Module 3: Introduction to Machine Learning with Python

Case Study

Case Study

Domain – Retail/Fashion

focus – Optimize sales

Business challenge/requirement

Fyntra is the largest online clothing company in USA. It sells clothing online, but they also have in-store style and clothing advice sessions. Customers come into the store, have sessions/meetings with a personal stylist, then can go home and order either on a mobile app or website for the clothes they want.

Company wants to decide whether to focus the effort on mobile app experience or its website. As a drastic measure it is also evaluating to shut down the website.

You as a ML expert in the team will help the company make the right decision

Key issues
Clearly establish a correlation among the parameters supplied in data

Considerations
NONE

Data volume
Approx 500 records – file FyntraCustomerData.csv (All data is fake so do not worry about privacy)

Additional information
- NA

Business benefits
Increase in profits as the focus on the optimal sales channel will result into the higher top line and the higher bottom line

1. Compute -- Use seaborn to create a jointplot to compare the Time on Website and Yearly Amount Spent columns. Is there a correlation?
2. Compute – Do the same as above but now with Time on App and Yearly Amount Spent. Is this correlation stronger than 1st One?
3. Compute -- Explore types of relationships across the entire data set using pairplot . Based off this plot what looks to be the most correlated feature with Yearly Amount Spent?
4. Compute – Create linear model plot of Length of Membership and Yearly Amount Spent. Does the data fits well in linear plot?
5. Compute – Train and Test the data and answer multiple questions -- What is the use of random_state=85?
6. Compute – Predict the data and do a scatter plot. Check if actual and predicted data match?
7. What is the value of Root Mean Squared Error?
8. Final Question – Based on coefficients interpret company should focus more on their mobile app or on their website

Follow the Python Notebook **Module_1_Linear_Regression.ipynb** for solution