیت یا شکست استارتاپ‌ها در طول زمان به‌دست آید.

در نهایت، مقایسه بین صنایع مختلف و تحلیل ویژگی‌های خاص هر صنعت می‌تواند منجر به ایجاد مدل‌های پیش‌بینی بهتر برای هر صنعت خاص شود. به‌طور مثال، صنعت حمل‌ونقل ممکن است ویژگی‌های خاص خود را داشته باشد که در سایر صنایع مشاهده نمی‌شود. از این‌رو، توصیه می‌شود که تحقیقات آینده مدل‌هایی را برای پیش‌بینی موفقیت استارتاپ‌ها در صنایع مختلف طراحی کنند و ویژگی‌های خاص هر صنعت را به‌طور دقیق در نظر بگیرند.

اجرای این پیشنهادات می‌تواند به ارتقای دقت مدل‌های پیش‌بینی موفقیت استارتاپ‌ها کمک کند و امکان تحلیل بهتر عواملی که بر موفقیت یا شکست استارتاپ‌ها تأثیر می‌گذارند را فراهم آورد.

## پیشنهادات کاربردی

نتایج این تحقیق می‌تواند تأثیرات مهمی بر تصمیم‌گیری مدیران، سرمایه‌گذاران و فعالان حوزه حمل‌ونقل داشته باشد. یکی از کاربردهای کلیدی این پژوهش در سرمایه‌گذاری و ارزیابی استارتاپ‌های حمل‌ونقل است. سرمایه‌گذاران می‌توانند با استفاده از مدل‌های داده‌کاوی معرفی‌شده در این پژوهش، ریسک‌های سرمایه‌گذاری را کاهش داده و استارتاپ‌های با احتمال موفقیت بالا را شناسایی کنند. این موضوع به‌ویژه برای شرکت‌های سرمایه‌گذاری خطرپذیر و صندوق‌های تأمین مالی استارتاپی اهمیت دارد. علاوه بر این، شرکت‌های حمل‌ونقل سنتی نیز می‌توانند از این مدل‌ها برای شناسایی فرصت‌های نوآوری استفاده کنند. با تحلیل روندهای موجود در بازار و بررسی عملکرد استارتاپ‌های موفق، شرکت‌های قدیمی می‌توانند با سرمایه‌گذاری در فناوری‌های نوین مانند هوش مصنوعی و لجستیک هوشمند، مزیت رقابتی خود را حفظ کنند. همچنین، مدیران استارتاپ‌های حوزه حمل‌ونقل می‌توانند از یافته‌های این پژوهش برای بهینه‌سازی استراتژی‌های رشد و توسعه کسب‌وکار خود استفاده کنند. مدل‌های معرفی‌شده می‌توانند به آن‌ها کمک کنند تا عوامل کلیدی موفقیت را شناسایی کرده و تصمیمات مبتنی بر داده بگیرند. این موضوع به‌ویژه در حوزه‌هایی مانند خودروهای اشتراکی، حمل‌ونقل هوشمند و تحویل کالا مبتنی بر هوش مصنوعی بسیار کاربردی است. علاوه بر این، نهادهای دولتی و تنظیم‌گران صنعت حمل‌ونقل نیز می‌توانند از این پژوهش برای بررسی تأثیر استارتاپ‌های حمل‌ونقل بر زیرساخت‌های شهری و توسعه پایدار بهره ببرند. استفاده از مدل‌های پیش‌بینی موفقیت استارتاپ‌ها می‌تواند به سیاست‌گذاران کمک کند تا برنامه‌های حمایتی هدفمندتری برای کسب‌وکارهای نوپا تدوین کنند و منابع را به شکلی کارآمدتر تخصیص دهند. در مجموع، این پژوهش می‌تواند به تحول دیجیتال در صنعت حمل‌ونقل، کاهش هزینه‌ها، افزایش کارایی و بهبود کیفیت خدمات برای مشتریان نهایی کمک کند و از این طریق، تأثیرات گسترده‌ای در کل زنجیره تأمین حمل‌ونقل ایجاد نماید.

## جمع بندی

این پژوهش به بررسی موفقیت استارتاپ‌های فعال در صنعت حمل‌ونقل پرداخته و با استفاده از الگوریتم‌های یادگیری ماشین، مدلی جامع برای پیش‌بینی میزان موفقیت این شرکت‌ها ارائه داده است. در این راستا، با تجزیه و تحلیل داده‌های متنوع شامل ویژگی‌های مالی، سرمایه‌گذاری، تیم مدیریتی و فناوری‌های مورد استفاده، عوامل کلیدی موفقیت شناسایی شده‌اند. نتایج پژوهش نشان داد که میزان سرمایه جذب‌شده، تعداد کارکنان، وضعیت عملیاتی و تعداد پتنت‌های ثبت‌شده از مهم‌ترین متغیرهایی هستند که بر موفقیت استارتاپ‌ها تأثیر می‌گذارند. همچنین، مشخص شد که ویژگی‌های جغرافیایی و حضور در اکوسیستم‌های استارتاپی نیز نقش مهمی در پیش‌بینی موفقیت این شرکت‌ها ایفا می‌کنند.

بررسی مدل‌های مختلف یادگیری ماشین نشان داد که تقویت گرادیان با دقت 91.44% بهترین عملکرد را در پیش‌بینی موفقیت استارتاپ‌های حوزه حمل‌ونقل داشته است. این الگوریتم توانسته است با ترکیب چندین مدل پایه، دقت پیش‌بینی را افزایش داده و روابط پیچیده میان متغیرهای مختلف را شناسایی کند. علاوه بر این، مقایسه نتایج پژوهش حاضر با تحقیقات پیشین نشان داد که استفاده از داده‌های گسترده‌تر و ترکیب ویژگی‌های صنعتی و مالی می‌تواند دقت مدل‌های پیش‌بینی را بهبود بخشد. بسیاری از مطالعات گذشته تنها به ویژگی‌های مالی و سرمایه‌گذاری پرداخته بودند، در حالی که این تحقیق تأثیر ویژگی‌های فناوری، اکوسیستم صنعتی و روندهای سرمایه‌گذاری را نیز بررسی کرده است.

این پژوهش دارای چندین نوآوری مهم بوده است، از جمله طراحی یک مدل جامع و چندبعدی که نه‌تنها ویژگی‌های مالی، بلکه متغیرهای مرتبط با صنعت، تیم مدیریتی، نوآوری و سرمایه‌گذاری را نیز در نظر گرفته است. همچنین، ارائه یک ابزار تصمیم‌گیری برای سرمایه‌گذاران از دیگر دستاوردهای این تحقیق بوده است که می‌تواند در کاهش ریسک‌های سرمایه‌گذاری و انتخاب استارتاپ‌های با پتانسیل بالا مؤثر باشد. علاوه بر این، تحلیل روندهای سرمایه‌گذاری در صنعت حمل‌ونقل نشان داد که استارتاپ‌های فعال در حوزه‌هایی مانند لجستیک هوشمند و خودروهای خودران، نرخ موفقیت بالاتری نسبت به سایر بخش‌های این صنعت دارند.

با این حال، این تحقیق دارای محدودیت‌هایی نیز بوده است، از جمله کمبود داده‌های جامع برای برخی از استارتاپ‌ها، که می‌تواند بر دقت مدل تأثیر بگذارد. همچنین، برخی متغیرهای کلیدی مانند تأثیرات اقتصادی کلان و سیاست‌های دولتی در مدل لحاظ نشده‌اند که می‌توانند بر موفقیت استارتاپ‌ها تأثیرگذار باشند. برای تحقیقات آینده، پیشنهاد می‌شود که از مجموعه داده‌های گسترده‌تری استفاده شود و همچنین، تأثیر متغیرهای خارجی مانند شرایط اقتصادی و روندهای جهانی نیز در مدل‌های پیش‌بینی گنجانده شوند.

در نهایت، این پژوهش نشان داد که استفاده از روش‌های داده‌کاوی و یادگیری ماشین در کنار تحلیل جامع متغیرهای کلیدی، می‌تواند ابزارهای مؤثری برای سرمایه‌گذاران، کارآفرینان و سیاست‌گذاران فراهم کند. یافته‌های این تحقیق می‌توانند به تصمیم‌گیری‌های هوشمندانه‌تر کمک کرده و راهکارهای بهتری برای سرمایه‌گذاری در استارتاپ‌های حوزه حمل‌ونقل و هوش مصنوعی ارائه دهند. از این رو، این پژوهش می‌تواند به‌عنوان یک مدل مرجع برای مطالعات آتی در حوزه پیش‌بینی موفقیت استارتاپ‌ها مورد استفاده قرار گیرد.

مراجع

مراجع

1. Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability*, *11*(1), 189.
2. Akhtar, M., & Moridpour, S. (2021). A review of traffic congestion prediction using artificial intelligence. *Journal of Advanced Transportation*, *2021*, 1-18.
3. Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.
4. Arroyo, J., Corea, F., Jimenez-Diaz, G., & Recio-Garcia, J. A. (2019). Assessment of machine learning performance for decision support in venture capital investments. *Ieee Access*, *7*, 124233-124243.
5. Banister, D., & Berechman, Y. (2001). Transport investment and the promotion of economic growth. *Journal of transport geography*, *9*(3), 209-218.
6. Behrooz, H., & Hayeri, Y. M. (2022). Machine Learning Applications in Surface Transportation Systems: A Literature Review. *Applied Sciences*, *12*(18), 9156.
7. Bento, F. R. d. S. R. (2017). *Predicting start-up success with machine learning* Universidade NOVA de Lisboa (Portugal)].
8. Bento, F. R. d. S. R. (2018). *Predicting start-up success with machine learning*
9. Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.
10. Chen, L., & Englund, C. (2015). Cooperative intersection management: A survey. *IEEE transactions on intelligent transportation systems*, *17*(2), 570-586.
11. Deias, A., & Magrini, A. (2023). The impact of equity funding dynamics on venture success: an empirical analysis based on Crunchbase data. *Economies*, *11*(1), 19.
12. Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2021). Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method. *arXiv preprint arXiv:2105.03360*.
13. Ezangina, I., & Malovichko, A. (2021). The venture capital market in a pandemic: Realities of time and future prospects. *Finance: Theory and Practice*, *25*(5), 92-116.
14. García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer.
15. Glasner, J. (2023). *Investors Have Stopped Feeding The Supply Chain*.
16. Goodfellow, I. (2016). Deep learning. In: MIT press.
17. Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157-1182.
18. Hagman, J. (2024). Unicorn IPOs: Underpricing, short-and long-term performance in the United States and Europe.
19. Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan kaufmann*, *340*, 94104-93205.
20. Hasan, U., Whyte, A., & Al Jassmi, H. (2019). A review of the transformation of road transport systems: are we ready for the next step in artificially intelligent sustainable Transport? *Applied System Innovation*, *3*(1), 1.
21. Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction. In: Springer.
22. Hughes, P., & Morgan, R. E. (2007). A resource-advantage perspective of product–market strategy performance & strategic capital in high technology firms. *Industrial Marketing Management*, *36*(4), 503-517.
23. James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.
24. Johnson, M. J. (2018). *Using Machine Learning Algorithms to Predict Technology-Based Small Business Commercialization* The George Washington University].
25. Khan, S., Adnan, A., & Iqbal, N. (2022). Applications of Artificial Intelligence in Transportation. 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET),
26. Kim, J., Kim, H., & Geum, Y. (2023). How to succeed in the market? Predicting startup success using a machine learning approach. *Technological Forecasting and Social Change*, *193*, 122614.
27. Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Morgan Kaufman Publishing*.
28. Krishna, A., Agrawal, A., & Choudhary, A. (2016). Predicting the outcome of startups: less failure, more success. 2016 IEEE 16th international conference on data mining workshops (ICDMW),
29. Kuhn, M. (2013). Applied predictive modeling. In: Springer.
30. Lee, D., & Lin, K.-C. (2020). How to transform sustainable energy technology into a unicorn start-up: technology review and case study. *Sustainability*, *12*(7), 3018.
31. Litman, T. (2017). Autonomous vehicle implementation predictions.
32. Liu, Y., Li, B., Yang, S., & Li, Z. (2024). Handling missing values and imbalanced classes in machine learning to predict consumer preference: Demonstrations and comparisons to prominent methods. *Expert Systems with Applications*, *237*, 121694.
33. Nguyen, H., Kieu, L. M., Wen, T., & Cai, C. (2018). Deep learning methods in transportation domain: a review. *IET Intelligent Transport Systems*, *12*(9), 998-1004.
34. Nikitas, A., Michalakopoulou, K., Njoya, E. T., & Karampatzakis, D. (2020). Artificial intelligence, transport and the smart city: Definitions and dimensions of a new mobility era. *Sustainability*, *12*(7), 2789.
35. Okrepilov, V. V., Kovalenko, B. B., Getmanova, G. V., & Turovskaj, M. S. (2022). Modern trends in artificial intelligence in the transport system. *Transportation Research Procedia*, *61*, 229-233.
36. Olugbade, S., Ojo, S., Imoize, A. L., Isabona, J., & Alaba, M. O. (2022). A review of artificial intelligence and machine learning for incident detectors in road transport systems. *Mathematical and Computational Applications*, *27*(5), 77.
37. Potanin, M., Chertok, A., Zorin, K., & Shtabtsovsky, C. (2023). Startup success prediction and VC portfolio simulation using CrunchBase data. *arXiv preprint arXiv:2309.15552*.
38. Prohorovs, A., Bistrova, J., & Ten, D. (2019). Startup success factors in the capital attraction stage: Founders’ perspective. *Journal of east-west business*, *25*(1), 26-51.
39. RAMEY, J. (2022). *Here’s Why Argo AI Is Shutting Down*.
40. Ross, G., Das, S., Sciro, D., & Raza, H. (2021). CapitalVX: A machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science*, *7*, 94-114.
41. Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.
42. Schwarting, W., Alonso-Mora, J., & Rus, D. (2018). Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, *1*(1), 187-210.
43. Sharchilev, B., Roizner, M., Rumyantsev, A., Ozornin, D., Serdyukov, P., & de Rijke, M. (2018). Web-based startup success prediction. Proceedings of the 27th ACM international conference on information and knowledge management,
44. Simon, J. P. (2016). How to catch a unicorn. *An exploration of the universe of tech companies with high*.
45. Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction MIT Press. *Cambridge, MA*, *22447*, 10.
46. Te, Y.-F., Wieland, M., Frey, M., Pyatigorskaya, A., Schiffer, P., & Grabner, H. (2023). Making it into a successful series a funding: An analysis of crunchbase and linkedin data. *The Journal of Finance and Data Science*, *9*, 100099.
47. Weking, J., Böttcher, T. P., Hermes, S., & Hein, A. (2019). Does business model matter for startup success? A quantitative analysis.
48. Whig, P., Velu, A., Nadikattu, R. R., & Alkali, Y. J. (2024). Role of AI and IoT in Intelligent Transportation. In *Artificial Intelligence for Future Intelligent Transportation* (pp. 199-220). Apple Academic Press.
49. Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining,
50. Witten, I. H., Frank, E., & Hall, M. A. (2011). What’s it all about. In *Data mining: Practical machine learning tools and techniques* (pp. 338). Morgan Kaufmann.
51. Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart transportation. *Future Internet*, *11*(4), 94.
52. Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. *Information Processing & Management*, *58*(4), 102555.
53. Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107-115.

پيوست‌ها

پيوست الف

**Abstract:**

Abstract …

**Keywords:**

Iran University of Science and Technology

School of Industrial Engineering

Thesis Title

A Thesis Submitted in Partial Fulfillment of the Requirement for the Degree of Master of Science (Doctor of Philosophy) in -------

By:

Student Name

Supervisor:

Dr. -------------

Advisor:

Dr. -------------

Dr. -------------

December 2021

Abduljabbar, R., Dia, H., Liyanage, S., & Bagloee, S. A. (2019). Applications of artificial intelligence in transport: An overview. *Sustainability*, *11*(1), 189.

Akhtar, M., & Moridpour, S. (2021). A review of traffic congestion prediction using artificial intelligence. *Journal of Advanced Transportation*, *2021*, 1-18.

Alpaydin, E. (2020). *Introduction to machine learning*. MIT press.

Arroyo, J., Corea, F., Jimenez-Diaz, G., & Recio-Garcia, J. A. (2019). Assessment of machine learning performance for decision support in venture capital investments. *Ieee Access*, *7*, 124233-124243.

Banister, D., & Berechman, Y. (2001). Transport investment and the promotion of economic growth. *Journal of transport geography*, *9*(3), 209-218.

Behrooz, H., & Hayeri, Y. M. (2022). Machine Learning Applications in Surface Transportation Systems: A Literature Review. *Applied Sciences*, *12*(18), 9156.

Bento, F. R. d. S. R. (2017). *Predicting start-up success with machine learning* Universidade NOVA de Lisboa (Portugal)].

Bento, F. R. d. S. R. (2018). *Predicting start-up success with machine learning*

Bishop, C. M., & Nasrabadi, N. M. (2006). *Pattern recognition and machine learning* (Vol. 4). Springer.

Chen, L., & Englund, C. (2015). Cooperative intersection management: A survey. *IEEE transactions on intelligent transportation systems*, *17*(2), 570-586.

Deias, A., & Magrini, A. (2023). The impact of equity funding dynamics on venture success: an empirical analysis based on Crunchbase data. *Economies*, *11*(1), 19.

Dellermann, D., Lipusch, N., Ebel, P., Popp, K. M., & Leimeister, J. M. (2021). Finding the unicorn: Predicting early stage startup success through a hybrid intelligence method. *arXiv preprint arXiv:2105.03360*.

Ezangina, I., & Malovichko, A. (2021). The venture capital market in a pandemic: Realities of time and future prospects. *Finance: Theory and Practice*, *25*(5), 92-116.

García, S., Luengo, J., & Herrera, F. (2015). *Data preprocessing in data mining* (Vol. 72). Springer.

Glasner, J. (2023). *Investors Have Stopped Feeding The Supply Chain*.

Goodfellow, I. (2016). Deep learning. In: MIT press.

Guyon, I., & Elisseeff, A. (2003). An introduction to variable and feature selection. *Journal of machine learning research*, *3*(Mar), 1157-1182.

Hagman, J. (2024). Unicorn IPOs: Underpricing, short-and long-term performance in the United States and Europe.

Han, J., Kamber, M., & Mining, D. (2006). Concepts and techniques. *Morgan kaufmann*, *340*, 94104-93205.

Hasan, U., Whyte, A., & Al Jassmi, H. (2019). A review of the transformation of road transport systems: are we ready for the next step in artificially intelligent sustainable Transport? *Applied System Innovation*, *3*(1), 1.

Hastie, T. (2009). The elements of statistical learning: data mining, inference, and prediction. In: Springer.

Hughes, P., & Morgan, R. E. (2007). A resource-advantage perspective of product–market strategy performance & strategic capital in high technology firms. *Industrial Marketing Management*, *36*(4), 503-517.

James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An introduction to statistical learning* (Vol. 112). Springer.

Khan, S., Adnan, A., & Iqbal, N. (2022). Applications of Artificial Intelligence in Transportation. 2022 International Conference on Electrical, Computer and Energy Technologies (ICECET),

Kim, J., Kim, H., & Geum, Y. (2023). How to succeed in the market? predicting startup success using a machine learning approach. *Technological Forecasting and Social Change*, *193*, 122614.

Kohavi, R. (1995). A study of cross-validation and bootstrap for accuracy estimation and model selection. *Morgan Kaufman Publishing*.

Krishna, A., Agrawal, A., & Choudhary, A. (2016). Predicting the outcome of startups: less failure, more success. 2016 IEEE 16th international conference on data mining workshops (ICDMW),

Kuhn, M. (2013). Applied predictive modeling. In: Springer.

Lee, D., & Lin, K.-C. (2020). How to transform sustainable energy technology into a unicorn start-up: technology review and case study. *Sustainability*, *12*(7), 3018.

Litman, T. (2017). Autonomous vehicle implementation predictions.

Liu, Y., Li, B., Yang, S., & Li, Z. (2024). Handling missing values and imbalanced classes in machine learning to predict consumer preference: Demonstrations and comparisons to prominent methods. *Expert Systems with Applications*, *237*, 121694.

Nguyen, H., Kieu, L. M., Wen, T., & Cai, C. (2018). Deep learning methods in transportation domain: a review. *IET Intelligent Transport Systems*, *12*(9), 998-1004.

Nikitas, A., Michalakopoulou, K., Njoya, E. T., & Karampatzakis, D. (2020). Artificial intelligence, transport and the smart city: Definitions and dimensions of a new mobility era. *Sustainability*, *12*(7), 2789.

Okrepilov, V. V., Kovalenko, B. B., Getmanova, G. V., & Turovskaj, M. S. (2022). Modern trends in artificial intelligence in the transport system. *Transportation Research Procedia*, *61*, 229-233.

Olugbade, S., Ojo, S., Imoize, A. L., Isabona, J., & Alaba, M. O. (2022). A review of artificial intelligence and machine learning for incident detectors in road transport systems. *Mathematical and Computational Applications*, *27*(5), 77.

Potanin, M., Chertok, A., Zorin, K., & Shtabtsovsky, C. (2023). Startup success prediction and VC portfolio simulation using CrunchBase data. *arXiv preprint arXiv:2309.15552*.

Prohorovs, A., Bistrova, J., & Ten, D. (2019). Startup success factors in the capital attraction stage: Founders’ perspective. *Journal of east-west business*, *25*(1), 26-51.

RAMEY, J. (2022). *Here’s Why Argo AI Is Shutting Down*.

Ross, G., Das, S., Sciro, D., & Raza, H. (2021). CapitalVX: A machine learning model for startup selection and exit prediction. *The Journal of Finance and Data Science*, *7*, 94-114.

Russell, S. J., & Norvig, P. (2016). *Artificial intelligence: a modern approach*. Pearson.

Schwarting, W., Alonso-Mora, J., & Rus, D. (2018). Planning and decision-making for autonomous vehicles. *Annual Review of Control, Robotics, and Autonomous Systems*, *1*(1), 187-210.

Sharchilev, B., Roizner, M., Rumyantsev, A., Ozornin, D., Serdyukov, P., & de Rijke, M. (2018). Web-based startup success prediction. Proceedings of the 27th ACM international conference on information and knowledge management,

Simon, J. P. (2016). How to catch a unicorn. *An exploration of the universe of tech companies with high*.

Sutton, R. S., & Barto, A. G. (1998). Reinforcement learning: an introduction MIT Press. *Cambridge, MA*, *22447*, 10.

Te, Y.-F., Wieland, M., Frey, M., Pyatigorskaya, A., Schiffer, P., & Grabner, H. (2023). Making it into a successful series a funding: An analysis of crunchbase and linkedin data. *The Journal of Finance and Data Science*, *9*, 100099.

Weking, J., Böttcher, T. P., Hermes, S., & Hein, A. (2019). Does business model matter for startup success? A quantitative analysis.

Whig, P., Velu, A., Nadikattu, R. R., & Alkali, Y. J. (2024). Role of AI and IoT in Intelligent Transportation. In *Artificial Intelligence for Future Intelligent Transportation* (pp. 199-220). Apple Academic Press.

Wirth, R., & Hipp, J. (2000). CRISP-DM: Towards a standard process model for data mining. Proceedings of the 4th international conference on the practical applications of knowledge discovery and data mining,

Witten, I. H., Frank, E., & Hall, M. A. (2011). What’s it all about. In *Data mining: Practical machine learning tools and techniques* (pp. 338). Morgan Kaufmann.

Zantalis, F., Koulouras, G., Karabetsos, S., & Kandris, D. (2019). A review of machine learning and IoT in smart transportation. *Future Internet*, *11*(4), 94.

Żbikowski, K., & Antosiuk, P. (2021). A machine learning, bias-free approach for predicting business success using Crunchbase data. *Information Processing & Management*, *58*(4), 102555.

Zhang, C., Bengio, S., Hardt, M., Recht, B., & Vinyals, O. (2021). Understanding deep learning (still) requires rethinking generalization. *Communications of the ACM*, *64*(3), 107-115.

1. Artificial Intelligence [↑](#footnote-ref-1)
2. Machine Learning [↑](#footnote-ref-2)
3. Crunchbase.com [↑](#footnote-ref-3)
4. Decision Tree (DT) [↑](#footnote-ref-4)
5. Artificial Neural Networks (ANN) [↑](#footnote-ref-5)
6. Random Forest (RF) [↑](#footnote-ref-6)
7. Support Vector Machine (SVM) [↑](#footnote-ref-7)
8. Logistic Regression [↑](#footnote-ref-8)
9. Gradient Boosting [↑](#footnote-ref-9)
10. Naïve Bayes [↑](#footnote-ref-10)
11. Argo.AI [↑](#footnote-ref-11)
12. Car sharing [↑](#footnote-ref-12)
13. Delivery Services [↑](#footnote-ref-13)
14. Business Understanding [↑](#footnote-ref-14)
15. KPIs [↑](#footnote-ref-15)
16. Data Understanding [↑](#footnote-ref-16)
17. Data Preparation [↑](#footnote-ref-17)
18. Outliers [↑](#footnote-ref-18)
19. Modeling [↑](#footnote-ref-19)
20. Decision Tree (DT) [↑](#footnote-ref-20)
21. Artificial Neural Networks (ANN) [↑](#footnote-ref-21)
22. Random Forest (RF) [↑](#footnote-ref-22)
23. Support Vector Machine (SVM) [↑](#footnote-ref-23)
24. Logistic Regression [↑](#footnote-ref-24)
25. Gradient Boosting [↑](#footnote-ref-25)
26. Evaluation [↑](#footnote-ref-26)
27. Cross-Validation [↑](#footnote-ref-27)
28. Accuracy [↑](#footnote-ref-28)
29. Precision [↑](#footnote-ref-29)
30. Recall [↑](#footnote-ref-30)
31. Deployment [↑](#footnote-ref-31)
32. Success Prediction Model [↑](#footnote-ref-32)
33. **Venture Capital** [↑](#footnote-ref-33)
34. True Positive Rate [↑](#footnote-ref-34)
35. False Positive Rate [↑](#footnote-ref-35)
36. **Mergers and Acquisitions** [↑](#footnote-ref-36)
37. Initial Public Offering (IPO) [↑](#footnote-ref-37)
38. Artificial Intelligence (AI) [↑](#footnote-ref-38)
39. Machine Learning [↑](#footnote-ref-39)
40. Data Mining [↑](#footnote-ref-40)
41. clustering [↑](#footnote-ref-41)
42. Dimensionality Reduction [↑](#footnote-ref-42)
43. Startup [↑](#footnote-ref-43)
44. Funding Rounds [↑](#footnote-ref-44)
45. Mergers and Acquisitions [↑](#footnote-ref-45)
46. Initial Public Offering - IPO [↑](#footnote-ref-46)
47. Crowdfunding [↑](#footnote-ref-47)
48. True Positive Rate [↑](#footnote-ref-48)
49. Light Gradient Boosting [↑](#footnote-ref-49)
50. Shapley [↑](#footnote-ref-50)
51. Comprehensive Backtesting Algorithm [↑](#footnote-ref-51)
52. Cross-Industry Standard Process for Data Mining [↑](#footnote-ref-52)
53. Synthetic Minority Oversampling Technique [↑](#footnote-ref-53)
54. Overfitting [↑](#footnote-ref-54)
55. Multicollinearity [↑](#footnote-ref-55)
56. Pearson Correlation Coefficient [↑](#footnote-ref-56)
57. Spearman Correlation Coefficient [↑](#footnote-ref-57)
58. Overfitting [↑](#footnote-ref-58)
59. Feature Importance [↑](#footnote-ref-59)
60. Weak Learners [↑](#footnote-ref-60)
61. Loss Function [↑](#footnote-ref-61)
62. Mean Squared Error [↑](#footnote-ref-62)
63. R-squared [↑](#footnote-ref-63)
64. Boosting [↑](#footnote-ref-64)