

راهنمای پروژه درس داده‌کاوی

نیمسال اول 1401-1402

پژوهش درس داده‌کاوی را می‌توان به صورت فردی یا در گروه‌های دو نفره انجام داد. مراحل انجام پروژه به شرح زیر است:

1. یک مجموعه داده‌ها (Data Set) متعلق به یک شرکت یا حوزه تخصصی (مدیریت کسب و کار، مدیریت ارتباط با مشتری، فضای وب، حوزه سلامت، اقتصاد و نظایر آن) تهیه کنید. در مورد مجموعه داده همه موارد زیر باید برقرار باشد:
الف) تعداد اشیاء یا رکوردهای در دسترس از 300 کمتر نباشد؛ ب) تعداد ویژگی‌ها از 4 کمتر نباشد؛ ج) حاصل ضرب تعداد اشیاء در تعداد خصیصه‌ها از 1500 کمتر نباشد.
تبصره: اگر مجموعه داده مربوط به یک نمونه موردی واقعی در دسترس نباشد، می‌توان یکی از دو کار زیر را اجرا نمود:
الف) می‌توان از مجموعه داده‌های عمومی موجود در فضای وب استفاده کرد؛ با این تفاوت که حداقل‌ها تغییر می‌کند: تعداد اشیاء یا رکوردهای در دسترس از 100,000 کمتر نباشد؛ ب) تعداد ویژگی‌ها از 4 کمتر نباشد؛ ج) حاصل ضرب تعداد اشیاء در تعداد خصیصه‌ها از یک میلیون کمتر نباشد.
ب) می‌توان از سه مجموعه داده عمومی، بدون محدودیت کمک گرفت که روی هر سه مجموعه داده بتوان مسأله مشابهی را حل کرد (بند 2).
2. یک مسأله/هدف را متناسب با مجموعه داده‌هایی که در اختیار دارید، انتخاب کنید (اهدافی مانند افزایش فروش، کاهش هزینه‌های فروش، افزایش بهره‌وری نیروی انسانی، کاهش زمان توقف دستگاه‌ها، کاهش زمان انتظار مشتری، افزایش رضایت مشتری).
3. روی مجموعه داده‌های خود، کلیه آماده‌سازی‌های مورد نیاز را انجام دهید.
4. داده‌های در اختیار را خلاصه‌سازی و بصری‌نمایی و روی نتایج آن تحلیل‌های لازم را ارائه کنید. از نمودارها و شکل‌ها و گراف‌های بصری مناسب و متنوع استفاده شود.
5. از تکنیک‌های داده‌کاوی برای کشف الگوهای جذاب استفاده نمایید. (الف) اگر از روش‌های طبقه‌بندی استفاده می‌کنید، لاقلاً 6 تکنیک طبقه‌بندی مورد استفاده قرار گیرد. (ب) اگر پیش‌بینی متغیرهای پیوسته را انجام می‌دهید، از بین چهار روش شبکه‌های عصبی، رگرسیون، رگرسیون SVM و سری‌های زمانی دست کم دو مورد را استفاده کنید. (ج) اگر از روش‌های خوشه‌بندی استفاده می‌کنید، لاقلاً چهار روش خوشه‌بندی را به کار گیرید. (د) اگر از رویکرد قواعد پیوند استفاده می‌کنید، لاقلاً دو تکنیک را مورد استفاده قرار دهید. (ه) اگر از ترکیب خوشه‌بندی و طبقه‌بندی استفاده می‌کنید، استفاده از یک روش خوشه‌بندی و سه تکنیک طبقه‌بندی کفایت می‌کند. (و) اگر از ترکیب خوشه‌بندی با پیش‌بینی متغیرهای پیوسته استفاده می‌کنید، استفاده از یک روش خوشه‌بندی و یک روش شبکه‌های عصبی، رگرسیون، رگرسیون SVM و سری‌های زمانی کفایت می‌کند. (ز) اگر از روش شبکه‌های بی‌زی استفاده می‌کنید (BBN)، این روش معادل سه روش طبقه‌بندی ارزیابی می‌شود.
6. تکنیک‌های مختلف به کار گرفته شده را ارزیابی نمایید و بهترین روش‌ها را ارائه نمایید. در صورت نیاز، تنظیم پارامترها به دقت انجام شود.
7. نتایج حاصل را تحلیل و بررسی و تفسیر نمایید.
8. راهکارهای مدیریتی حاصل از تحلیل نتایج را استخراج و ارائه نمایید.

توجه: در تهیه گزارش، اصول گزارش‌نویسی و آیین نگارش مورد توجه ویژه قرار گیرد.

تهیه گزارش نهایی

گزارش شما باید به زبان فارسی و به شکل یک مقاله علمی و شامل بخش‌های زیر باشد:

- 1- عنوان
 - 2- چکیده
 - 3- مقدمه: شامل بیان مسأله، کاربردها در دنیای واقعی، ضرورت و اهمیت موضوع، ویژگی‌های کار انجام شده در این تحقیق، ساختار گزارش.
 - 4- پیشینه پژوهش: معرفی پیشینه موضوع با تکیه بر پژوهش‌های مهم انجام شده، تحلیل پیشینه موضوع با تأکید بر شکاف‌های تحقیقاتی و جایگاه این تحقیق در پیشینه موضوع.
 - 5- مدلسازی: بیان شیوه مدل‌سازی و گام‌های طی شده برای حل مسأله
 - 6- نتایج و تحلیل: معرفی سازمان (نمونه موردی)، مجموعه داده در اختیار و ویژگی‌های آن، تنظیم پارامترها، نتایج کلی به دست آمده از اجرای تکنیک‌ها به تفکیک آموزش و آزمون و به ازای شاخص‌های مختلف ماتریس آشفستگی، نرخ صحت، نرخ یادآوری، F-Measure، SSE، MSE و.... سپس، نتایج به دست آمده تحلیل و بهترین تکنیک‌ها معرفی گردد. نتایج حاصل مورد بررسی قرار گیرد و سعی شود کاربردهای موثری برای آنها در سازمان ارائه شود.
 - 7- جمع‌بندی: در این بخش موارد زیر به ترتیب ارائه شود: مروری بسیار مختصر بر فرایند طی شده و نتایج و یافته‌های اصلی؛ بیان نوآوری این پژوهش، نقاط ضعف و محدودیت‌های این پژوهش و ارائه مسیر برای پژوهش‌های آتی
- فهرست مراجع: فراموش نکنید که فهرست مراجع نباید، شماره برای عنوان داشته باشد. نحوه ارجاع‌دهی باید از یک استاندارد مانند سیستم ارجاع‌دهی هاروارد¹ پیروی کند. برای مشاهده نحوه ارجاع‌دهی هاروارد پیوندهای زیر را ببینید:
- <http://libweb.anglia.ac.uk/referencing/harvard.htm>
http://education.exeter.ac.uk/dll/studyskills/harvard_referencing.htm
- گزارش نهایی تا ساعت 24 تاریخ 1401/11/12 باید آماده و ارسال گردد. گزارش نهایی خود را به نشانی hamidreza.koosha@gmail.com ارسال نمایید. مسئولیت دیر رسیدن یا نرسیدن گزارش بر عهده خود دانشجویان گرامی است.

قالب گزارش‌ها

در تهیه گزارش به موارد زیر نیز توجه داشته باشید:

- 1- قلم (فونت) های انتخاب شده باید مناسب باشند. برای متن فارسی، ترجیحاً قلم‌های انتخابی از نوع B مانند B Nazanin یا B Mitra یا B Lotus یا B یا B باشند. تفاوت این قلم‌ها با قلم‌هایی مانند Nazanin یا Mitra در این است که قلم‌های نوع B به هنگام تبدیل فایل به pdf مشکلی ایجاد نمی‌کنند؛ این در حالی است که برخی از نرم‌افزارها در تبدیل قلم‌هایی مانند Nazanin (قلم‌های بدون B) مشکل دارند.

¹ Harvard System of Referencing

- 2- اندازه قلم‌ها مناسب باشند. قلم‌های بیش از حد بزرگ یا کوچک انتخاب نشوند.
- 3- فاصله خطوط را به نحو مناسبی تنظیم کنید. اگر فاصله خطوط، کم در نظر گرفته شود، نوشته‌ها فشرده و غیر خوانا خواهند بود. اگر فاصله خطوط، بیش از حد زیاد در نظر گرفته شود، فضای خالی موجود در صفحه، ظاهر نامناسبی برای آن، ایجاد خواهد کرد.
- 4- عنوان‌ها شماره‌گذاری شوند و زیرعناوین نیز با توجه به عناوینی که در زیر آنها قرار می‌گیرند، شماره‌گذاری شوند.
- 5- تا جایی که امکان دارد، آیین نگارش فارسی را رعایت کنید و واژه‌های فارسی را مورد استفاده قرار دهید.
- 6- دقت داشته باشید که در گزارش پیش از استفاده از علائمی مانند نقطه (.)، ویرگول (،)، و دونقطه (:) و ... هیچ فاصله‌ای بین متن و آنها نباید باشد. همچنین پس از این علائم، باید یک فاصله قرار داده شود و سپس متن نوشته شود. به عنوان نمونه " باشد. همچنین" درست است؛ اما "باشد.همچنین" و " باشد . همچنین" نادرست می‌باشند.
- 7- بین متن داخل پرانتز ()، کروشه []، گیومه « » و آکولاد { } و خود آنها، هیچ فاصله‌ای نیست، اما کلمات خارج آنها (منظور همان پرانتز و غیره است) از دو طرف یک جای خالی، فاصله دارند.
- 8- شکل‌ها و جداول حتماً باید شماره و عنوان داشته باشند. در ضمن هیچ شکل و جدولی نباید در گزارش وجود داشته باشد، مگر آنکه در متن به آن ارجاع داده شده باشد و در مورد آن توضیحی ارائه شود. شماره و عنوان هر شکل در پایین آن شکل و شماره و عنوان هر جدول در بالای آن می‌آید. شماره‌گذاری جدول‌ها و شماره‌گذاری شکل‌ها به صورت جداگانه انجام می‌پذیرد. تمامی منحنی‌ها، نمودارها، تصویرها، ترسیم‌ها و غیره تحت نام «شکل» شماره‌گذاری می‌شوند.
- 9- شماره و عنوان جداول و شکل‌ها باید با فونت B nazanin سیاه وسط چین با اندازه 10 در بالای جدول یا زیر شکل نوشته شود.
- 10- باید بین جدول یا شکل و عنوان جدول یا شکل و متن در قبل و بعد از آن یک سطر فاصله وجود داشته باشد.
- 11- تا جایی که امکان دارد، در متن واژه‌ها با الفبای لاتین ذکر نشوند (به جز علائم اختصاری). در صورتی که در متن واژه‌ای را استفاده می‌کنید که مایل هستید، معادل لاتین آن نیز ذکر شود، واژه لاتین را به صورت زیرنویس² بیاورید.
- 12- در مورد واژه‌های دو بخشی که باید کنار هم قرار گیرند دقت لازم، مبذول گردد. به عنوان مثال ننویسید "می توان" یا "می شود"؛ درست آن این است: "می توان" یا "می شود".
- 13- سعی کنید که متن گزارش هم‌تراز³ باشد.
- 14- ارجاع‌دهی در متن را فراموش نکنید. هر جمله یا عبارتی که از جای دیگر می‌آوردید چه به صورت الکترونیک و چه به صورت کاغذی باید ارجاع‌دهی انجام دهید. گزارش‌های بدون ارجاع‌دهی فاقد ارزش تلقی می‌شوند و بررسی نخواهند شد.

پیروز باشید

حمیدرضا کوشا

² Footnote

³ justify