



امتحان عملی اول درس یادگیری ماشین

استاد : دکتر رضا رضانی
طراح سوالات : علی معینیان

سلام به دوستان گرامی 😊 به اولین امتحان عملی خودتون خوش اومدید. تنها خواهشی که ازتون دارم اینه که سوالات رو با دقت مطالعه کنید. مهم ترین کاری که باید بکنید همینیه 😊

قوانین امتحان

✓ زمان امتحان: 2 ساعت، سوالات به گونه‌ای طراحی شده که در زمان مقرر قابل حل باشند.

✓ دسترسی‌ها: در زمان امتحان، دسترسی به اینترنت و تمامی ابزارهای هوش مصنوعی **مجاز است** ولی به هیچ عنوان حق استفاده از ابزارهای ارتباطی مانند تلگرام، واتساپ و ایمیل را **ندارید**.
اگر این برنامه‌ها باز هستند، از تمام این برنامه‌ها **خارج شوید** به نحوی که در Taskbar هم نباشند و در مرورگر هم آنها را باز نکنید.

✓ نحوه نظارت: قبل از شروع امتحان، برنامه SystemApps.exe را دانلود و اجرا کنید.
در ابتدا، شماره دانشجویی، نام و نام خانوادگی را در این برنامه وارد کنید. این برنامه تمامی کارهای شما را لاگ می‌کند. در صورت استفاده از هر نرم‌افزار ارتباطی، سیستم لاگ را ثبت کرده و نتیجه امتحان شما صفر خواهد بود.

✓ فایل‌های نهایی که باید تحویل دهید عبارتند از:
• لاگ‌های سیستم: فایل‌های allRecords.txt و activeRecords.txt در صفحه دسکتاپ شما ذخیره شده.
• نوت‌بوک پاسخ برای هر 3 سوال به صورت کامل به همراه تحلیل‌ها و پاسخ سوالات.
• تحویل هر کدام از این موارد بعد از زمان امتحان هیچ نمره‌ای نخواهد داشت.

✓ همانطور که قبلاً هم بیان شده بود، تحلیل شما از خروجی‌هاست که اهمیت دارد. پس نشان دادن چند نمودار رنگی یا چند چارت خوشگل در خروجی نمره خاصی ندارد. حتماً و حتماً تحلیل داشته باشید.

✓ در نهایت دوباره تکرار می‌کنم که ابزارهای هوش مصنوعی در کنار شما هستند اما خودتان سوالات را با دقت بخوانید تا به خوبی درک کنید که سوال چیست؟

در نهایت اگر سوالی داشتید از من بپرسید، به شما در حل سوالات کمک خواهم کرد.
موفق باشید ❤️



شما به عنوان Lead Data Analyst در شرکت Urban Hype استخدام شده‌اید. این شرکت، بزرگترین پلتفرم فروش آنلاین کتونی و لباس است. مدیر مارکتینگ، خانم جسیکا، وارد قسمت IT شرکت شده و به شما خبر می‌دهد که **کمپین بزرگ نزدیکه!**

جسیکا هیچ ایده‌ای ندارد که باید چه محصولی رو پروموت کنند؟ نمیدونن روی چه برندهایی تمرکز کنند؟ یا اصلا ترند بازار به کدام سمت داره میره؟ جسیکا فقط به دیتاست برای شما آورده. به جسیکا کمک کنید تا به دیدگاه درستی در خصوص داده‌هایی که داره برسه.

ماموریت 1: با دیتاست آشنا بشید. ستون‌ها رو بررسی کنید. فرمت ستون‌ها و دیتای داخلشون چیه؟ آیا فرمت دیتاها استاندارد؟

ماموریت 2: شروع به تمیز کردن داده‌ها کنید. به هر روشی که بلد هستید بررسی کنید وضعیت داده‌ها از نظر تمیز بودن و Missing Value چگونه؟ اگر وجود داره راه چاره چیه؟ برامون از راه حلتون بنویسید.

ماموریت 3: اگر بخواهید Feature‌های جدید اضافه کنید، چه Feature‌هایی خواهد بود؟ چرا؟ اصلا نیاز هست این Feature رو اضافه کنید یا نه؟ آیا برای تحلیل دقیق‌تر در بعد زمان، نیاز دارید ستون Date رو کمی تغییر روش ایجاد کنید؟

ماموریت 4:

- روند فروش را در ماه‌های موجود بررسی کنید و ماه‌های پرفروش را مشخص کنید (از نمودار مناسب استفاده کنید).
- با توجه به اینکه در حال حاضر کمپین مهمی داریم، گزارشی از 5 کالای پرفروش هر ماه بدهید.
- بررسی کنید آیا روز خاصی وجود دارد که مردم بیشتر خرید کنند؟ مثلا آخر هفته‌ها؟ یا هفته اول هر ماه؟
- یک نمودار دایره‌ای یا همون Pie Chart بکشید که سهم هر برند رو از کل درآمد مشخص کنه.
- محصولات رو به سه دسته تقسیم کنید و رفتار خرید بر اساس جنسیت و بر اساس کشور رو در اون تحلیل کنید:
 - محصولات Economy یعنی زیر 150 دلار
 - محصولات Premium یعنی بین 150 تا 300 دلار
 - محصولات Exclusive یعنی بالای 300 دلار
- با توجه به بررسی‌هایی که کرده‌اید، در یک گزارش یک خطی باید به جسیکا بگویید: بهترین تاریخ کمپین است که کاربرهامون دوست دارن به شیوه پرداخت داشته باشند. تمرکز رو بزارید روی محصولات (Economy/Pre/Exclu) در کشورهای حتما برندهای (3 برند برتر فروش) رو در انبار داشته باشید و از هرکدوم (3 کالای پرفروش) رو شارژ کنید.

سوال 2: دیتاست Students_Exam_Score.csv

فرض کنید شما مشاور ارشد سازمان ملی سنجش و آموزش کشور هستید و قراره که روی یک پروژه جدید کار کنید ! رئیس سازمان ملی سنجش کشور قراره گزارش جامعی از وضعیت عملکرد دانشجویان بده و شما به عنوان مسئول این پروژه باید گزارش جامعی از وضعیت عملکردی دانشجویان بدهید.

- در ابتدا Feature جدیدی به نام Final Score بسازید که میانگین ReadingScore، MathScore و WritingScore باشد.
 - حتما مقادیر را از نظر مقدار Null بودن و منطقی بودن اعداد داخل دیتاست بررسی کنید. مثلا بررسی کنید آیا اعداد همگی بین 0 تا 100 هستند؟ ممکن است اشتباهاتی صورت گرفته باشد.
 - دانشجویان را با توجه به Final Score در سه دسته قرار دهید:
 - ضعیف: نمره زیر 45
 - متوسط: نمره 45 تا 65
 - خوب: 65 تا 100
 - بررسی کنید چند درصد از دانشجویان در کدام دسته بندی قرار میگیرند؟ از نظر جنسیت چطور؟
 - بررسی کنید آیا رابطه‌ای میان تحصیلات والدین و دسته‌ای که دانشجویان در آن طبقه‌بندی شده وجود دارد؟
 - آیا می‌توانید ارتباط بین Final Score و TestPrep را تشریح کنید؟ آیا اصلا رابطه‌ای میان این دو Feature هست؟
 - آیا نوع غذای مصرفی دانشجو بر عملکرد نهایی او موثر بوده است؟
- با توجه به این مورد آیا افزایش کیفیت غذای دانشجویان می‌تواند باعث افزایش عملکرد آنان شود؟

سوال 3: دیتاست dirty_cafe_sales.csv

- چند ماه از تاسیس کافه Smile (بالای میدان فارابی - مرداوچ 😊) گذشته و نیاز داریم که یه بررسی از دیتاهای موجود داشته باشیم تا فروش رو بررسی کنیم. ناگهان متوجه می‌شویم که سیستم صندوق‌داری قاطی کرده و وضعیت داده‌ها داغوووونه. حسابدارها گیج شدن، داده‌ها پر از مقادیر Null و خطای Error هست. خلاصه بگم-براتون، همگی بدبخت شدیم.
- شما تنها امید این کافه هستید. هر کاری بلد هستید بکنید تا بتوانید دیتاست موجود رو به وضع بهتری برسانید. حق ندارید داده‌ای رو به صورت رندوم پر کنید، مگر اینکه آخرین گزینه باشه.
- میتوانیم با مهندسی معکوس قیمت سطرها بفهمیم Item چی بوده؟ شاید بشه فهمید
- در کل باید بدانید که ما در این دیتاست هم Missing Value داریم، هم Invalid Value که با UNKNOWN و ERROR می‌توانید پیداشون کنید. مقادیر Price رو هم باید از نظر وجودشون و منطقی بودنشون چک کنید.
- در نهایت، باید به مدیریت کافه Smile گزارش بدید که چطوری این حجم از داده‌های کثیف رو توانستید یک پله ارتقا بدید و داده‌های تمیز تری رو داشته باشیم.