Project Meeting Minutes 5 - Week 10 & 11

Project Name: Multi Mart Revenue forecasting

Below is the summary of the analysis/visualization project we completed in 3rd semester.

The project focused on visualizing and analyzing sales and revenue data of Multi Mart Retail store spanning the period of 2019 to 2023. The primary objective was to provide a comprehensive understanding of sales performance, revenue generation, and customer behavior analysis to support informed decision-making, particularly regarding the potential expansion of its Loyalty Card program into new regions.

Git Hub Project Repository: DAB_Grp7_Capstone_Project

Week 10 and Week 11 Minutes of meeting are as follows:

Week 10:

Alisha James

- Used GradientBoostingRegressor and RandomForestRegressor for totalpurchases as target variable from the output of Lazy Predict Model.
- Performed GridSearch to identify the combination of best possible hyperparameter values for model tuning.
- Performed correlation between all the variables to identify any other potential target variable and features for the same.
- Identified a high value of negative correlation between 'hasloyaltycard' and 'feedbackscore' of -0.87 which was used for further machine learning models keeping 'hasloyaltycard' as the target.

Ikram Patel & Sujata Biswas

- Interim Presentation
    - Prepared presentation to showcase activities completed till 9th week in the project and upcoming activities.
    - Showcased a highlight of the final product.
- As per the output from Lazy Predict, tried GradientBoostingRegressor and RandomForestRegressor using totalpurchases as a target.
- Further performed GridSearch to identify best hyperparameters to tune the model.
- Researched on the correlation of variables again to identify if there are any other variables which can be used for prediction.
- Identified a high negative correlation between hasloyaltycard and feedback score of -0.87 and applied Machine learning models for predicting hasloyaltycard.

- This will be helpful to predict whether the customer will buy Loyalty card or not. So the store management can decide on the basis of certain parameters to suggest a loyalty card to the customer.
- Feature importance performed to be used with hasloyaltycard and below features were used to predict hasloyaltycard.
    - Totalrevenue
    - Totalpurchase
    - Frequency
    - Feedbackscore
    - Discountused
    - Referralsource
    - Churnindicator
- The following metrics were calculated to examine the model performance.
    - **R-squared** value using Random Forest model: 0.7845644755437593
    - **Precision:** 0.9067164179104478
    - **Recall:** 0.9067164179104478
    - **F1 Score:** 0.9067164179104478

Srikanth Ayyalasomayajula & Gayathri Manju Jayasena Kurup

- Prepared the design for final product, User Interface web portal.
- Identified the columns or questions which need to be added in the UI.
- Researched on Machine learning model deployment process using joblib and connecting the model data to user interface using flask.
- Researched on hosting the website so that it can be accessible using a URL from web instead of just local machine.

Group Work

- Attached is the analysis document attached.

Grp7_Draft_Analysis_
doc_upd.docx

Deliverables till date:

- Interim Report

Multi Mart Revenue
forecasting_Interim_R

- Interim Presentation

PDF
capestone_interim_gr
oup07.pptx.pdf

Week 11:

Date and Time	Location	Attendees	Professor
26th Mar'24 Tuesday 10:00 AM – 10:30 AM	In person meeting	Alisha James (0811919) Gayathri Manju Jayasena Kurup (0836679) Ikram Patel (0822315) Sujata Biswas (0832706) Srikanth Ayyalasomayajula (808545)	Abiodun Sodiq Shofoluwe

Discussion: Following points were discussed as part of the above meeting

- Since the target hasloyaltycard has a negative correlation with feedbackscore, which means the customer with no loyalty card is giving good feedback whereas the ones holding are giving low feedback scores.
- And it was observed that the data has more customers without loyaltycard (~8632) then with loyaltycard (~1368)
- Data is imbalance as per the above stats which is influencing the accuracy score.
- Need to identify statistical techniques such as k-fold Cross-Validation to balance the data and then apply Machine learning models.
- Also, another method was discussed to reduce the data set to 4000 records and apply Machine learning models on it to get the accurate scores.

Next Meeting schedule: 2nd Apr'24, Tuesday