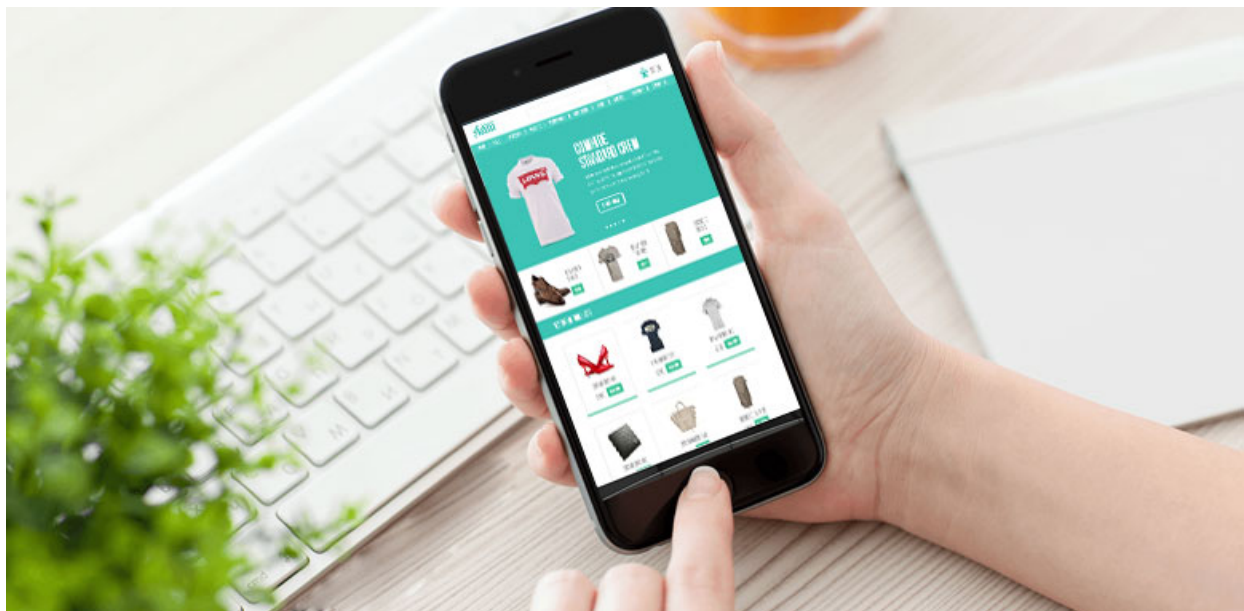


تشخیص ترجیح رابط کاربری به کمک الگوریتم‌های یادگیری ماشین



مقدمه:

رابط کاربری (UI) طرح گرافیکی یک برنامه است که شامل دکمه‌هایی است که کاربران روی آن کلیک می‌کنند، متنی که می‌خوانند، تصاویر، لغزنده‌ها، فیلدهای ورودی متن و بقیه مواردی که کاربر با آنها تعامل دارد. رابط کاربری شامل طرح‌بندی صفحه، انتقال‌ها، انیمیشن‌های رابط و هر ریز تعامل است. هر نوع عنصر بصری، تعامل یا انیمیشن همگی باید طراحی شوند. یک شرکت تجارت الکترونیک اخیراً رابط کاربری وبسایت خود را تغییر داده است، اما طراحان این وبسایت نمی‌دانند که چگونه می‌توان رابط کاربری جدید خود را ارزیابی کرد. آیا مشتریان آن را دوست دارند؟ آیا آنها رابط کاربری جدید را ترجیح می‌دهند یا قدیمی؟

آنها همچنین بازخوردهای مشتریان خود را بر اساس برخی نظرسنجی‌های از پیش سازماندهی شده جمع‌آوری کرده‌اند. از شما به عنوان مهندسان یادگیری ماشین خواسته شده است که یک مدل یادگیری ماشین بسازید تا اولویت رابط کاربری کاربر را بر اساس اطلاعات تعامل UI آنها شناسایی کنید.

مجموعه دادگان:

دیتاستی که در این تمرین قرار است از آن استفاده شود شامل اطلاعات حاصل از نظرسنجی کاربران یک وبسایت خرید اینترنتی است. ستون‌هایی که در این مجموعه داده قرار دارد به شرح زیر است:

شماره ID کاربر	CustomerID
سن کاربر	Age
شهر محل زندگی کاربر	City
ناحیه محل زندگی کاربر	State
تعداد کل سفارش‌هایی که کاربر در سایت ایجاد کرده است.	No_of_orders_placed
تاریخ آخرین سفارشی که کاربر ثبت کرده است.	Last order placed_date
ایا کاربر یک کاربر ویژه است؟ (۰ یا ۱)	Is premium_member
امتیاز کاربر به قسمت لباس‌های زنانه (بین ۰ تا ۱۰)	Women's Clothing
امتیاز کاربر به قسمت لباس‌های مردانه (بین ۰ تا ۱۰)	Men's Clothing
امتیاز کاربر به قسمت لباس‌های بچگانه (بین ۰ تا ۱۰)	Kid's Clothing
امتیاز کاربر به قسمت وسایل خانه (بین ۰ تا ۱۰)	Home & Living
امتیاز کاربر به قسمت لوازم آرایشی (بین ۰ تا ۱۰)	Beauty
امتیاز کاربر به قسمت وسایل الکترونیکی (بین ۰ تا ۱۰)	Electronics
اینکه آیا کاربر UI قبلی را ترجیح داده است یا جدید؟ (New_UI یا OLD_UI)	Preferred_Theme

پیش‌پردازش داده:

در پوشه اصلی تمرین، یک فایل نوتبوک به نام User-UI-Preference.ipynb وجود دارد که تمامی مراحل لازم انجام این تمرین در آن توضیح داده شده است. دو فایل train.csv و test.csv در اختیار شما نیز قرار داده شده است که به ترتیب مربوط به داده‌ی آموزش و تست است.

پیش پردازش داده چیست؟ پیش پردازش داده ها مرحله ای جدایی ناپذیر در یادگیری ماشینی است، زیرا کیفیت داده ها و اطلاعات مفیدی که می توان از آن استخراج کرد به طور مستقیم بر توانایی مدل ما برای یادگیری تأثیر می گذارد. بنابراین، بسیار مهم است که داده های خود را قبل از وارد کردن آنها به مدل خود، از قبل پردازش کنیم. روش های فراوانی برای پیش پردازش داده در ادبیات یادگیری ماشین و علم داده وجود دارد که در این تمرین به برخی از آنها خواهیم پرداخت که عبارتند از:

الف) پرکردن داده های miss شده

ب) نرمال سازی و استاندارد سازی داده

ج) آماده سازی داده های غیر عددی (categorical) برای مدل های پیش بینی کننده

توضیحات لازم مربوط به آنها در نوتبوک موجود است.

طراحی و آموزش مدل های یادگیری ماشین:

پس از آماده سازی مجموعه داده خام به داده قابل استفاده، نوبت به پیاده سازی الگوریتم های یادگیری ماشین می رسد. در این تمرین می بایست یکی از الگوریتم های KNN و Naïve Bayes به دلخواه خودتان انتخاب کنید و بدون استفاده از کتابخانه های موجود مانند Scikit-learn آنها را از ابتدا پیاده سازی و ارزیابی کنید.

سپس برای رسیدن به دقت های بهتر باید از حداقل دو الگوریتم نظارتی از پیش پیاده سازی شده در کتابخانه Scikit-learn استفاده کنید. توضیحات لازم در فایل نوتبوک آمده است.

ارزیابی مدل:

پس از پیاده سازی مدل های متفاوت می بایستی مدل های خود را به کمک داده ارزیابی (validation) ارزیابی کنید. اینکه داده ارزیابی چگونه بدست آمده است در نوتبوک توضیح داده شده است. برای ارزیابی از متریک های Precision, Accuracy, Recall و F1-score استفاده کنید.

موارد تحویل:

لطفا به موارد زیر حتما دقت کنید:

1. یک فایل گزارش به فرمت PDF نیز به همراه نوتبوک تکمیل شده می بایستی آپلود کنید که شامل فرضیات، نتایج ارزیابی و دیگری توضیحات لازمی که از شما در نوتبوک خواسته شده باشد.
2. سلول های نوتبوک شما باید از پیش اجرا شده باشد و نتایج اجرای آن در نوتبوک باقی مانده باشد. اگر نباشد شما نمره آن قسمت از دست خواهید داد.

3. اگر هر قسمت از کدهای شما قابل اجرا نباشد شما نمره آن قسمت را از دست می‌دهید لذا به کد نوشته خود دقت لازم را کنید.

لطفاً به نکات زیر توجه فرمایید:

- فرمت نام‌گذاری فایل‌ی که آپلود می‌کنید حتماً به صورت **[student id][student name].zip** باشد.
- در صورت مشاهده هرگونه تقلب نمره **صفر** برای تکلیف در نظر گرفته می‌شود.
- سوالات و ابهامات خود را از **mohammad99hashemi@gmail.com** بپرسید.