

پروژه پایان دوره (شماره ۲) اردلان دانشور

توضیحات کلی:

طبق صورت سوال یک دیتاست به ما داده شده است که در آن اطلاعات مشتریان قرار دارد.

هدف اصلی این است که یک مدل آموزش دهیم که پیش بینی کند آیا مشتری جدیدی که وارد فروشگاه می شود جزو مشتریان ما باقی می ماند یا خیر؟

برنامه در دو فایل نوشته شده که در فایل اول با استفاده از jupyterlab مدل آموزش داده می شود و در یک فایل پیکل ذخیره می شود.

فایل دوم py. از کتاب خانه streamlit استفاده شده است. افراد غیر فنی آشنایی با پایتون و ... به همین دلیل یک برنامه تحت وب ساده برای آن ها درست کردم که فروشندگان به سادگی با وارد کردن اطلاعات مشتریان بتوانند نتیجه پیش بینی رو مشاهده کنند.

نمونه ای از برنامه نهایی

Ardalans churn predictor! :D

This is an online app for our marketing team

Whith this app they can predict if a new customer will stay with us or not

Please enter the customers information

Gender

☒ Male

☐ Female

Enter customers credit score

Nationality

☒ German

☐ French

☐ Spanish

How old is the customer?

Whats the customers tenure?

Whats the customers bank balance?

Whats the customers number of products

Does he/she have a credit card?

قدم اول:

ابتدا تمامی کتابخانه های استفاده شده درون برنامه را import می کنیم.

قدم دوم:

با استفاده از پانداز فایل دیتاست را به داخل برنامه import می کنیم.

قدم سوم:

ابتدا برای داشتن یک ایده کلی در مورد دیتاست ستون ها و instance های داخل آن ها به صورت رندوم مشاهده می کنیم. (۳ مورد)

در همین مرحله در ابتدای کار می توانیم چند تا از ستون ها را حذف کنیم زیرا برای مدل ما کاربردی نیستند. سپس چک می کنیم که چه ستون هایی هستند که مقدار غیر

عددی دارند. مشاهده می کنیم که ستون های ملیت و جنسیت دارای مقادیر غیر عددی هستند. در مرحله بعد بر روی آنان عملیات انکودینگ انجام می دهیم. علاوه بر این مشاهده می شود که در هیچ ستونی null value وجود ندارد که کار ما را بسیار ساده تر می کند.

قدم چهارم:

ابتدا قبل از انکود کردن و اسکیل کردن داده ها یک کپی از دیتاست می گیریم که بتوانیم با اسکیل مناسب آن را روی نمودار رسم کنیم. سپس داده ها را انکود و اسکیل می کنیم. (من برای راحتی خودم و عادت خودم دوباره دیتاست را دیتافریم می کنم.)

قدم پنجم:

چند سوال کلی می ورسیم و جوابشان را بر روی نمودار مشاهده می نماییم.

بین تمامی مشتریان ما چند نفر آقا و چند نفر خانم هستند؟
بین آقایان و خانم ها چه کسانی درآمد بیشتری دارند؟ مشاهده می شود که آنقدر جواب ها نزدیک هستند بهتر اس سوال را برای صد مشتری برتر بر اساس درآمد بررسی کنیم. با توجه به نمودار با دست آمده درآمد خانم ها مقداری بیشتر است. شهروندان چه کشوری پول بیشتری داخل حساب بانکی خود دارند؟ به نظر می آید فرانسوی ها پولدار تر باشند D:
آیا فرانسوی ها در بین ۱۰۰ مشتری برتر ما نیز قرار دارند؟
افراد چه کشوری قابل اعتماد ترند؟ جواب کوتاه : فرانسه

قدم ششم:

در این قسمت می خواهیم مدل خود را آموزش دهیم. می خواهیم از `pycaret` استفاده کنم برای این کار نیازی نبود داده ها را خودم انکود یا اسکیل کنم و یا حتی `correlation` بین ستون ها را به دست آورم ولی ترجیح دادم این مراحل را خودم انجام بدم.

قدم هفتم:

مدل را آموزش میدهم

قدم هشتم:

بهترین مدل را انتخاب می کنیم و پارامترهای آن را به دست می آوریم.

سپس میزان خطای مدل و ستون های تاثیر گذار را مشاهده می کنیم.

در آخر `confusion matrix` را رسم می کنیم که یک دید به نسبت عملکرد مدل داشته باشیم.

در آخر این مرحله مدل آموزش داده شده را در یک فایل `pickle` ذخیره می کنیم که بتوانیم از مدل در برنامه تحت وب خود استفاده کنیم.

قدم نهم:

یک فایل پایتون باز می کنیم و با استفاده از `streamlit` ورودی های فروشندگان خود را دریافت می کنیم و سپس با مدل ایمپورت شده نتیجه را به آن ها نمایش می دهیم که آیا مشتری پیش ما می ماند یا خیر

با تشكر از وقت و توجه شما