****

دانشکده فنی و مهندسی

کارشناسی ارشد مهندسی کامپیوتر نرم‌افزار

گروه مهندسی کامپیوتر و فناوری اطلاعات

**گزارش تحقیق درس سمینار**

**موضوع:**

**مروری بر چگونگی آنالیز داده‌های تولید شده توسط اینترنت اشیاء**

**نگارش:**

**علی مهرائی**

**استاد:**

**دکتر احمد فراهی**

دی‌ماه 1399



**چکیده**

در کلان داده‌ها آنچه که داده را ارزشمند می کند پردازشیست که روی داده ای خام صورت می‌گیرد تا به داده‌هایی با ارزش افزوده برسیم. با پیشرفت سریع فضای ابری، پوشش کلان داده به تدریج گسترش یافته است و توجه وسیعی را از تمام جنبه های زندگی به خود جلب کرده است. در روند توسعه اجتماعی مدرن، تجزیه و تحلیل کلان داده به تدریج در برنامه ریزی توسعه آینده، ارزیابی ریسک و ادغام وضعیت توسعه بازار اعمال می‌شود. کلان داده و هوش مصنوعی مهمترین تغییرات فناوری در دنیای کنونی رایانه ها هستند، و این فناوری ها تأثیر بسزایی در تمام مراحل زندگی ما داشته اند. برخی از مطالعات نشان داده اند که در بسیاری از موارد، هر چه تعداد داده‌ها بیشتر باشد، نتایج تحلیل با روشهای هوش مصنوعی دارای کارایی بیشتری می‌شوند. بر این اساس، این مقاله خلاصه ای از دو مقاله است که در در هر کدام از مقاله ها یک مساله را بر اساس یک روش هوش مصنوعی تحلیل خواهیم نمود، خط مشی اصلی این نوشتار یک الگوریتم مشتری آنلاین را بر اساس الگوریتم یادگیری ماشین[[1]](#footnote-1) برای تحلیل کلان داده[[2]](#footnote-2) IOT[[3]](#footnote-3) بدون ساختار[[4]](#footnote-4) پیشنهاد می‌کند و از آن در سناریوهای تحلیل کلان داده استفاده می کند. و در کنار این مساله، تحلیل و پیش بینی ترافیک شهری را بر اساس یک مدل ترافیکی را در یک منطقه شهری مورد بررسی قرار خواهیم داد.

**کلمات کلیدی:** آنالیز داده[[5]](#footnote-5)، اینترنت اشیاء، ترافیک شهری، داده‌کاوی[[6]](#footnote-6)، فروش آنلاین، کلان داده، هوش مصنوعی[[7]](#footnote-7)، یادگیری ماشین.

فهرست مطالب

[عنوان](#_Toc413125201) شماره صفحه

[فصـل اول 1](#_Toc44936604)

[مقـدمـه 1](#_Toc44936605)

[مقدمه 2](#_Toc44936606)

[1-1 تعریف مسئله و بیان سؤال‌های اصلی گزارش تحقیق 2](#_Toc44936607)

[1-2 سابقه انجام تحقیق 4](#_Toc44936608)

[1-3 هدف‌ها 6](#_Toc44936609)

[1-4 چه كاربردهایی از انجام این تحقیق متصور است؟ 6](#_Toc44936610)

[1-5 روش انجام تحقیق 7](#_Toc44936611)

[۱-6 مراحل انجام تحقیق 7](#_Toc44936612)

[1 - 7 ساختار گزارش تحقیق 7](#_Toc44936613)

[فصـل دوم 8](#_Toc44936614)

[مفاهیم عمومی و انواع ساختارهای کلان داده 8](#_Toc44936615)

[مقدمه 9](#_Toc44936616)

[۲ – ۱ اینترنت اشیاء 9](#_Toc44936617)

[۲ – ۲ انواع ساختارهای کلان داده 10](#_Toc44936618)

[2-3 مدل « جوّ رفتاری رانندگان » EDCB 12](#_Toc44936619)

[2-4 تجزیه تحلیل کلان داده‌ها 13](#_Toc44936620)

[2 – 5 جمع بندی 14](#_Toc44936621)

[فصـل سـوم 15](#_Toc44936622)

[مروری بر کارهای انجام‌ شده 15](#_Toc44936623)

[مقدمه 16](#_Toc44936624)

[۳- ۱ استفاده از هوش مصنوعی در تحلیل داده‌های مالی 16](#_Toc44936625)

[۳- 2 هرس اطلاعات با استفاده از هوش مصنوعی در کلان داده‌ها 16](#_Toc44936626)

[۳-3 روشهای ادغام داده‌های به دست آمده از منابع مختلف 16](#_Toc44936627)

[۳- 4 مدیریت اطلاعات به دست آمده از اینترنت 17](#_Toc44936628)

[3- 5 تحلیل اطلاعات به دست آمده از اینترنت 18](#_Toc44936629)

[۳-6 جمع بندی 19](#_Toc44936630)

[فصـل چهارم 20](#_Toc44936631)

[روشهای تحلیل کلان داده 20](#_Toc44936632)

[مقدمه 21](#_Toc44936633)

[4 -1 یادگیری ماشین 21](#_Toc44936634)

[4 -2 پردازش داده‌های بدون ساختار 23](#_Toc44936635)

[4 -3 پردازش داده‌های مرتبط با ترافیک شهری 24](#_Toc44936636)

[4 - 4 جمع بندی 26](#_Toc44936637)

[فصـل پنجم 27](#_Toc44936638)

[جمع‌بندی و پیشنهادها 27](#_Toc44936639)

[مقدمه 28](#_Toc44936640)

[۵ – ۱ نتایج حاصل از تحقیق 28](#_Toc44936641)

[۵ – ۲ پیشنهادها 29](#_Toc44936642)

[مراجع 30](#_Toc44936643)

[مراجع 31](#_Toc44936644)

[واژه نامه 33](#_Toc44936643)

[واژه نامه فارسی به انگلیسی 31](#_Toc44936644)

[واژه نامه انگلیسی به فارسی 36](#_Toc44936644)

**فهرست اشکال**

[عنوان](#_Toc413125201) صفحه

|  |  |
| --- | --- |
| ***شکل 2-1 – دیاگرام OTA******شکل 2-2 – قالب پروفایل سازی هوشمند رفتار******شکل 4-1 – زمان و تعداد گره‌های آزمایش******شکل 4-2 – نام و دیتا و اندازة گره‌های آزمایش******شکل 4-3 - مقایسه نتایج بین بانک اطلاعاتی سنتی و الگوریتم آنالیز ترمینال آنلاین******شکل 4-4 – دیاگرام OTA*** | 131425252526 |
|  |  |

**فهرست جداول**

[عنوان](#_Toc413125201) صفحه

|  |  |
| --- | --- |
| ***جدول 4-1 – زمان و تعداد گره‌های مورد استفاده***  | 26 |

**فهرست علائم اختصاری**

|  |  |
| --- | --- |
| **APNOMS: Asia Pacific Network Operations** **and Management Symposium** | **سمپوزیوم مدیریت و عملیات شبکه آسیا پسیفیک** |
| **EDCB: Environment-driven commuter behavioral** | **جو رفتاری رانندگان** |
| **IoT: Intenet of Things** | **اینترنت اشیا** |
| **OTA: Online Terminal Analysis** | **آنالیز آنلاین ترمینال** |
| **RFID: Radio Friquency Identification** | **شناسة مبتنی بر فرکانس رادیویی** |
| **SVM: Support vector machine** | **ماشین تصمیم برداری** |

## فصـل اول

## مقـدمـه

## مقدمه

اولین بار در سال 1982 میلادی با ارائة یک مدل بهینه شده از دستگاه‌های فروش کوکاکولا[[8]](#footnote-8) برای اتصال به اینترنت، در دانشگاه Carnegie Mellon آمریکا، مفهوم شبکه‌ای از ماشین‌های هوشمند ارائه گردید (Palermo, 2014) و بعد از آن در ابتدای دهه نود میلادی، Mark Weiser در مقاله‌ای تحت عنوان: «کامپیوتر برای قرن بیست و یک» مفهومی از شبکه‌هایی که به غیر از کامپیوتر، دیگر وسایل الکترونیکی و الکتریکی نیز به آن متصل می‌شوند را ارائه نمود. (Weiser, 1991)

در اواسط دهه نود میلادی رضا راجی در مقاله‌ای که با عنوان: «بسته‌هایی کوچک از داده‌ها به مجموعة بزرگی از گره‌ها به منظور ادغام و اتوماسیون همه چیز از لوازم خانگی تا کل کارخانه» که در نشریه «طیف IEEE[[9]](#footnote-9)» به چاپ رساند مفهوم شبکه‌های غیر کامپیوتری را با شرح بیشتری ارائه نمود. (راجی، 1994)

اما در نهایت این Ashton بود که در سال پایانی قرن بیستم عبارت «اینترنت اشیاء[[10]](#footnote-10)» را به کار برد. (Ashton, 1999)

با گسترش اینترنت در اقصی نقاط جهان، و افزایش ضریب نفوذ آن در جوامع و به تبع آن به وجود آمدن شبکه ای از اشیاء[[11]](#footnote-11) که از طریق اینترنت با یکدیگر متصل هستند که در حقیقت همان اینترنت اشیاء می‌باشد، به طور طبیعی حجم بسیار بزرگی از داده‌ها جمع آوری شده و از آنجاییکه با افزایش ظرفیت مراکز داده[[12]](#footnote-12)، قابلیت ذخیره سازی داده‌های با حجم بالا نیز فراهم گردیده. به این حجم عظیم از داده، کلان داده[[13]](#footnote-13) می گویند و کلان داده به دلیل نوآوری مدامی که در همه زمینه های زندگی وجود دارد به تدریج برای انسان دارای ارزش بیشتری می‌شود. مفهوم کلان داده در سال 1980 میلادی توسط Toffler بیان گردید و امروز در همه جنبه های زندگی ما تأثیرهای زیادی دارد و بسیاری از تصمیم گیریهای درست تجاری و سیاسی بر مبنای اطلاعات استخراج شده از روشهای تحلیلی این کلان داده‌ها خواهد بود. (Hou, et. al. 2019)

این کلان داده‌هایی که توسط شبکة اینترنت اشیاء تولید می‌شوند ساختارهای متفاوتی دارند و بر اساس روشهای داده‌کاوی[[14]](#footnote-14) تحلیلهای گوناگونی نیز بر روی آنها صورت می گیرد، همچنین در صنایع مختلف و به شیوه های مختلف این تحلیلها صورت گرفته و مورد استفاده قرار می‌گیرد، اینکه با چه مقصدی بخواهیم از این داده‌ها استفاده کنیم، بسیار با اهمیت است.

## 1-1 تعریف مسئله و بیان سؤال‌های اصلی گزارش تحقیق

به گزارش موسسهGartner [[15]](#footnote-15) در سال 2020 میلادی چیزی نزدیک بیست میلیارد دستگاه (یا به عبارتی دقیقتر: شیء[[16]](#footnote-16)) به اینترنت متصل هستند (وبسایت موسسه Gartner، 2020)، و به همین دلیل، به صورت خودکار حجم زیادی از داده‌ها توسط این اشیاء تولید می گردد، گاهی نتایج حاصله از جمع آوری داده‌ها برای تصمیم گیری های بلند مدت مورد نیاز است، و گاهی برای تصمیم گیری های بلادرنگ، مثلاً شرکت های فناوری اطلاعات مثل علی بابا، Tencent، و.... برای بازده مالی حاصل از کلان داده‌ها سرمایه گذاریهایی انجام داده اند تا بتوانند در سیستمهایی مثل تاکسی های خطی، دوچرخه های اشتراکی و.... از آنها بهره گیری کنند. (Hou, et. al. 2019)

در مقابل نیز داده‌هایی نظیر آنچه از سنسورهای بررسی رفت و آمد خودروها در خیابان ها استحصال می شوند، باید جهت برنامه ریزی بلادرنگ سیستمهای اعلان وضعیت ترافیک مورد استفاده گردند. بدیهی است در مواردی مثل ترافیک خودروها، مواردی نظیر رفتار انسانی و شرایط محیطی نیز علاوه بر اطلاعات دریافتی از سنسورهای بررسی[[17]](#footnote-17) در تولید داده‌های اولیه موثرند. (Bandaragoda, et. al., 2020)

داده‌های به دست آمده از اینترنت اشیا، در زمینه ای مختلفت و براساس فرآیند های مختلف تولید می‌شوند و فرآیندهای داده‌کاوی زیادی روی آنها و برای مقاصد مختلف انجام می‌شود.

به عنوان نمونه، امروزه و بر خلاف دهه های گذشته، کلان داده‌ها جزئی جدایی ناپذیر در سرویسهای مالی هستند و بسیاری از سازمانهای وابستگی زیادی به این داده‌ها دارند چرا که به صورت روزانه صدها میلیون تراکنش مالی در جهان انجام می‌شود و این تراکنشها علاوه بر اینکه تاثیر بسیار زیادی در زمینة محصولات و خدمات مالی از خود به جای میگذارند، در زمینه هایی همچون تحلیل ریسک و بازار های مالی بسیار به کار گرفته می‌شوند. (Morshadul Hasan, et. al., 2020)

مفهوم «اینترنت اشیا» در سالهای بعد از معرفی شدنش، به تدریج به دو نوع تبدیل شد : یکی دلالت بر اینترنت اشیا و دیگری بیانگر دستگاههایی که در دست کاربران برای انتقال اطلاعات، تبادل و ارتباط قرار گرفته مثل گوشیها ی هوشمند. و به طور عمده به استفاده از سنسورها، کدهای دو بعدی و سایر فناوری ها که برای دستیابی به اطلاعات و دسترسی به محصولات و استفاده از اینترنت اشیا و شبکه‌های ارتباطی برای انتقال و ذخیره سازی استفاده می‌شوند، اشاره دارد. واضح است که رابطة اینترنت اشیا با کلان داده همیشه مکمل و تفکیک ناپذیر بوده است. کلان داده شامل داده‌های ساخت یافته، نیمه ساخت یافته و بدون ساختار است و داده‌های بدون ساختار به طور فزاینده ای به بخش اصلی داده تبدیل می شوند. بر اساس گزارش های سازمانهای رسمی، 80% از داده‌های به دست آمده، داده‌های بدون ساختار هستند و رشد نمایی آنها 60% در سال است. (Hou, et. al. 2019)

بحث در زمینه ی کلان داده‌ها، چگونگی استحصال این کلان داده‌ها از اینترنت اشیاء و روشهای تحلیل این کلان داده‌ها و همچنین کاربرد این تحلیلها بسیار گسترده است و تقریبا با هر صنعت و موضوعی گره خورده است و تحقیقات بسیار زیادی در این زمینه نیز صورت گرفته و طبیعی هست که در این گزارش، هدف ما بررسی تمام ابعاد ماجرا نیست، بلکه هدف ما بررسی دو نمونه از تحقیقاتی هست که قبلا در این زمینه انجام شده، با در نظر گرفتن ملاحظات و محدودیت ها و شرایطی که محققین مربوطه در نظر گرفته اند، ابتدا یک مورد که داده‌های آن دارای منابع مشخص و در یک قالب مشخص هست را بررسی خواهیم نمود و درگام بعد یک مورد که داده‌های آن بسیار جیم و فاقد ساختار هست را مورد توجه قرار خواهیم داد:

الف : تحلیل برای پیش بینی وضعیت آینده یک سیستم ترافیکی در یک منطقه شهری، با استفاده از روشهای هوش مصنوعی که داده‌های آن با استفاده از سنسورهای ترافیکی و پاره ای عوامل دیگر به دست می آید.

ب : و در گام بعدی یک کلان دادة بدون ساختار را که از یک سیستم فروش آنلاین استحصال شده را با استفاده از هوش مصنوعی برای برنامه ریزی توسعه آینده، ارزیابی ریسک و ادغام وضعیت توسعه بازار بررسی خواهیم نمود.

دراین گزارش و بر اساس مقاله های مرجع که در بخش 1-6 همین فصل معرفی خواهند شد به سوالات زیر پاسخ خواهیم داد :

1. کلان داده‌های مبتنی بر اینترنت اشیاء از چه روشهایی به دست می آیند، چه ساختاری دارند و چه عواملی بر آنها تأثیرگذار هست؟
2. با چه شیوه ای میتوان داده‌های به دست آمده از اینترنت اشیاء را جهت تصیم گیری آنی[[18]](#footnote-18) تجزیه و تحلیل نمود؟
3. با چه الگوریتمی میتوان کلان داده‌های بدون ساختار را جهت پیش بینی وضعیت آینده، تجزیه و تحلیل نمود؟

در فصل دوم به سوال یک گزارش تحقیق پاسخ خواهیم داد و در فصل سوّم موضوع مورد بحث در سوالات دوّم و سوّم را بسط داده و در نهایت در فصل چهارم پاسخ برای سوالات دوّم و سوّم ارائه خواهیم نمود.

## 1-2 سابقه انجام تحقیق

بسیاری از شرکت ها و دولتمردان برای تصمیم گیری برای آینده علاوه بر تحلیل داده‌های کنونی، برای آینده نگری و آماده سازی بستر لازم برای فعالیت ها و سرمایه گذاریها و توسعه کسب و کارها نیازمند نوعی از تحلیل هستند که بر روی کلان داده‌ها صورت بگیرد و وضعیت داده‌های استحصال شده در آینده را پیش بینی کند.

بررسی جزء به جزء کلان داده‌ها، به ویژه اینکه از نوع داده‌های بدون ساختار باشد، با روشهای معمول کاری ساده نیست و حتی در حالت ساخت یافته و نیمه ساخت یافته نیز، شاید در جائیکه مسأله زمان مطرح باشد، روشهای معمول برای آنالیز داده‌های اولیه ممکن است کارآیی لازم را نداشته باشند و لذا باید راهکاری برای این مهم اندیشیده شود. در این زمینه، یعنی در مواجهه با کلان داده‌ها در زمینه هایی همچون فعالیت های مالی، سیستمهای تشخیص بدافزار، خرید اینترنتی، ترافیک خیابانهای شهری و... نیز فعالیت هایی صورت گرفته است.

در سال 1993 میلادی، Martin-del-Brio و همکارش با استفاده از شبکه‌های عصبی[[19]](#footnote-19) روشهایی برای تجزیه و تحلیل یک سری داده‌های مالی ارائه داد. (Martin-del-Brio, et. al., 1993)

یک از اتفاقاتی که در مواجهه با کلان داده‌ها باید صورت بگیرد، حذف اطلاعات بی فایده می‌باشد که اصطلاحاً به آن هرس داده‌ها گفته می‌شود که این امر در بدست آمدن نتایج شفاف کمک بسیار بزرگی خواهد بود، در این زمینه نیز تحقیات متعددی به ویژه با استفاده هوش مصنوعی صورت گرفته، و به عنوان نمونه میتوان از روشهایی که جهت بهینه سازی و هرس داده‌ها با استفاده شبکه‌های عصبی توسط Stepniewski و همکارش ارائه شد اشاره نمود. (Stepniewski, et. al., 1997)

همچنین در همان سالها، توسط Hernandez و همکارش روشهایی دیگر برای ادغام داده‌هایی که از منابع مختلف دریافت می‌شوند ارائه شد که نحوی که داده‌های زاید در نظر گرفته نشوند و یا به عبارتی دادده های تمیزی به دست آید.( Hernandez, et. al., 1998)

با گسترش استفاده از اینترنت و در همان سالهایی که بحث اینترنت اشیا، هم کم کم پا میگرفت، شیوه هایی برای بررسی الگوهای وبگردی کاربران اینترنت توسط Cooley و همکارانش ارائه شد که اگرچه به صورت مستقیم با داده‌کاوی ارتباطی ندارد اما در تبیین شناخت ما از یک عامل بسیار تاثیر گذار در تولید داده‌های مرتبط با اینترنت، یعنی عامل انسانی بسیار کارگشاست. (Cooley, et. al., 1999)

در سالهای بعدی نیز میتوان مقالات بسیاری را یافت که در زمینة تجزیه و تحلیل اطلاعات با استفاده از شیوه های هوش صنوعی به نگارش در آمده، اما اینترنت اشیاء تا اواخر دهه اول قرن بیست و یکم در گستره جهانی تاثیر گذار نبود و در سال 2013 میلادی با فراگیر شدن اینترنت اشیاء، K.S.Gill در مقاله‌ای با عنوان عجیب : «اینترنت اشیاء! بعد چه؟ » ضمن تحلیل و بررسی عصر اینترنت اشیا، و با تغییر شعری از Eliot [[20]](#footnote-20) شاعر پرآوازة آمریکایی، به حجم زیاد داده که به واسطة اینترنت اشیاء تولید می‌شود اشاره می‌کند و به صورت تلویحی اهمیت چگونگی برخورد ما با این نوع داده‌ها را بیان می‌کند:

«دانش کجاست،

وقتیکه در کلان داده‌ها گم میشویم؛

خِرَد کجاست،

وقتی که در مجاز افتاده ایم؛

انسان کجاست،

 وقتیکه گرفتار اینترنت اشیاء هستیم! ». (Gill, 2013)

در سال 2013 میلادی طی بررسی ای که توسط Ju و همکارانش صورت گرفت، مباحث مدیریت حجم عظیم دیتا که به واسطة اینترنت اشیاء تولید می شوند و در «سمپوزیوم مدیریت و عملیات شبکة آسیا پسیفیک[[21]](#footnote-21)» مطرح شده بودند را طی یک مقاله جمع بندی و ارائه نمودند. (Ju, et. al., 2013)

در کنار گسترش اینترنت و اینترنت اشیاء و با توسعة ساختارهای ابری[[22]](#footnote-22)برای ذخیره و پردازش داده‌ها و رشدِ نمایی تولیدِ دادها در سالهای بعد، کارهای تحقیقی در زمینه نیز رو به رشد می گذارند، که نتایج جستجو در پایگاه داده‌های دانش[[23]](#footnote-23) موید این ادعاست.

تحقیقاتی که مشخصا در زمینه اینترنت اشیاء و داده‌های تولید شده توسط آن صورت گرفته اند، بیشتر از میانة دهة دوّم قرن جاری آغاز گشته اند :

Zhou و همکارانش در سال 2016 میلادی، مواردی نظیر منابع تولید داده، ترافیک تولید شده توسط این داده‌ها، سرویسهای مورد استفاده، محتواهای تولید شده در اینترنت اشیاء و حتی مباحثی نظیر محاسبات ابری[[24]](#footnote-24) را مورد بررسی قرار دادند. (Zhou, et. al., 2016)

داده‌های خام و حجیمی که توسط اینترنت اشیاء تولید و جمع آوری می‌شود، را شاید بتوان بر اساس متدهای چند مستاجری[[25]](#footnote-25) در پردازش ابری مورد تجزیه و تحلیل قرار داد، که با توجه به تحقیقاتی که توسط Dalčeković و همکارانش صورت گرفت نشان داده شد که این روش محاسبة ابری در موارد غیر بحرانی[[26]](#footnote-26) کاربردی خواهد بود. (Dalčeković, et. al., 2016)

یکی از مساله هایی که در مواجهه با کلان داده‌ها در اینترنت اشیاء با آن برخورد می کنیم، تخمین هزینه بهیه سازی داده‌های به دست آمده توسط اینترنت اشیاء به وسیلة محاسبات ابری می‌باشد که این خود موضوع تحقیقی هست که توسط Femminella و همکارانش در سال 2018 میلادی صورت گرفت. ( Femminella, et. al., 2018)

یکی از روشهای مهم در تحلیل دادههای بزرگ که توسط اینترنت اشیاء ایجاد می‌شود استفاده از فناوری پردازش توزیع شده یا خوشهای است. هدوپ[[27]](#footnote-27) و آپاچی اسپارك[[28]](#footnote-28) دو سیستم کاربردی برای تحلیل دادههای بزرگ بخصوص داده‌های تولید شده توسط اینترنت اشیاء میباشند. از این جهت در یک کار تحقیقاتی، اینترنت اشیاء، کلان داده وسیستمهای خوشه ای مورد بررسی قرار گرفتند. نتایج این تحقیقات که توسط روشنک و همکارانش در سال 1397 خورشیدی انجام شده نشان میدهد که برای پردازش کلان داده حاصل از اینترنت اشیاء و تحلیل ترافیک آن؛ استفاده از سیستمهای پردازش توزیع شده مانند هدوپ و آپاچی اسپارك نقش مهم و حیاتی دارد و سیستم اسپارك به علت استفاده از حافظه اصلی نسبت به هدوپ که از حافظه ثانویه استفاده مینماید سرعت بیشتری دارد. (روشنک، دهنوی، کفّاش و گلاب پور، 1397)

در زمینة شیوه های تحلیل برای کلان داده‌ها که اکثراً هم بدون ساختار مشخص هستند، توسط رهبری و همکارش تحقیقاتی با تمرکز بر تجزیه تحلیل ها مرتبط با داده غیرساخت یافته است صورت گرفته. که این تحقیق نیاز برای توسعه شیوه های تحلیلی مناسب و موثر با به کار بردن حجم های عظیم داده‌های ناهمسان به شکل متن غیر ساختاری، صوت و ویدیو را مورد تاکید قرار میدهد. ( رهبری و اوسطی، 1398)

همچنین در سال 2019 میلادی توسط Eddaoudy و همکارش روشهایی برای تصمیم سازی آنی براساس کلان داده‌های حاصل از اینترنت اشیاء ارائه شد که این روشها از هوش مصنوعی با روش یادگیری ماشین[[29]](#footnote-29) استفاده میکنند و در پیش بینی امراض کاربرد دارند. (Eddaoudy, et. al., 2019)

در مواردی نظیر انجام خرید اینترنتی توسط کاربران و اطلاعات استحصال شده از آمارهای کنونی نیز باید پارامترهایی بررسی شوند که نتایج پیش بینی رفتار خریدکنندگان تا حدود زیادی به واقعیت نزیک باشد، تحقیقات صورت گرفته نشان میدهد به ویژه زمانیکه داده‌های جمع آوری شده حجیم بوده و دارای ساختار مشخصی نباشد استفاده از هوش مصنوعی در این زمینه کارگشاست. (Hou, et. al. 2019)

در زمینه اطلاعات ترافیک خودروها در خیابانهای شهری علاوه بر اینکه از سنسورهای ترافیکی به دست می آید، عوامل خارجی مثل رفتار انسانی و شرایط جوی نیز بر نوع اطلاعاتی که داریم و پیش بینی که بخواهیم بر اساس اطلاعات موجود انجام دهیم تأثیرگذار خواهد بود و بدیهی است روشی که در پیش بینی باید مورد استفاده قرار گیرد، باید همه جوانب را در نظر داشته باشد و به ویژه باید به صورت آنی نتایج تحلیل را به ما بازگرداند دراین زمینه نیز استفاده از هوش مصنوعی نتایج مثبتی در پی داشته است. (Bandaragoda, et. al., 2020)

## **1-3 هدف‌ها**

در این بررسی، روشهای استحصال نتایج دلخواه از یک سری کلان داده که از اینترنت اشیاء حاصل می‌شوند را در دو زمینه بررسی خواهیم نمود، ابتدا اینکه از هوش مصنوعی برای پیش بینی ترافیک شهری به صورت آنی و براساس داده‌های به دست آمده از منابع مختلف به ویژه اینترنت اشیاء شامل سنسورهای ترافیکی استفاده خواهیم نمود و دوّمین مورد اینکه با استفاده از هوش مصنوعی به روش یادگیری ماشین برای پیش بینی بازار فروش با استفاده از داده‌های به دست آمده از خریدهای آنلاین مشتریان را که عموما در ردة داده‌های بدون ساختار قرار میگیرند را بررسی خواهیم نمود، سپس به صورت خلاصه راجع به کل فرآیند استحصال داده بحث خواهیم کرد و در نهایت مدلی برای موارد فوق ارائه خواهیم داد.

## 1-4 چه كاربردهایی از انجام این تحقیق متصور است؟

تصمیم گیریهای درست بر اساس داده‌های درست انجام می شوند و داده‌های درست خود اطلاعاتی هستند که در بسیاری از مواقع از داده‌های دیگر استخراج شده اند و بنابراین حداقل در موضوع مورد مطالعه این گزارش می توان جهت بهینه سازی ترافیک شهری، به ویژه در اطراف مراکز خرید، سرگرمی و دیگر محل های پر تردد شهری استفاده نمود.

همچنین از روشهای استحصال اطلاعات از داده‌های فروش های اینترنتی می توان، به کسب. کارهای آنلاین کمک نمود تا برنامه ریزی دراز مدت بهتری صورت دهند تا صاحبات کسب و کار آنلاین، میزان رضایت مشتریان خود را افزایش دهند.

## 1-5 روش انجام تحقیق

روش انجام این تحقیق به‌ صورت کتابخانه­ای همراه با ارائه مستندات می‌باشد. منابع مورد استفاده شامل مقالات، تحقیقات علمی و پژوهشی، کتب، وبسایتهای خبری و آماری و جستجوهای اینترنتی، در زمینه­ی‌ روش‌های جمع آوری داده، عوامل تاثیر گذار بر داده‌ها، هرس داده‌ها و بهینه سازی آنها، اینترنت اشیاء، کلان داده و ارتباط آن با اینترنت اشیاء، روشها و اهداف تجزیه و تحلیل کلان داده‌ها، استفاده از هوش مصنوعی در تجزیه و تحلیل کلان داده‌ها، انواع ساختار کلان داده و داده‌کاوی میباشد که در نهایت استفاده از روش­های فوق در دو زمینه ترافیک شهری و فروش آنلاین مهمترین هدف پژوهش کنونی می‌باشد.

## ۱-6 مراحل انجام تحقیق

در ابتدا ضمن بررسی مفاهیم پایه مانند اینترنت اشیا، داده‌کاوی، کلان داده و ساختارهایش، به شناسایی الگوریتمهای مرتبط با تجزیه تحلیل کلان داده‌ها خواهیم پرداخت و با در نظر داشتن منابع :

(Hou, et. al. 2019)

(Bandaragoda, et. al., 2020)

ساختار های چگونگی آنالیز کلان داده‌ها را در دو زمینه ترافیک شهری و فروش آنلاین بررسی خواهیم نمود و علاوه برآن منابع متعدد دیگری را نیز به صورت ضمنی بررسی خواهیم نمود.

## 1 – 7 ساختار گزارش تحقیق

در فصل جاری ضمن طرح پرسشهای اصلی گزارش تحقیق به بیان اهداف کلی و مباحث مطرح ‌شده می‌پردازیم.

در فصل دوم به مفـاهیـم عمـومـی و روش­های تولید و انواع ساختارهای کلان داده به صورت عمومی و نیز کلان دادة حاصل از اینترنت اشیاء خواهیم پرداخت.

در فصل سوم به بیان کارهای انجام شده در زمینه جمع آوری داده از اینترنت اشیاء، روشهای بهینه سازی کلان داده‌ها و الگوریتمهای مبتنی بر هوش مصنوعی در تحلیل داده‌ها خواهیم پرداخت.

در فصل چهارم ابتدا به بیان کاربردهای تجزیه و تحلیل کلان داده‌ها در زمینه بهسازی و پیش بینی ترافیک شهری و همچنین پیش بینی فروش آنلاین پرداخته شده و ضمن بررسی روشهای تحلیل کلان داده‌ها، در حوزه هایی نظیر امور مالی و تشخیص امراض نیز نتایج تحقیقات انجام شده را بررسی خواهیم نمود.

در فصل پنجم به نتایج گرفته شده از تحقیق پرداخته شده و پیشنهاداتی برای کارهای آینده بیان شده است.

## فصـل دوم

## مفاهیم عمومی و انواع ساختارهای کلان داده

## مقدمه

از زمانیکه در سال 13980 میلادی مفهوم کلان داده توسط Toffler ارائه شد، چشم اندازهای توسعه آن گسترش یافته و به همه جنبه های زندگی، کار و تحقیق رسوخ کرده است، امروزه با ارتقا فناوری اطلاعات، اطلاعات پراکنده جمع آوری و به تدریج در داده‌های بزرگ و پیچیده متمرکز می‌شوند. توسعه سریع کلان داده مزایای بزرگی را برای صنعت فناوری پیشرفته به ارمغان آورده است و عادت ها را تا حدی برای کاربران تغییر داده و توجه بسیاری از شرکتهای دارای قدرت اقتصادی را به خود جلب کرده است. در سال 2017، شرکت های مبتنی بر فناوری اطّلاعات مانند علی بابا[[30]](#footnote-30) برای بهره مندی از بازده مالی حاصل از کلان داده‌ها، در تحقیقات بزرگ سرمایه گذاری کردند. اهمیت استراتژیک فناوری کلان داده فقط مربوط به تسلط بر حجم زیادی از داده‌ها نیست، بلکه تمرکز بر ارزش افزوده این داده‌های مهم است به عبارت دیگر کلان داده با صنعت مقایسه می‌شود، کلید اصلی سود آوری در این صنعت، انجام پردازش داده‌ها و تقویت قابلیت های محاسبه داده برای دستیابی به ارزش داده است. از نظر فنی، ارتباط بین کلان داده و محاسبات ابری به همان اندازه ی پشت و روی یک سکه جدا نشدنی است کلان داده توسط یک رایانه قابل پردازش نیست و باید از یک معماری توزیع شده استفاده کرد این برنامه قابلیت توزیع داده‌کاوی را برای مقادیر زیادی از داده‌ها فراهم می‌کند. با این حال، برای پردازش توزیع شده، ذخیره سازی داده‌های توزیع شده و فناوری های محاسبات ابری و مجازی سازی باید به محاسبات ابری اعتماد کند. با ظهور فضای ابری، « کلان داده » توجه بیشتر و بیشتری را به خود جلب کرده است. تیم تحلیلگر معتقد است که کلان داده اغلب برای توصیف میزان زیادی از داده‌های بدون ساختار و داده‌های نیمه ساختار یافته ایجاد شده توسط شرکت ها استفاده می‌شود. هنگام بارگیری یک پایگاه داده رابطه ای برای تجزیه و تحلیل، این داده‌ها وقت و هزینه زیادی را می طلبند. تجزیه و تحلیل کلان داده با محاسبات ابری همراه است. از آنجا که تجزیه و تحلیل کلان داده به صورت بلادرنگ نیاز به چهارچوبی مانند « نگاشت کاهش[[31]](#footnote-31) » دارد تا بتواند کار را به ده ها، صدها یا حتی هزاران رایانه توزیع کند بنابراین، مطالعه الگوریتم های یادگیری ماشین در زمینه کلان داده نقش مهمی در ارتقا توسعه کشور، تشکیلات اقتصادی و جامعه دارد. امروزه یادگیری ماشین نقش اساسی در پردازش کلان داده‌ها دارد. به عنوان مثال، در بازی « alphaGo » در مقابل بازیکن مشهور بازی گو، « Kejie » در سال 2017، این بازی با نتیجه 3-0 به نفع ماشین « alphaGo » به پایان رسید. این قضیه یک نماد مهم یادگیری ماشین است. یادگیری ماشین به محدودیت های عوامل انسانی غلبه می کند، به طور مؤثر داده‌ها را از طریق شبکه‌های عصبی، درختان تصمیم گیری و علم یادگیری عمیق پردازش می کند و عملیات داده را بهبود می بخشد. (Hou, et. al. 2019)

## ۲ – ۱ اینترنت اشیاء

مفهوم " اینترنت اشیا " از سال 1990 آغاز شد و در اصل از اینترنت سرچشمه گرفت با این حال، در قدمهای بعدی، به تدریج به دو نوع تبدیل شد : یکی پسوند مبتنی بر اینترنت اشیا و دیگری آنچه با گسترش کاربران بین پروژه ها برای انتقال و تبادل اطلاعات و ارتباط نامیده می‌شود مانند شبکه‌های اجتماعی و سیستمهای خرید و فروش آنلاین. به طور کلی به استفاده از سنسورها، کدهای دو بعدی و سایر فناوری ها برای دستیابی به اطلاعات و دسترسی به محصولات و استفاده از اینترنت اشیا و شبکه‌های ارتباطی برای انتقال و ذخیره سازی اشاره دارد با توسعه سریع فناوری شبکه حیوانات[[32]](#footnote-32)، سابقه کلان داده می تواند منابع بیشتری را برای اینترنت اشیا فراهم کند.از سوی دیگر، توسعه فناوری اینترنت اشیا به ورود سریعتر به عصر کلان داده کمک خواهد کرد. دیده می شود که رابطه اینترنت اشیا با کلان داده همیشه مکمل و تفکیک ناپذیر بوده است. (Hou, et. al. 2019)

اینترنت اشیا یک الگوی کاربردی جدید در زمینه فناوری برای دستیابی سریع اطلاعات از راه دور از طریق فناوری های ارتباطی بی سیم مدرن است. اینترنت اشیا به " اینترنت از طریق همة اتصالات " اشاره دارد. یعنی اطلاعات با بارگیری اطلاعات در یک دستگاه سنجش اطلاعات مانند [[33]](#footnote-33)RFID به دست می آید. شبکه ای که هوشمندانه اطلاعات همه جا را شناسایی کند، می تواند از طریق اتصالات اینترنتی به همه جا انتقال یابد. مدل کاربردی فناوری، مبتنی بر دستگاه های جمع آوری اطلاعات فراگیر مانند برچسب های RFID، سنسورها، درایوها و تلفن های همراه، برای رسیدن به هدف نهایی : ارتباط و همکاری است. (Hou, et. al. 2019)

در توسعه علم و فناوری مدرن، استفاده از محاسبات ابری، فناوری شبکه سازی و به اشتراک گذاری منابع اطلاعاتی و سایر فناوری های بالا به منظور ارتقاء مؤثر سطح هوش و سطح سود مدیریت شهری در سطح شهر و بهتر شدن زندگی ساکنان شهری است. این روند نفوذ به دولت شهری، جامعه، اقتصاد و نهادها، شیوة شهرهای هوشمند است. بهترین نمونه استفاده از کلان داده‌ها و فناوری های IOT در زمینه مراقبت های بهداشتی، شناسایی بیماران توسط فناوری RFID است که برای تطبیق، موقعیت یابی بیمار، جمع آوری علائم حیاتی و مدیریت نظارت استفاده می شود. به طور خاص این قضیه بیماران را ملزم می کند که هنگام پذیرش در بیمارستان، ساعت الکترونیکی داشته باشند تا از اطلاعات هویتی بیمار مطلع شوند. در پوشش شبکه شناسایی فرکانس، پزشکان می توانند از فناوری شناسایی فرکانس برای شناسایی، سازماندهی، ردیابی و ثبت هویت بیمار در هر زمان و هر مکان استفاده کنند. اینترنت اشیاء و کلان داده‌ها از ابتدا رابطه ای جدا نشدنی دارند. (Hou, et. al. 2019)

## ۲ – ۲ انواع ساختارهای کلان داده

کلان داده شامل داده‌های ساختاری، نیمه ساختار یافته و بدون ساختار است و داده‌های بدون ساختار به طور فزاینده ای به بخش اصلی داده تبدیل می شوند. بر اساس گزارش های رسمی، 80% از داده‌های موجود در شرکتها، داده‌های بدون ساختار هستند و رشد نمایی آنها 60% در سال است. تجزیه و تحلیل کلان داده از طریق الگوریتم های یادگیری ماشین، دوره جدیدی از تجزیه و تحلیل کلان داده را باز می کند.جمع آوری و استفاده از داده‌های به ظاهر دشوار با استفاده از روشهای هوش مصنوعی ساده می‌گردد. کلان داده به دلیل ایجاد نوآوری مدام در همه زمینه های زندگی به تدریج برای انسان دارای ارزش بیشتری می‌گردد. (Hou, et. al. 2019)

قدرت پردازش و محاسبه معماری سیستم تحلیلی سنتی موجود با تأثیر رشد سریع مقیاس کلان داده‌ها و پیچیدگی روبه رو است، بر اساس گزارش تحقیق، حجم داده‌ها در زمینه های مختلف در حال گسترش است و مقیاس گردآوری داده‌ها اندازه گیری شده است. از GB و TB به EB و ZB[[34]](#footnote-34) افزایش یافته است و انواع زیادی از داده‌ها وجود دارد. علاوه بر طیف گسترده ای از منابع داده، انواع داده‌ها متنوع هستند و ساختار داده‌ها نه تنها داده‌هایی با ساختار سنتی، بلکه داده‌های بدون ساختار را نیز شامل می‌شوند. این امر باعث می شود که راه حل های سنتی ذخیره سازی داده‌ها برای ساختارهای داده فعلی مناسب نباشد و الزامات آنها برای قابلیت های پردازش داده افزایش یابد. داده‌های بدون ساختار معمولاً نمی توانند به طور مستقیم محتوای آن را درک کنند و باید توسط نرم افزار مربوطه باز شوند. این مسأله برای بازیابی اطلاعات در آینده مشکلات زیادی را به همراه می آورد. علاوه بر این، داده‌ها به آسانی قابل درک نیستند. داده‌های بدون ساختار، ساختار مشخصی ندارند، نمی توانند استاندارد سازی شوند و مدیریت آنها آسان نیست، بنابراین جستجو، ذخیره، به روز رسانی و استفاده از این داده‌های بدون ساختار به یک سیستم هوشمندانه نیاز دارد. اسناد اداری، متن، تصاویر و اطلاعات صوتی و تصویری در همه قالب ها داده‌های بدون ساختار هستند. (1) از نظر متن، فن آوری جستجوی کامل متن مبتنی بر تطابق کلمات است و تحقق درخواست دشوار است. جستجوی هوشمند برای بهبود اثر بازیابی از فرهنگ لغت تقسیم کلمات، فرهنگ لغت مترادف و فرهنگ لغت همگن استفاده می کند و تجزیه و تحلیل زمینه بازیابی کاربر و فناوری بازخورد مربوط به کاربر را برای کمک به پرس و جو ترکیب می کند. این کار باعث می شود اطلاعات دانش هوشمندی را در اختیار کاربران قرار دهد و در نهایت اطلاعات معتبری را به طور دقیق به کاربر بازگرداند. شرط لازم برای تحقق کارکردها برای استفاده از یک بخش متن، فرکانس کلمه، تجزیه و تحلیل متن برای تجزیه و تحلیل متن، خوشه بندی متن تجزیه و تحلیل معنایی، متن کاوی و سایر تکنیک های استخراج ویژگی متن لازم است. (2) تصویر، استخراج ویژگی تصویر بر اساس فناوری آنالیز تصویر است. استخراج ویژگی تصویر استفاده از قابلیت های استخراج رایانه است. استخراج ویژگی تصویر شامل سه سطح زیر است : تابع اصلی بصری سازی، ویژگی های اصلی تصویر مانند رنگ، لبه، شکل، بافت، چیدمان و غیره را استخراج می کند. ویژگی های شیء میانی ویژگی هایی هستند که تصاویر را براساس دانش و استدلال منطقی استخراج می کنند [ مانند اشیا خاص یا شخصیت ها ]، در حالتی که از انتزاع پیشرفته برای انجام استخراج ویژگی از ویژگی های انتزاعی یک تصویر بخواهیم استفاده کنیم، مثل جمله رویداد های خاص، محتوای خاص یا ویژگی های تصویر، به پشتیبانی خارجی بیشتری نیاز خواهد بود.(3) تکنیک های تحلیل صوتی، شامل استخراج ویژگی های صوتی، طبقه بندی صوتی و موارد دیگر هستند. در استخراج ویژگی های صوتی، اطلاعات مربوطه به انرژی دامنه فرکانس، نسبت انرژی زیر باند، نرخ عبور صفر، پهنای باند و موارد مشابه در صوتی که درج شده است، میانگین نسبت انرژی زیر باند، ترافیک طیف و سایر محتویات جهت استخراج ویژگی مربوطه به کار میروند. از ویژگی های استخراج شده می توان برای تطبیق و تشخیص صوتی استفاده کرد. (4) ویدئو، ویدئو در حال حاضر پیچیده ترین نوع است، و داده‌های ویدئویی متداول ممکن است حاوی اطلاعات غنی مانند صدا، تصویر و متن باشد. در عین حال، از آنجا که هر پرونده ویدئویی بسیار بزرگ تر از داده‌های دیگر است. مشکل پیچیده و متغیر است. در تحیلیل ویدئو میتوان از روشهای به کار رفته در هر یک از دسته های قبلی بهره گرفت. به عنوان مثال، از تکنیک های تشخیص تصویر می توان برای استخراج فریم های کلیدی از یک فیلم استفاده کرد، و نتایج به دست آمده می تواند به عنوان خلاصه تصویر ویدئو مورد استفاده قرار گیرد. یا می توان یک شاخص تصویر برای این نکات کلیدی برای اجرای یک سرویس فهرست بندی ویدئو ایجاد کرد. در تجزیه و تحلیل فناوری داده‌های بدون ساختار، روش اصلی استخراج ویژگی ها از داده‌های بدون ساختار است و ویژگی های به دست آمده معمولاً داده‌هایی با ابعاد بالا هستند. (Hou, et. al. 2019)

## 2-3 مدل « جوّ رفتاری رانندگان » EDCB[[35]](#footnote-35)

با توجه به اینکه یکی از مواردی که در این مقاله پیرامون آن صحبت خواهیم کرد بررسی یک مدل ترافیک شهری و استخراج دیتا از اطلاعات به دست آمده از منابع مختلف من‌جمله سنسورهای IoT خواهد بود، پس به ارائة مدلی می‌پردازیم که بر اساس آن بتوانیم به نتایج مطلوب‌تری در زمینه روان سازی ترافیک شهری دست یابیم. مدل رفتاری رانندگان شهری به دو بخش عمدة: رفتار روزمره و رفتارهای آنی تقسیم می‌شود که برای راحتی و درک بهتر مفاهیم بعدی یک مدل به عنوان پوشش دهندة دو مدل رفتاری رانندگان ارائه می‌دهیم به نام «جوّ رفتاری رانندگان» یا همان EDCB که طبیعی است بخشی از رفتار رانندگان که به صورت روزمره صورت می‌گیرد، بخش ثابت این مدل و بخش دیگری که راننده در آن به صورت بلادرنگ تصمیم گیری می‌کند، بخش پویای این مدل خواهد بود، برای درک بهتر این مدل باید این موارد را در نظر بگیریم: (1) جوّ ترافیکی[[36]](#footnote-36) : که شامل رفتار رانندگان، شرایط خودرو، وضع خیابان‌ها، چراغ‌های راهنمایی، انواع جاده و چهارراه‌ها و امثالهم می‌باشد (2) عوامل محیطی خارجی[[37]](#footnote-37) : عواملی نظیر تعطیلات عمومی، تعطیلات مدارس، رویدادهای اجتماعی، تعمیرات جاده‌ای، و از این دست عوامل (3) آنالیز رفتاری[[38]](#footnote-38) : تحلیل بلادرنگ شرایط رفتاری ترافیکی قابلیت برداشت صحیح از مدل مورد اشارة EDCB را با استفاده از ماژول‌هایی که در ادامه از آن نام خواهیم برد فراهم خواهد ساخت. (4) پروفایل راننده[[39]](#footnote-39) : علاوه بر موارد ذکر شده، سابقة راننده که در چه شرایطی چه رفتاری از خود نشان داده نیز باید برای تصمیم گیری های آتی نگهداری شود. در شکل 1-2 مدل پیشنهادی EDCB نشان داده شده است. (Bandaragoda, et. al., 2020)

*شکل 2-1 – دیاگرام OTA* (Bandaragoda, et. al., 2020)

در ادامه یک قالب پروفایل سازی هوشمند رفتار[[40]](#footnote-40) پیشنهاد می‌گردد که شامل سه لایة است (1) لایة مدل دیتا[[41]](#footnote-41) : که در این لایه کلیة داده‌های مورد نیاز،از روشهای گوناگون مبتنی از اینترنت اشیا و دیگر عوامل جمع آوری می شوند. (2) لایة هوش مصنوعی[[42]](#footnote-42) : که خود شامل دو ماژول است و فرآیند تحلیل داده‌ها در این لایه صورت خواهد گرفت، این دو ماژول عبارتند از « ماژول پروفایل جریان ترافیک[[43]](#footnote-43) » و « ماژول تحلیل مسیر آمد و شد[[44]](#footnote-44) » (3) لایة پروفایل رفتار راننده[[45]](#footnote-45) : که در این لایه الگوهای رفتاری رانندگان بر اساس مدل EDCB شناسایی خواهند گردید. این سه لایه در شکل 2-2 نشان داده شده اند. (Bandaragoda, et. al., 2020)



*شکل 2-2 – قالب پروفایل سازی هوشمند رفتار* (Bandaragoda, et. al., 2020)

## 2-4 تجزیه تحلیل کلان داده‌ها

مبنای نظری تکنولوژی آنالیز کلان داده‌ها، داشتن مقدار زیادی از داده‌های اولیه است، یعنی داده‌هایی با منابع دقیق، داده‌های غنی و ارتباطاتی طبیعی. تئوری تجزیه و تحلیل کلان داده‌ها عمدتاً شامل دو استراتژی تجزیه و تحلیل است: تجزیه و تحلیل خوشه ای[[46]](#footnote-46) و تجزیه و تحلیل وابستگی[[47]](#footnote-47)، و روش تجزیه و تحلیل پیش بینی مبتنی بر اینهاست. در حال حاضر، فناوری های پردازش کلان داده‌ها عمدتاً شامل فناوری محاسبات توزیع شده، فناوری محاسبات حافظه و فناوری پردازش جریان هستند. زمینه هایی که این سه فناوری در آنها استفاده شده اند متفاوت هستند. فناوریهای محاسبات درون حافظه برای رسیدگی به موضوعاتی مانند خواندن داده‌های کارآمد و پردازش آنلاین بلادرنگ توسعه یافته است. فناوری جریان، جریان های داده واقعی، مداوم و بدون کنترل را حل می کند. از فناوری محاسبات توزیع شده می توان برای تجزیه مشکلات در بسیاری از کارهای کوچک که به چندین فرآیند رایانه ای اختصاص یافته اند، استفاده کرد. منبع آزاد هدوپ به فناوری اصلی رایانه توزیع شده تبدیل شده است، که از مقیاس پذیری، استفاده از تجهیزات مناسب و قابلیت اطمینان بالای خوبی برخوردار است. فناوری محاسبات توزیع شده برای منابع داده توزیع شده در مجموعه شرکت های انرژی کاربرد دارد. فناوری محاسبات درون حافظه داده‌های در مقیاس بزرگ را برای انجام عملیات پرس و جو و تجزیه و تحلیل در حافظه قرار می دهد. فناوری محاسبات حافظه برای خواندن و نوشتن دیسک از زمان زیادی استفاده نمی کند، که سرعت محاسبه را افزایش می دهد. همراه با روشهای تحلیلی مانند آنالیز همبستگی مقطعی، داده‌ها می توانند به سطح بالایی تصفیه شوند و از اطلاعات پایه اصلی گسسته، بی نظم و پیچیده استفاده کامل کنند. پس از جمع آوری، تجزیه و تحلیل و جمع آوری مقدار زیادی از داده‌های اولیه، از طریق تجزیه و تحلیل خوشه ای، یک منبع جمع آوری داده نسبتاً پایدار حاصل می شود. با ارتباط انواع و سطوح مختلف اطلاعات، می توانیم داده‌ها را پس از خوشه بندی نزدیک تر با داده‌های بین دسته های مختلف، ایجاد کنیم. فراهم کردن منبع معتبر از اطلاعات داده‌های مرجع و صرفه جویی در وقت در فرایندهای پیچیده تجزیه و تحلیل داده‌ها، آسان تر است.

## 2 – 5 جمع بندی

هنگامی که مقدار زیادی از داده‌های اصلی از طریق پایگاه داده جمع آوری و به آن ها دسترسی پیدا می کند، فناوری رایانه می تواند، ترکیب و فیلتر مقادیر زیادی از داده‌های بی‌ربط را انجام دهد. با این حال برای بهینه سازی ارتباطات سیار (موبایل)، ارائه پشتیبانی تصمیم گیری اولیه و تعیین شده برای بهینه ساز ضروری است. در شبکه ارتباطات سیار در مناطقی که صدها میلیون کاربر دارد و میزان داده بسیار زیاد است. در از روش درخت تصمیم گیری برای پردازش مقدار زیادی داده‌ها جهت کار بهینه سازی روزانه استفاده می شود. الگوریتم درخت تصمیم گیری با طبقه بندی هدفمند داده‌های زمینه ای اطلاعات درستی را پیدا می کند. بزرگترین مزیت این است که می توان با استفاده از یک زبان ساده، طبقه بندی سریع و مشروح سازی مناسب، برای تحلیل و پردازش داده‌ها در مقیاس بزرگ بهره گرفت.

در این فصل نگاهی داشتیم به اینترنت اشیاء، چگونگی استحصال دیتا از مراجع مختلف جهت بررسی و داده‌کاوی و ارائة روشهایی به صورت کلی برای بررسی و تحلیل داده‌های به دست آمده و نیز اینکه مثلا چگونه می‌شود با ارائه یک مدل برای یک ساختار تولید کنندة دیتا، مثل ترافیک شهری، داده‌ها به شکلی قابل تحلیل جمع آوری نمائیم، در ادامه گزارش تحقیق و در فصل بعدی، نگاهی خواهیم داشت به کارهای انجام شده در این زمینه.

## فصـل سـوم

## مروری بر کارهای انجام‌ شده

## مقدمه

این تحقیقات مبتنی بر وضعیت جدید فعلی توسعه جامعه است و نقش مهمی در ارتقا پیشرقت بهتر جامعه دارد. از طریق تجزیه و تحلیل جامع از پیشینه تحقیق از داده‌های بزرگ و وضع موجود از تحقیقات یادگیری ماشین، مشخص می‌شود که استفاده مؤثر از نتایج تحقیق در زمینه یادگیری ماشینی می تواند مشکل کلان داده را بهتر حل کند. به منظور بهبود تراکم ارزش داده‌های عظیم ساختار یافته و حذف داده‌های زاید این مقاله از داده‌های بدون ساختار به عنوان نمونه استفاده می کند و الگوریتم های مربوط به یادگیری دستگاه برای انجام پیش پردازش، پردازش کاهش ابعاد و آموزش مدل پیش بینی استفاده می کند. و در بانک اطلاعاتی سنتی برای مقایسه کارایی تجزیه و تحلیل داده‌ها نتایج خوبی حاصل شده است.

و همچنین در زمینه مطالعات ترافیک شهری باید گفت : تجزیه و تحلیل ترافیک یک زمینه پر کاربرد است که در بسیاری از تحقیقات از جمله آنالیز اقتصادی، انتخاب و اجرای کنترل ترافیک و ارزیابی کیفیت خدمات و تسهیلات حمل و نقل موضوع تحقیقات بسیاری بوده است. مطالعات مربوط به تحلیل ترافیک و تحلیل جابجایی انسان به بسیاری از اشکال داده از جمله داده‌های شبکه‌های اجتماعی.، داده‌های تحرک انسانی موبایل و داده‌های حسگر ترافیک وابسته است.

## ۳- ۱ استفاده از هوش مصنوعی در تحلیل داده‌های مالی

در سال 1993 میلادی، Martin-del-Brio و همکارش با استفاده از شبکه‌های عصبی روشهایی برای تجزیه و تحلیل یک سری داده‌های مالی ارائه داد. در روشهای این محققین با بررسی وضعیت وامهای یک بانک و اینکه وام گیرندگان تا چه حدی توانایی باز پرداخت دارند توانستند با روشهای هوش مصنوعی رفتار وام گیرندگان در شعب با مختصات مشابه را پیش بینی کنند.( Martin-del-Brio, et. al., 1993)

## ۳- 2 هرس اطلاعات با استفاده از هوش مصنوعی در کلان داده‌ها

یک از اتفاقاتی که در مواجهه با کلان داده‌ها باید صورت بگیرد، حذف اطلاعات بی فایده می‌باشد که اصطلاحاً به آن هرس داده‌ها گفته می‌شود که این امر در بدست آمدن نتایج شفاف کمک بسیار بزرگی خواهد بود، در این زمینه نیز تحقیات متعددی به ویژه با استفاده هوش مصنوعی صورت گرفته، و به عنوان نمونه میتوان از روشهایی که جهت بهینه سازی و هرس داده‌ها با استفاده شبکه‌های عصبی توسط Stepniewski و همکارش ارائه شد اشاره نمود. در این روشها و با استفاده از هوش مصنوعی توانستند فرمی گرافیکی ارائه دهند که بر اساس آن نقاط خارج از ساختار فرم اصلی به عنوان نقاط زاید شناخته میشد و از دایره اطلاعات حذف میگردید.( Stepniewski, et. al., 1997)

## ۳-3 روشهای ادغام داده‌های به دست آمده از منابع مختلف

همچنین در همان سالها، توسط Hernandez و همکارش روشهایی دیگر برای ادغام داده‌هایی که از منابع مختلف دریافت می‌شوند ارائه شد که نحوی که داده‌های زاید در نظر گرفته نشوند و یا به عبارتی داده‌های تمیزی به دست آید به این صورت که با استفاده از هوش مصنوعی و بر اساس چگونگی پیاده سازی ساختار دیتا، از جنبه ها ی مختلفی این داده‌ها بررسی میگردید و سپس بر اساس هر روش یک سری گراف و نمودار تهیه میشد که بر اساس روشهای چند مرحله ی ادغام و حذف در نهایت امر به داده‌های پاکسازی شده و منسجمی میرسیدیم.( Hernandez, et. al., 1998)

## ۳- 4 مدیریت اطلاعات به دست آمده از اینترنت

یکی از مواردی که در ساختار های استحصال کلان داده‌ها بسیار حائز اهمیت است، چگونگی مدیریت حجم داده‌های زیاد است، به ویژه زمانی که با اینترنت و داده‌های تولید شده در آن سر و کار داریم، در این زمینه در سال 2013 میلادی طی بررسی ای که توسط Ju و همکارانش صورت گرفت، مباحث مدیریت حجم عظیم دیتا که به واسطة اینترنت اشیاء تولید می شوند و در «سمپوزیوم مدیریت و عملیات شبکة آسیا پسیفیک » مطرح شده بودند را طی یک مقاله جمع بندی و ارائه نمودند. (Ju, et. al., 2013)

در کنار گسترش اینترنت و اینترنت اشیاء و با توسعة ساختارهای ابری برای ذخیره و پردازش داده‌ها و رشدِ نمایی تولیدِ دادها در سالهای بعد، کارهای تحقیقی در زمینه نیز رو به رشد می گذارند، که نتایج جستجو در پایگاه داده‌های دانش موید این ادعاست.

تحقیقاتی که مشخصا در زمینه اینترنت اشیاء و داده‌های تولید شده توسط آن صورت گرفته اند، بیشتر از میانة دهة دوّم قرن جاری آغاز گشته اند :

Zhou و همکارانش در سال 2016 میلادی، مواردی نظیر منابع تولید داده، ترافیک تولید شده توسط این داده‌ها، سرویسهای مورد استفاده، محتواهای تولید شده در اینترنت اشیاء و حتی مباحثی نظیر محاسبات ابری را مورد بررسی قرار دادند. در روشهای ژو و همکاران، با بهره گیری از آمارهای مرتبط با مکانهای شلوغ در کشور چین، و آمارهایی مرتبط با جابجایی موبایل ها در شبکه موبایل و یا حتی مسیرهای حرکت تاکسیهای درون شهری و نیز آمارهای آب و هوایی و با استفاده از نود های محاسباتی قدرتمند توانستند نگرش های قابل توجهی در مواجهه با اینگونه داده‌ها ارائه کردند. (Zhou, et. al., 2016)

داده‌های خام و حجیمی که توسط اینترنت اشیاء تولید و جمع آوری می‌شود، را شاید بتوان بر اساس متدهای چند مستاجری در پردازش ابری مورد تجزیه و تحلیل قرار داد، که با توجه به تحقیقاتی که توسط Dalčeković و همکارانش صورت گرفت نشان داده شد که این روش محاسبة ابری در موارد غیر بحرانی کاربردی خواهد بود. در این روشها، سعی بر این بود که وقتی با داده‌های حجیم سر و کار داریم، به نوعی جمع آوری و ذخیر سازی دیتا صورت بگیرد که نتایج تحلیل و آنالیز داده‌ها بر اساس یک سری دادة بروز و قابل اعتماد باشد و فرآیند ذخیره سازی دیتا و تحلیل آن به قدری دچار پیچیدگی نگردد که داده‌های جدید در این میان نادیده گرفته شوند.( Dalčeković, et. al., 2016)

یکی از مساله هایی که در مواجهه با کلان داده‌ها در اینترنت اشیاء با آن برخورد می کنیم، تخمین هزینه بهیه سازی داده‌های به دست آمده توسط اینترنت اشیاء به وسیلة محاسبات ابری می‌باشد که این خود موضوع تحقیقی هست که توسط Femminella و همکارانش در سال 2018 میلادی صورت گرفت. در روش Femminella ، ابتدا یک مدل سازی صورت گرفت که طی آن فرآیند جمع آوری داده‌ها از بستری که مبتنی بر اینترنت اشیاء بود به صورت ساده بیان گردید و با استفاده از مدلهای ریاضی، این فرآیند ساده به یک ساختار ابری تعمیم یافت. ( Femminella, et. al., 2018)

## 3- 5 تحلیل اطلاعات به دست آمده از اینترنت

یکی از روشهای مهم در تحلیل دادههای بزرگ که توسط اینترنت اشیاء ایجاد می‌شود استفاده از فناوری پردازش توزیع شده یا خوشهای است. هدوپ و آپاچی اسپارك دو سیستم کاربردی برای تحلیل دادههای بزرگ بخصوص داده‌های تولید شده توسط اینترنت اشیاء میباشند. از این جهت در یک کار تحقیقاتی، اینترنت اشیاء، کلان داده وسیستمهای خوشه ای مورد بررسی قرار گرفتند. نتایج این تحقیقات که توسط روشنک و همکارانش در سال 1397 خورشیدی انجام شده نشان میدهد که برای پردازش کلان داده حاصل از اینترنت اشیاء و تحلیل ترافیک آن؛ استفاده از سیستمهای پردازش توزیع شده مانند هدوپ و آپاچی اسپارك نقش مهم و حیاتی دارد و سیستم اسپارك به علت استفاده از حافظه اصلی نسبت به هدوپ که از حافظه ثانویه استفاده مینماید سرعت بیشتری دارد. در این مقاله سه فناوری مهم و کاربردی امروزه مانند اینترنت اشیاء، کلان داده و معماری پردازش آپاچی اسپارك مورد بررسی و تحلیل قرار گرفته و تلاش شده که این مفاهیم در ارتباط با هم توضیح داده شوند زیرا اینترنت اشیاء حجم بالایی از داده تولید نموده و از طرفی این داده‌ها برای پردازش نیاز به بستر مناسب دارند که یکی از آنها آپاچی اسپارك است. امروزه حجم اطلاعات و دادهها به شکل سرسام آوری در حال افزایش است. یکی از دلیل اصلی رشد دادهها دسترسی افراد به اینترنت، شبکه‌های اجتماعی، انجمنهای گفتگو و غیره است. پردازش و تحلیل این حجم انبوه اطلاعات در بسیاری از کاربردها مانند موتورهای جستجوگر و شبکه‌های اجتماعی فرآیندی چالش برانگیز و مهم است. به علت حجم بالای اطلاعات و دادهها، پردازش این دادهها که به دادههای بزرگ معروف هستند با ابزارهای سابق غیرممکن است. دادههای بزرگ مفهومی جدید است و تلاشهای بسیاری برای گسترش روشهای تحلیلی آنها بکار گرفته می‌شود بخصوص اینکه این دادهها به صورت جریان دادهها نیز در نظر گرفته می‌شوند که دارای روشهای پردازش نظیر هدوپ، نگاشت کاهش، سیستمهای توصیه گر، ترکیب روشهای خوشه بندی و نگاشت کاهش و غیره میباشند. نفوذ به شبکه اینترنت یک چالش مهم و جهانی است که توسط هکر انجام می‌شود و هدف از انجام نفوذ به شبکه ایجاد اختلال درسرویسهای شبکه و سرقت اطلاعات است. نفوذ به شبکه در واقع یک ناهنجاری در ترافیک شبکه است که میتواند کارایی شبکه را دچار اختلال نماید و اطلاعات کاربران در اختیار هکر قرار گرفته شود. شبکه اینترنت اشیاء یک شبکه بزرگ از گرههای هوشمند است که اخیراً ارایه شده است و دارای ترافیک بالایی است و برای تشخیص دادن ناهنجاری در آن نیاز به پردازش کلان داده است. روشهای یادگیری ماشین توانایی بالایی برای مقابله با این حجم بالا از ترافیک ندارند از این رو نیاز است که تجزیه و تحلیل ترافیک این شبکه ها در بستر مناسب مانند آپاچی اسپارك که یک سیستم توزیع شده برای تحلیل کلان داده است انجام شود. با توجه به نقش سیستمهای تشخیص نفوذ در بهبود امنیت شبکههای کامپیوتری در پژوهش آتی از این فناوریهای برای تشخیص نفوذ به شبکه اینترنت اشیاء برای کاربران مشخص گردد. (روشنک و همکاران، 1397)

در زمینة شیوه های تحلیل برای کلان داده‌ها که اکثراً هم بدون ساختار مشخص هستند، توسط رهبری و همکارش تحقیقاتی با تمرکز بر تجزیه تحلیل ها مرتبط با داده غیرساخت یافته است صورت گرفته. که این تحقیق نیاز برای توسعه شیوه های تحلیلی مناسب و موثر با به کار بردن حجم های عظیم داده‌های ناهمسان به شکل متن غیر ساختاری، صوت و ویدیو را مورد تاکید قرار میدهد. داده‌های بزرگ اصطلاحی است که به حجم بسیار زیاد داده‌ها اطلاق می شود، که این داده‌ها از منابع مختلفتی همچون کامپیوترها، دستگاه های تلفن همراه، سنسورها و حتی لوازم خانگی تولید می شوند و حوزه های مختلفی از قبیل صتنعت، تجارت، سلامت و غیره را در بر می گیرد. روند تولید این داده‌ها به صورت پویا و با حجم بسیار بالایی در حال رشد است. در این میان آنچه به عنوان چالش مطرح می شود، چگونگی مدیریت و تحلیل این حجم بالای داده است که در صورت عدم بهره گیری صحیح از آن به موجودیتی بی فایده تبدیل خواهد شد که تنها برای سازمان کسب و کار هزینه ی نگهداری و ذخیره سازی در بر خواهد داشت تمرکز اصلی این مقاله تجزیه و تحلیل انواع داده‌های ساختارنیافته به منظور به دست آوردن بینش ارزشتمند و معتبر از داده‌های بزرگ است که در این راستا تکنیک های تجزیه و تحلیل برای متن، صدا، ویتدئو و داده رسانه های اجتماعی، مورد بررسی قرار گرفته. ( رهبری و اوسطی، 1398)

همچنین در سال 2019 میلادی توسط Eddaoudy و همکارش روشهایی برای تصمیم سازی آنی براساس کلان داده‌های حاصل از اینترنت اشیاء ارائه شد که این روشها از هوش مصنوعی با روش یادگیری ماشین استفاده میکنند و در پیش بینی امراض کاربرد دارند. (Eddaoudy, et. al., 2019)

در مواردی نظیر انجام خرید اینترنتی توسط کاربران و اطلاعات استحصال شده از آمارهای کنونی نیز باید پارامترهایی بررسی شوند که نتایج پیش بینی رفتار خریدکنندگان تا حدود زیادی به واقعیت نزیک باشد، تحقیقات صورت گرفته نشان میدهد به ویژه زمانیکه داده‌های جمع آوری شده حجیم بوده و دارای ساختار مشخصی نباشد استفاده از هوش مصنوعی در این زمینه کارگشاست. (Hou, et. al. 2019)

در زمینه اطلاعات ترافیک خودروها در خیابانهای شهری علاوه بر اینکه از سنسورهای ترافیکی به دست می آید، عوامل خارجی مثل رفتار انسانی و شرایط جوی نیز بر نوع اطلاعاتی که داریم و پیش بینی که بخواهیم بر اساس اطلاعات موجود انجام دهیم تأثیرگذار خواهد بود و بدیهی است روشی که در پیش بینی باید مورد استفاده قرار گیرد، باید همه جوانب را در نظر داشته باشد و به ویژه باید به صورت آنی نتایج تحلیل را به ما بازگرداند دراین زمینه نیز استفاده از هوش مصنوعی نتایج مثبتی در پی داشته است. (Bandaragoda, et. al., 2020)

## ۳-6 جمع بندی

در این فصل به مروری بر کارهای انجام شده پرداختیم و در ضمن آن و *در پاسخ به اولین سوال تحقیق* به روشهای استحصال کلان داده‌ها ، ساختار آنها و همچنین عوامل تاثیر گذار برآنها اشاره ای داشتیم .همان‌طور که گفته شد روش‌های مختلف آنالیز داده‌ها در مواجهه با حجم زیادی از داده، به شیوه های گوناگونی صورت می‌گیرد و بیش از همه، نقش الگوریتمهای مبتنی بر هوش مصنوعی در این زمینه داده­کاوی در زمینه اینترنت اشیاء مورد استفاده قرار گرفته‌اند. صحت و دقت الگوریتم‌های به‌کار رفته متفاوت می­باشد و همواره از بین چند الگوریتم یک الگوریتم که دقت بالاتری دارد که مورد توجه قرار می‌گیرد.

## فصـل چهارم

## روشهای تحلیل کلان داده

## مقدمه

از سال ۱۹۵۰ به بعد که رایانه، در تحلیل و ذخیره­سازی داده­ها به کار رفت؛ حجم اطلاعات ذخیره شده بر روی رایانه­ها پس از دهه­های مختلف رو به افزایش هست و همزمان با پیشرفت فناوری اطلاعات، حجم داده­ها در پایگاه داده­ها همچنان با سرعت بیشتری در حال حرکت رو به جلوست و بنابراین نیاز مند این هستیم که با روشهای منطقی و استدلال انسانی و با بهره گیری از توان ماشینهای محاسباتی و به عبارتی دقیقتر : « هوش مصنوعی » به روشهایی دست پیدا کنیم که نتایج خوبی از تحلیل داده‌ها کسب نماییم.

## 4 -1 یادگیری ماشین

یادگیری ماشین یک حوزه ی تحقیقاتی داغ در علوم رایانه فعلی و رشته های هوش مصنوعی است. صنعت بطور یکنواخت استاندارد " یادگیری ماشین " را تعیین نمی کند، اما یادگیری ماشین بطور کلی الگویی از فرآیندهای شناختی انسان و فرآیندهای یادگیری است که قدرت محاسباتی رایانه ها را برای انجام شبیه سازی رفتارهای انسانی و به دست آوردن دانش جدید یا الگوریتم های مهارت ترکیب می کند. از دانش قبلی و داده‌های آموزشی برای راهنمای یادگیری استفاده می کند و بطور مداوم ساختارهای دانش موجود را برای بهبود عملکردش میپذیرد. در سالهای اخیر، بسیاری از الگوریتمهای یادگیری ماشین بطور گسترده ای در فرآیندهای مهندسی و تحقیقات علمی از جمله خوشه بندی داده‌ها [[48]](#footnote-48)، ماشین پشتیبانی بردار[[49]](#footnote-49)، رگرسیون غیر خطی[[50]](#footnote-50)، شبکه‌های عصبی[[51]](#footnote-51)، الگوریتمهای ژنتیک[[52]](#footnote-52) و غیره مورد استفاده قرار گرفته است.. یادگیری ماشین نقش بزرگی در تحقیقات کلان داده‌ها دارد. به عنوان مثال، موفقیت گوگل در پردازش متن به دلیل یادگیری ماشین است و وقتی یک انباره ذخیره سازی کلان داده[[53]](#footnote-53) ایجاد می شود، دانش زیادی در زمینه شبکه ای عصبی و یادگیری با و بدون نظارت برای استفاده از خوشه های هدوپ[[54]](#footnote-54) لازم است. در عین حال، سیستم توصیه محصولات آمازون نیز ترکیبی از کلان داده‌ها و یادگیری ماشین است. تجزیه و تحلیل عمیق برای تجزیه و تحلیل کلان داده‌ها نیز مبتنی بر آنالیز آماری و یادگیری ماشین است، توسعه یادگیری ماشین عمدتاً شامل دو جهت تحقیق است : اول، مطالعه مکانیسم یادگیری : محور اصلی تحقیق مکانیسم یادگیری، مطالعه تکنیک های یادگیری ماشین است. با توسعه و تغییر در محیط کلان داده‌ها. تجزیه و تحلیل داده‌ها نیازهای کاربردی بالایی در توسعه بسیاری از زمینه های جامعه دارد. از طریق یادگیری ماشین، می توان به سرعت دانش مربوطه را به دست آورد و توسعه فناوری ماشین را ارتقا بخشید. در محیط توسعه کلان داده‌ها، یادگیر ماشین باید نقش مهم یادگیری را برجسته کند، به تدریج دامنه واقعی یادگیری ماشین را گسترش داده و تجزیه و تحلیل داده‌ها را بر اساس یادگیری ماشین انجام دهد، به صورت کارآمد بخش های مختلفی از داده‌ها را پردازش کند و اهداف اساسی یادگیری ماشین را روشن کند. دومین مورد، مطالعه کاربرد منطقی اطلاعات است : تمرکز بر یافتن اطلاعات با ارزش تر از یک مخزن گسترده مدیریت داده‌هاست. در محیط توسعه کلان داده‌ها، به تدریج بازده تولید افزایش یافته و تعداد کلی و انواع داده‌ها تغییرات اساسی را پشت سر گذاشته اند. علاوه بر تجزیه و تحلیل عمیق انواع مختلف ردیف های مهم جدید داده‌ها، مانند تجزیه و تحلیل داده‌های متنی، جستجوی محتوای تصاویر و پردازش داده‌های تصویر، به گونه ای که دستگاه یادگیری ماشین به سمت متنوع سازی توسعه جامعه اقدام کند. در حال حاضر، انتخاب منطقی روش های یادگیری نیمه تأیید شده برای تقویت کیفیت داده‌های آموزش و تقویت توانایی یادگیری، مسأله اصلی است که دغدغه ادارات مربوطه است.

کلان داده‌ها برای هوش مصنوعی اساسی است و تبدیل کلان داده‌ها به دانش یا بهره وری به طور جدایی ناپذیری با یادگیری ماشین در ارتباط است. می توان گفت که یادگیری ماشین هسته اصلی هوش مصنوعی و راه اساسی برای اطمینان از این است که ماشین ها دارای هوش انسانی هستند. وظیفه یادگیری ماشین کشف اطلاعاتی است که بر اساس حجم کلان داده‌های موجود و مفید باشد. هر چه اطلاعات بیشتر پردازش شود، یادگیری ماشین بیشتر می تواند مزایای آن را نشان دهد. این مشکل را می توان با تهیه کلان داده‌ها یا بهبود عملکرد بسیار زیاد، مانند شناخت زبان، طراحی تصویر و پیش بینی وضعیت آب و هوا، حل کرد. روشهای یادگیری K نزدیکترین همسایه[[55]](#footnote-55) مطابق با قوانین خاصی، ابتدا یک نمونه طبقه بندی یا تست خواهد شد و سپس با نمونه های اصلی مقایسه خواهد شد. سپس نزدیک ترین نمونه K را از نتایج مقایسه انتخاب شده و بعد باید محاسبه شود که کدام داده نمونه K در چه تعداد دفعات ظاهر می شود. سپس چه نوع داده ای باید طبقه بندی شود، تعداد c کلاس. $w\_{1}$, $w\_{2}$, … , $w\_{c}$ مشکل تشخیص الگو که هر نوع نمونه از دستة Ni را دارد که :

 Ni(i = 1, 2,... , c). تابع تفکیک کننده که می تواند $w\_{i} $ را مشخص کند این است :

$g\_{i}\left(x\right)=min\left‖x-x\_{i}^{k}\right‖$ k= 1,2,…,$N\_{i}$

برای نمونه های ناشناخته x، به سادگی، فاصله ماهالانوبیس[[56]](#footnote-56) بین نمونه های x و N از دسته شناخته شده را مقایسه کنید : d=$\sqrt{\left(x\_{u}-m\right)^{T}C^{-1}\left(x\_{u}-m\right)}$

که در آن m و c به ترتیب میانگین و کواریانس s است. مشخص گردیده که x برابر با نمونه نزدیک به آن است. این الگوریتم مزایای زیر را دارد : درک ساده و آسان است. نیازی به مدل سازی و آموزش نیست ؛ و اجرای آن بسیار آسان است. برای طبقه بندی رویدادهای نادر مشکلات طبقه بندی بسیار مناسب است. با این حال الگوریتم همچنین دارای کاستی نیز می‌باشد. این الگوریتم یک الگوریتم تنبل[[57]](#footnote-57) و با مصرف حافظه بالا *است*. هنگامی که نمونه آزمون طبقه بندی می شود، مقدار محاسبه بزرگ است و عملکرد کم است. تفسیر ضعیف است و درخت تصمیم گیری و قوانین دیگر قابل ارائه نیست. الگوریتم ماشین بردار پشتیبانی یکی از الگوریتم های یادگیری کلاسیک ماشین است و در هر دو روش تحلیل تئوریک و کاربردهای علمی نتایج خوبی کسب کرده است. یک خط مستقیم برای تقسیم داده‌ها به دو دسته استفاده می شود. این خط به عنوان یک تابع تمایز خطی استفاده می شود و به این صورت بیان می‌گردد : $g\left(x\right)=ω^{T}x+b$

*این رابطه معادل یک هایپرپلان است و معادله بهینه رده بندی هایپرپلان به شرح زیر خواهد بود :*

$ω^{T}x+b=0$

*نمونه با استفاده از نقشه برداری غیر خطی تبدیل به مکان می شود و داده‌های نمونه از فضای نمونه کم بعدی به بعد خطی از بعد ویژگی تبدیل می شوند و سپس اهداف از طبقه بندی خطی به دست می آیند. در نهایت، عملکرد طبقه بندی می تواند به صورت زیر بیان شود :*

$$f\left(x\right)=\sum\_{i=1}^{n}ω\_{i}φ\left(x\_{i}\right)+b$$

*طبق نظریه پرسپترون[[58]](#footnote-58)، عملکرد طبقه بندی برای به دست آوردن نمونه ای که توسط محصول فرم ارائه می‌شود به این صورت خواهد بود :*

$f\left(x\right)=\sum\_{i=1}^{n}α\_{i}y\_{i}<φ\left(x\_{i}\right)$ *,* $φ\left(x\right)>+b$

*علاوه بر الگوریتم SVM، الگوریتم خوشه بندی کلاسیک K-means برای یادگیری ماشین یک الگوریتم خوشه بندی طبقه بندی شده است که می تواند برای تجزیه و تحلیل داده‌ها نیز استفاده شود. این الگوریتم فاصله بین هر شیء و نقطه مرکز تعریف شده را محاسبه می کند و مختصات نقطه مرکز را مطابق با الگوریتم بهینه می کند تا بهترین نتیجه خوشه بندی را به دست آورد :* $J\_{c}=\sum\_{i=1}^{k}\sum\_{p\in C\_{i}}^{}\left‖p-M\_{i}\right‖^{2}$

*یادگیری ماشین به طور عمده شامل مراحل زیر است، (1) انتخاب نوع تجربه آموزش برای ارائه بازخورد مستقیم و غیر مستقیم برای تصمیم گیری سیستم. (2) انتخاب عملکرد هدف برای بهبود عملکرد یادگیری.*

الگوریتم تجزیه و تحلیل کلان داده مبتنی بر IOT مبتنی بر ماشین، ارائه شده در این مقاله به الگوریتم آنالیز ترمینال آنلاین تعلق دارد. طراحی خاص به شرح زیر است : الگوریتم آنالیز ترمینال آنلاین[[59]](#footnote-59) به مدل و فرم داده ورودی وارد می‌شوند تا نتیجه نهایی داده‌ها را به دست آورد که این روش عمدتاً برای طراحی داده‌های بدون ساختار کاربرد دارد. برای سناریوهای مستقیم برنامه محور، نمونه مجموعه آموزش انتخاب شده [[60]](#footnote-60)OTA از داده‌های بدون ساختار است. OTA از فاصله گره مجاور به عنوان یک پارامتر وزنی برای ارزیابی همبستگی استفاده می کند. شکل 4-1 روند خواندن پرونده الگوریتم آنالیز ترمینال آنلاین را نشان می دهد.

## 4 -2 پردازش داده‌های بدون ساختار

در چهار آزمایش تجزیه و تحلیل داده‌ها، تعداد گره‌های حسگر و زمان استفاده شده در آزمایش در شکل 4-2 نشان داده شده است.

همانطور که از جدول 4-1 مشاهده می شود با افزایش تعداد گره‌ها، زمان صرف شده برای پردازش داده‌ها نیز افزایش یافته است به ویژه زمانی که تعداد گره‌ها به 240000 افزایش یافته است. زمان پردازش از پیشرفت بسیار بالایی برخوردار است به گونه ای که نتایج تجزیه و تحلیل داده‌های الگوریتم آنالیز ترمینال آنلاین در طی فرآیند اجرا می تواند بطور جامع ارزیابی شود و آزمایش به تجزیه و تحلیل گره نام و گره داده ادامه می دهد. نتایج در شکل 4-3 نشان داده شده است. در شکل 4-4 مشاهده می شود که اندازه گره نام با اندازه گره داده همبستگی مثبتی دارد و میزان نتایج تجزیه و تحلیل داده‌های بدست آمده توسط اجرای OTA را می توان ارزیابی کرد. از نظر تعداد عملیات پردازش در ثانیه، تعداد گره‌ها در چهار مرحله برابر با تعداد گره‌های جدول 1-4 است. همانطور که از شکل 4-3 مشاهده می شود کارایی تجزیه و تحلیل داده‌ها در پایگاه های داده سنتی کمتر از الگوریتم های ترمینال آنلاین است. به ویژه هنگامی که تعداد گره‌ها در شبکه‌های حسگر زیاد باشد، کارایی الگوریتم های ترمینال آنلاین بسیار بیشتر از پایگاه داده‌های سنتی است و منجر به نتایج خوبی می شود.

*شکل 4-3 - مقایسه نتایج بین بانک اطلاعاتی سنتی و الگوریتم آنالیز ترمینال آنلاین*

*(*Hou, et. al. 2019*)*

*شکل 4-2 – نام و دیتا و اندازة گره های آزمایش*

*(*Hou, et. al. 2019*)*

*شکل 4-1 – زمان و تعداد گره های آزمایش*

*(*Hou, et. al. 2019*)*

## 4 -3 پردازش داده‌های مرتبط با ترافیک شهری

با استفاده از مدلها و روشهای محاسباتی که در مدلهای فوق الذکر و در بخش پیشین به آنها اشاره شد میتوان با بررسی وضعیت گذشته و حال حاضر مسیرهای رفت و آمد خودرو در یک منطقه شهری، الگویی به دست داد که به روان سازی جریان ترافیک کمک شود و در نهایت با استفاده از (1) مجموعه داده‌ها[[61]](#footnote-61). (2) مدلسازی داده‌ها[[62]](#footnote-62) (3) نواحی منتخب[[63]](#footnote-63) (4) آنالیز رفتار پایدار راننده[[64]](#footnote-64) (5)تدابیر با استفاده از تجزیه و تحلیل مسیرهای تردد[[65]](#footnote-65) و (6) تحلیل نوسان رفتار راننده[[66]](#footnote-66) به نتایج مطلوبی در بهینه سازی وضع ترافیک دست یافت.



*شکل 4-4 – دیاگرام OTA* (Hou, et. al. 2019)

به منظور تجزیه و تحلیل عمیق عملکرد الگوریتم آنالیز ترمینال آنلاین که در این مقاله مورد بررسی قرار گرفته است. این مقاله به تجزیه و تحلیل عملکرد داده‌های اصلی مبتنی بر حسگر اینترنت اشیا برای تجزیه و تحلیل کلان داده می پردازد. با توجه به حجم زیادی از اطلاعات داده کاربر، یک بسته بزرگ داده برای آزمایش داده‌ها و سپس پیکربندی این پلت فرم ایجاد شده است. ساختن یک پلت فرم کلان داده، از Hadoop 1.03 , Ubuntu linux 10.04 , SunJava6 استفاده می کند. هدوپ از طریق [[67]](#footnote-67)SSH می تواند گره‌های از راه دور و گره‌های محلی را مدیریت کند. پس از اتمام پیکربندی، داده‌های عملیاتی کاملاً مورد تجزیه و تحلیل قرار می گیرند. در جدول 4-1 زمان و تعداد گره‌های مورد استفاده برای هر تجزیه و تحلیل نشان داده شده است.



*جدول 4-1 (*Hou, et. al. 2019)

## 4 - 4 جمع بندی

در این فصل به مروری بر روش یادگیری ماشین و کاربرد آن در تحلیل داده ای فروش آنلاین پرداختیم و ضمن معرفی داده‌های بدون ساختار، یک سری فرآیندهای مرتبط با داده ای ترافیک شهری را بررسی نمودیم.

با توجه داشته باشیم که با گسترش روز افزون داده ای تولید شده از منابع مختلف، دیگر و مثل سابق، داده‌هایی که برای تصمیم گیری در اختیار ما هست را نمیتوان به شکلی یکدست تعریف نمود، بلکه باید با روشهایی خاص ضمن پاکساری و بهینه سازی داده‌ها از موارد زاید، به نوعی داده‌ها را تحلیل نماییم که علاوه بر اینکه وقت زیادی از ما گرفته نشود، نتایج مثبت و مورد اطمینانی حاصل گردد که در این فصل، مفصلا راجع به یکی از روشهای تجزیه و تحلیل اطلاعات از داده‌ها، یعین یادگیری ماشین صحبت کردیم ضمن اینکه *پاسخ به سوالات دوم و سوم تحقیق* یعنی شیوة تصمیم گیری آنی با استفاده از تجزیه و تحلیل کلان داده‌های حاصل از اینترنت اشیا ( روشهای هوش مصنوعی ) و همچنین استفاده از الگوریتم های یادگیری ماشین، ( با یک مثال در مورد ترافیک شهری ) برای پیشبینی وضعیت آینده را مورد بررسی قرار دادیم.

##

## فصـل پنجم

## جمع‌بندی و پیشنهاد­ها

##

## مقدمه

بانک های اطلاعاتی رابطه ای ده ها سال بر اساس فناوری مدیریت داده‌های ساخت یافته ایجاد شده و اکنون به تکامل رسیده است. با این حال داده‌های بدون ساختار تقریباً کل اطلاعات را تشکیل می دهند. پیچیدگی آن بسیار بالاتر از داده‌های ساخت یافته است بنابراین، چگونگی مدیریت موثر داده‌های بدون ساختار. اولویت اصلی مدیریت داده‌ها است. مدیریت داده‌های بدون ساختار از طریق داده‌های ساختیافته ی حاصل از تبدیل داده‌های ناهمگن یک روش بسیار موثر است. در کل، در فرآیند توسعه کلان داده‌ها، مقدار اطلاعات داده به سرعت در حال افزایش است، و الگوریتم یادگیری سنتی تک دستگاه نمی تواند الزامات اساسی توسعه جامعه را برآورده کند. الگوریتم یادگیری گسترده ماشین به صورت موازی می تواند نیازهای توسعه عصر کلان داده‌ها را برآورده کند و می تواند به طور موثر با نیازهای اساسی توسعه هوش مصنوعی سازگار باشد. ترویج نوسازی اجتماعی محور توسعه آینده یادگیری ماشینی است. در این مقاله کاربرد الگوریتم های یادگیری ماشین در تجزیه و تحلیل داده‌های بدون ساختار از طریق پشتیبانی از فناوری ذخیره کلان داده‌ها، فناوری تجزیه و تحلیل کلان داده‌ها و فناوری پردازش کلان داده‌ها مورد بررسی قرار گرفته است.

## ۵ – ۱ نتایج حاصل از تحقیق

در این گزارش با بررسی سابقه اینترنت اشیا و نیز پیشینه ی تولید و جمع آوری اطلاعات در اینترنت سعی گردید که یک نمای کلی از وضعیت روز، در این زمینه ارائه گردد.

در مجموع، آنچه که بیش از همه مورد اهمیت واقع می‌شود، توان استنتاج از حجم عظیمی از داده‌هاست که بر اساس اصول داده‌کاوی و با استفاده از تکنیکهای هوش مصنوعی انجام می‌شود و بیشترین کاربرد نتایج حاصل از این تحقیقات و محاسبات و آنالیزها، استفاده آنها در امور مدیریتی است به صورتیکه باعث بهیه شدن زندگی روزمره و یز تسهیل فرآیندهای کاری کنونی گردد.

با توجه به تحقیقات انجام شده دراین زمینه، که این تحقیقات در زمینه های مختلفی صورت گرفته، مشخص گردید که میتوان بسیار از فرآیند های مالی و خرید های آنلاین را به شکلی بهینه نمود که کاربران یا به عبارتی خریداران بتوانند از شرایط بهتری استفاده کنند و همچنین صاحبان کسب و کارهای آنلاین با هدف گیری مناسب، در آمد بیشتری کسب نمایند.

ضمن اینکه بر اساس این تحقیقات، متوجه میشویم که بسیاری از پیچیدگیهای زندگی شهری، مثل فرآیندهای مالی، حمل و نقل و غیره، با استفاده از هوش مصنوعی در تحلیل داده‌ها، میتوان به نتایجی دست یافت که ضمن پیش بینی رفتار عوامل تاثیر گذار در آینده، با دقت نسبتا مناسبی تسهیلاتی برای استفاده کنندگان ایجاد نمود، مثلا با پیشبینی وضعیت ترافیکی، و ارائه گزارش به رانندگان درون شهری میتوان از به وجود آمدن حجم بالای ترافیک جلوگیری نمود.

## ۵ – ۲ پیشنهادها

در فرآیند های کاری آینده، ضمن در نظر داشتن آنچه تا کنون انجام شده، میتوان در زمینه هایی نظیر :

* بهبود ساختار های بیمه ای
* بهبود عملکرد های آموزشی در مدارس و دانشگاهها
* بهبود وضعیت خدمت رسانی در مناطق شهری – مثل حمل و نقل شهری – هتل ها و..
* بهبود وضعیت خدمات رسانی سرویسهای ارتباط جمعی
* بهبود برنامه سازی برای رسانه
* بهبود وضعیت سرمایه گذاری، به ویژه خدمات سرمایه گذاری خطر پذیر
* و...

و به خصوص در کشور عزیزمان ایران، از این روشها بهره گیری نمود.

ضمن اینکه ایجاد تخصص و گسترش فن آوری هوش مصنوعی و بهره یری از آن در این ایام کرونایی که بیشتر خدمات به سوی آنلاین شدن گام برمیدارند، بی شک دارای آینده ای روشن و پر بازده خواهد بود.

## مراجع

## مراجع

روشنک، نفیسه - دهنوی، سیما - کفّاش، علی - گلاب پور، امین (1397) **«مروری بر فناوری اینترنت اشیاء با رویکرد پردازش کلان داده در بستر سیستمهای خوشه ای»** دومین کنفرانس ملی کامپیوتر، فناوری اطلاعات و کاربردهای هوش مصنوعی- 1397

رهبری گاوگانی، مینا - اوسطی عراقی، نفیسه (1398) **«کلان داده ابزاری جهت تجریه و تحلیل داده‌های غیرساختیافته در اینترنت اشیاء»** چهارمین کنگره ملی تحقیقات بنیادین در مهندسی کامپیوتر و فناوری اطلاعات - 1398

*Ashton, K. (2009).* ***"That 'Internet of Things' Thing"****. 2009-Retrieved 9 May2017.*

*Cooley, Robert- Mobasher, Bamshad and Srivastava, J aideep (1999).* ***"Data Preparation for Mining World Wide Web - Browsing Patterns"****. Knowledge and Information Systems 1 (1999) 5-32.*

*Dalčeković ,Nikola & Vukmirović ,Srđan & Stoja ,Sebastijan & Milošević ,Nemanja (2016)* ***"Enabling the IoT paradigm through multi-tenancy supported by scalable data acquisition layer"****. (2016) Ann. Telecommun. DOI 10.1007/s12243-016-0523-x*

*Eddaoudy, Abderrahmane and Maalmi, Khalil (2019)* ***"A new Internet of Things architecture for real‑time prediction of various diseases using machine learning on big data environment"****.﻿J Big Data (2019) 6:104* [*https://doi.org/10.1186/s40537-019-0271-7*](https://doi.org/10.1186/s40537-019-0271-7)

*Femminella ,Mauro & Pergolesi ,Matteo & Reali ,Gianluca (2016)* ***"IoT, big data, and cloud*** *computing* ***value chain: pricing issues and solutions"****. Annals of Telecommunications* [*https://doi.org/10.1007/s12243-018-0643-6*](https://doi.org/10.1007/s12243-018-0643-6)

*Frank Palermo (2014)* ***"Internet of Things Done Wrong Stifles Innovation"****. InformationWeek. 7 July 2014. Retrieved 10 November 2014*

*Gill , Karamjit S. (2013)* ***"The Internet of things! then what”*** *.Published online: 27 October 2013 - Springer-Verlag London 2013*

*HERNANDEZ , MAURICIO A. and J. STOLFO, SALVATORE (1998)* ***- “Real-world Data is Dirty: Data Cleansing and The Merge/Purge Problem” -*** *Data Mining and Knowledge Discovery, 2, 9–37 (1998)*

*Ju , Hongtaek et al (2012).* ***"******Management in the Big Data & IoT Era: A Report on APNOMS 2012"*** *J Netw Syst Manage (2013) 21:517–524 - DOI 10.1007/s10922-013-9270-8.*

*Martin-del-Brio, Bonifacio and Serrano-Cinca, Carlos (1993)* ***- “Self-organizing Neural Networks for the Analysis and Representation of Data: Some Financial Cases” -*** *Neural Computing and Applications – Springer,Verlag London Limited (1993)1:193-206*

*Mattern, Friedemann; Floerkemeier, Christian (2010).* ***"From the Internet of Computer to the Internet of Things"*** *(PDF). Informatik-Spektrum. 33 (2): 107–121. Bibcode:2009InfSp..32..496H. doi:10.1007/s00287-010-0417-7. Retrieved 3 February 2014.*

*Raji, R.S. (1994).* ***"Smart networks for control"****. IEEE Spectrum. 31 (6): 49–55. doi:10.1109/6.284793.*

*Rui Hou, YanQiang Kong, Bing Cai, Huan Liu (2020) –* ***“Unstructured big data analysis algorithm and simulation of Internet of Things based on machine learning”*** *- Neural Computing and Applications* [*https://doi.org/10.1007/s00521-019-04682-z - 2020*](https://doi.org/10.1007/s00521-019-04682-z%20-%202020)

*Slawomir W. Stepniewski , Andy J. Keane (1997)* ***- “Pruning Backpropagation Neural Networks Using Modern Stochastic Optimisation Techniques” -*** *Neural Computing and Applications – Springer,Verlag London Limited (1997)5:76-98*

*Tharindu Bandaragoda ,Achini Adikari, Rashmika Nawaratn, Dinithi Nallaperuma, Ashish Kr. Luhach, Thimal Kempitiya, Su Nguyen,Damminda Alahakoon,Daswin De Silva, Naveen Chilamkurti(202) ,****“ Artificial intelligence based commuter behaviour profiling framework using Internet of things for real-time decision-making”*** *- Neural Computing and Applications* [*https://doi.org/10.1007/s00521-020-04736-7*](https://doi.org/10.1007/s00521-020-04736-7) *- 2020*

*Carnegie Mellon University(2014)****, The "Only" Coke Machine on the Internet"****. Retrieved on 10 November 2014.*

*Weiser, Mark (1991).* ***“The Computer for the 21st Century”*** *(PDF). Scientific American. 265 (3): 94–104. Bibcode:1991SciAm.265c..94W.*

*Zhou ,Zhangbing - Tsang , Kim-Fung – v, Zhuofeng - Gaaloul ,Walid (2016).* ***“Data intelligence on the Internet of Things”*** *Published online: 25 April 2016\_ Springer-Verlag London 2016*

واژه‌نامه

**واژه‌نامه فارسی به انگلیسی**

|  |  |
| --- | --- |
| exabyte  | اگزا بایت |
| K-nearest neighbor | الگوریتم k نزدیکترین همسایه |
| Lazy algorithm | الگوریتم تنبل |
| Genetic algorithm | الگوریتمهای ژنتیک |
| Big data storage warehouses | انباره ذخیره سازی کلان داده |
| IoT – Internet of Things | اینترنت اشیاء |
| Apache spark | آپاچی اسپارک |
| AlphaGo | آلفاگو |
| OTA - Online Terminal Analysis | آنالیز آنلاین ترمینال |
| Unstructured | بدون ساختار |
| Knowledge database | پایگاه دانش |
| Perceptron | پرسپترون |
| Traffic flow profiling | پروفایل جریان ترافیک |
| Commuter profile | پروفایل رانندگان |
| Terabyte | تترا بایت |
| cluster analysis | تحلیل خوشه ای |
| Data analysis | تحلیل داده‌ها  |
| Behavior analysis | تحلیل رفتاری |
| Trajectory analysis | تحلیل مسیر آمد و شد |
| Association analysis | تحلیل وابستگی |
| Real time decision-making | تصیم سازی بر خط |
| Environment-driven commuter behavioral | جو رفتاری رانندگان |
| Multi-tenancy | چند مستاجری |
| Data clustering | خوشه بندی داده‌ها |
| Hadoop clusters | خوشه بندی هادوپ |
| Data mining | داده‌کاوی |
| Carnegie Mellon University | دانشگاه کارنگی ملون |
| Nonlinear regression | رگرسیون غیر خطی |
| zettabyte | زتا بایت |
| APNOMS: Asia Pacific Network Operations and Management Symposium | سمپوزیوم مدیریت و عملیات شبکه آسیا پسیفیک |
| IoT sensors | سنسور های اینترنت اشیاء |
| Animal network | شبکه حیوانات |
| Neural networks | شبکه‌های عصبی |
| Network of Things | شبکه ی اشیاء |
| RFID - Radio Frequency Identification | شناسه فرکانس رادیویی |
| Thing | شیء |
| External environment factors | عوامل محیطی خارجی |
| Non critical | غیر بحرانی |
| Cloud | فضای ابری |
| intelligent behavior profiling framework | قالب پروفایل سازی هوشمند رفتار |
| Big data | کلان داده |
| Gigabyte  | گیگیا بایت |
| Commuter behavior profiling layer | لایه پروفایل رفتاری رانندگان |
| Data modelling layer | لایه مدلسازی داده |
| Artificial intelligence layer | لایه هوش مصنوعی |
| Support vector machine - SVM | ماشین تصمیم برداری |
| Coca Cola vending machine | ماشین فروش کوکا کولا |
| Mahalanobis | ماهالا نوبیس |
| Dataset | مجموعه داده |
| Cloud computing | محاسبات ابری |
| Traffic environment | محیط ترافیک |
| Data modeling | مدلسازی داده |
| Data Centers | مراکز داده |
| Gartner Inc. | موسسه گارتنر |
| MapReduce | نگاشت کاهش |
| Hadoop | هادوپ |
| Artificial intelligence  | هوش مصنوعی |
| Machine learning | یادگیری ماشین |

**واژه‌نامه انگلیسی به فارسی**

|  |  |
| --- | --- |
| AlphaGo | آلفاگو |
| Animal network | شبکه حیوانات |
| Apache spark | آپاچی اسپارک |
| APNOMS: Asia Pacific Network Operations and Management Symposium | سمپوزیوم مدیریت و عملیات شبکه آسیا پسیفیک |
| Artificial intelligence  | هوش مصنوعی |
| Artificial intelligence layer | لایه هوش مصنوعی |
| Association analysis | تحلیل وابستگی |
| Behavior analysis | تحلیل رفتاری |
| Big data | کلان داده |
| Big data storage warehouses | انباره ذخیره سازی کلان داده |
| Carnegie Mellon University | دانشگاه کارنگی ملون |
| Cloud | فضای ابری |
| Cloud computing | محاسبات ابری |
| cluster analysis | تحلیل خوشه ای |
| Coca Cola vending machine | ماشین فروش کوکا کولا |
| Commuter behavior profiling layer | لایه پروفایل رفتاری رانندگان |
| Commuter profile | پروفایل رانندگان |
| Data analysis | تحلیل داده‌ها  |
| Data Centers | مراکز داده |
| Data clustering | خوشه بندی داده‌ها |
| Data mining | داده‌کاوی |
| Data modeling | مدلسازی داده |
| Data modelling layer | لایه مدلسازی داده |
| Dataset | مجموعه داده |
| Environment-driven commuter behavioral | جو رفتاری رانندگان |
| exabyte  | اگزا بایت |
| External environment factors | عوامل محیطی خارجی |
| Gartner Inc. | موسسه گارتنر |
| Genetic algorithm | الگوریتمهای ژنتیک |
| Gigabyte  | گیگیا بایت |
| Hadoop | هادوپ |
| Hadoop clusters | خوشه بندی هادوپ |
| intelligent behavior profiling framework | قالب پروفایل سازی هوشمند رفتار |
| IoT – Internet of Things | اینترنت اشیاء |
| IoT sensors | سنسور های اینترنت اشیاء |
| K-nearest neighbor | الگوریتم k نزدیکترین همسایه |
| Knowledge database | پایگاه دانش |
| Lazy algorithm | الگوریتم تنبل |
| Machine learning | یادگیری ماشین |
| Mahalanobis | ماهالا نوبیس |
| MapReduce | نگاشت کاهش |
| Multi-tenancy | چند مستاجری |
| Network of Things | شبکه ی اشیاء |
| Neural networks | شبکه‌های عصبی |
| Non critical | غیر بحرانی |
| Nonlinear regression | رگرسیون غیر خطی |
| OTA - Online Terminal Analysis | آنالیز آنلاین ترمینال |
| Perceptron | پرسپترون |
| Real time decision-making | تصیم سازی بر خط |
| RFID - Radio Frequency Identification | شناسه فرکانس رادیویی |
| Support vector machine - SVM | ماشین تصمیم برداری |
| Terabyte | تترا بایت |
| Thing | شیء |
| Traffic environment | محیط ترافیک |
| Traffic flow profiling | پروفایل جریان ترافیک |
| Trajectory analysis | تحلیل مسیر آمد و شد |
| Unstructured | بدون ساختار |
| zettabyte | زتا بایت |

****

Payam Noor University

Faculty of Engineering

Department of

 Computer Engineering

 and

Information Technology

Course: Seminar

**A review on the Analysis algorithms of the big data**

**that produced by IoT**

**Ali Mehraei**

**949513880**

**Supervisor:**

**Dr. Ahmad Faraahi**

January 2021

1. Machine learning [↑](#footnote-ref-1)
2. Big data [↑](#footnote-ref-2)
3. Intenet of Things – اینترنت اشیاء [↑](#footnote-ref-3)
4. Unstructured [↑](#footnote-ref-4)
5. Data analysis [↑](#footnote-ref-5)
6. Data mining [↑](#footnote-ref-6)
7. Artificial intelligence [↑](#footnote-ref-7)
8. Coca Cola vending machine [↑](#footnote-ref-8)
9. IEEE spectrum [↑](#footnote-ref-9)
10. IoT – Internet of Things [↑](#footnote-ref-10)
11. Network of Things [↑](#footnote-ref-11)
12. Data Centers [↑](#footnote-ref-12)
13. Big data [↑](#footnote-ref-13)
14. Data Mining [↑](#footnote-ref-14)
15. Gartner Inc. [↑](#footnote-ref-15)
16. Thing [↑](#footnote-ref-16)
17. IoT sensors [↑](#footnote-ref-17)
18. Real time decision-making [↑](#footnote-ref-18)
19. Neural Networks [↑](#footnote-ref-19)
20. Thomas Stearns Eliot OM [↑](#footnote-ref-20)
21. APNOMS (Asia Pacific Network Operations and Management Symposium) [↑](#footnote-ref-21)
22. Cloud [↑](#footnote-ref-22)
23. Knowledge database [↑](#footnote-ref-23)
24. Cloud computing [↑](#footnote-ref-24)
25. Multi-tenancy [↑](#footnote-ref-25)
26. Non critical [↑](#footnote-ref-26)
27. Hadoop [↑](#footnote-ref-27)
28. Apache spark [↑](#footnote-ref-28)
29. Machine learning [↑](#footnote-ref-29)
30. Alibaba Co. [↑](#footnote-ref-30)
31. MapReduce [↑](#footnote-ref-31)
32. Animal network [↑](#footnote-ref-32)
33. RFID - Radio Frequency IDentificition [↑](#footnote-ref-33)
34. Gigabyte – Terabyte – Exabyte – Zettabyte [↑](#footnote-ref-34)
35. Environment-driven commuter

behavioural [↑](#footnote-ref-35)
36. Traffic environment [↑](#footnote-ref-36)
37. External environment factors [↑](#footnote-ref-37)
38. Behaviour analysis [↑](#footnote-ref-38)
39. Commuter profile [↑](#footnote-ref-39)
40. intelligent behaviour profiling

framework [↑](#footnote-ref-40)
41. Data modelling layer [↑](#footnote-ref-41)
42. Artificial intelligence layer [↑](#footnote-ref-42)
43. Traffic flow profiling [↑](#footnote-ref-43)
44. Trajectory analysis [↑](#footnote-ref-44)
45. Commuter behaviour profiling layer [↑](#footnote-ref-45)
46. cluster analysis [↑](#footnote-ref-46)
47. Association analysis [↑](#footnote-ref-47)
48. Data clustering [↑](#footnote-ref-48)
49. Support vector machine - SVM [↑](#footnote-ref-49)
50. Nonlinear regression [↑](#footnote-ref-50)
51. Neural networks [↑](#footnote-ref-51)
52. Genetic algorithm [↑](#footnote-ref-52)
53. Big data storage warehouses [↑](#footnote-ref-53)
54. Hadoop clusters [↑](#footnote-ref-54)
55. K-nearest neighbor [↑](#footnote-ref-55)
56. Mahalanobis [↑](#footnote-ref-56)
57. Lazy algorithm [↑](#footnote-ref-57)
58. Perceptron [↑](#footnote-ref-58)
59. OTA - Online Terminal Analysis [↑](#footnote-ref-59)
60. Online Terminal Analysis [↑](#footnote-ref-60)
61. Dataset [↑](#footnote-ref-61)
62. Data modeling [↑](#footnote-ref-62)
63. Selected area [↑](#footnote-ref-63)
64. Stabilized commuter behaviour analysis [↑](#footnote-ref-64)
65. Experiments using traffic trajectory analysis [↑](#footnote-ref-65)
66. Fluctuating commuter behaviour analysis [↑](#footnote-ref-66)
67. Secured Shell [↑](#footnote-ref-67)