Project Journal

Project Title: Data-Driven Insights for Banking: Financial and Consumer Behavior Analysis

Team Member Name: Sneha Mini Biju(x23323701)

Week 1

Date: 20/12/2024
Task Description:

- Created a histogram to compare account balances for accepted vs. not accepted proposals.
- Created a box plot of bank balances by job type, excluding zero balances. Showed that balance follows a lognormal distribution.

Time Spent: 5 hours

Challenges or Difficulties Encountered:

- Fine-tuning visualizations for clarity (e.g., adjusting axis limits and tick intervals).
- Understanding outliers in the balance column and their potential impact on the model.

Solutions or Actions Taken:

- Used descriptive statistics to understand the distribution and adjust plots accordingly.
- Customized visualizations to make them intuitive and meaningful for people.

Learning Outcomes:

- Strengthened data visualization skills and learned to interpret outlier behavior.

Week 2

Date: 23/12/2024
Task Description:

- Prepared the dataset for model training by dropping irrelevant columns (e.g., duration).
- Encoded categorical columns using LabelEncoder for job, marital, poutcome, month, contact, and education.
- Split the dataset into training and testing sets using train_test_split.
- Trained a RandomForestClassifier with 50 estimators and evaluated its accuracy.

Time Spent: 9 hours

Challenges or Difficulties Encountered:

- Balancing the dataset for better model performance.

- Understanding the feature importance scores generated by the classifier

Solutions or Actions Taken:

- Balanced the dataset by reviewing distributions and adjusting split ratios.
- Used feature importance plots to refine feature selection and improve model performance.

Learning Outcomes:

- Enhanced knowledge of Random Forest hyperparameters and the importance of balanced data.

Week 3

Date: 27/12/2024

Task Description:

- Created a SQLite database project.db to store the final processed data.
- Defined and executed SQL queries to create and populate the marketing table.
- Serialized the trained classifier and encoders using pickle for future use with Flask.
-
- Interpreted trends in heatmaps to identify key patterns:
    - High credit limit concentrations around specific "steps" (e.g., TWD 50,000, 200,000).
    - Clear age ranges associated with higher default frequencies.

Time Spent: 10 hours

Challenges or Difficulties Encountered:

- Learning SQLite syntax for inserting multiple rows efficiently.
- Verifying data integrity in the SQLite database after insertion.

Solutions or Actions Taken:

- Referred to SQLite documentation for best practices in batch insertion.
- Conducted validation checks post-insertion to ensure data consistency.

Learning Outcomes:

- Developed a deeper understanding of database integration and serialization techniques.

Week 4

Date: 3/01/2024

Task Description:

- Worked on the report sections: methodology, results, and evaluation.
- Ensured key points were well-articulated with clarity and in line with the project goals.
- Integrated insights from visualizations and model performance metrics.

Time Spent: 8 hours

Challenges or Difficulties Encountered:

- Methodology structuring to show logical progression.
- Condensing evaluation metrics into a summary.

Solutions or Actions Taken:

- Peer collaboration for feedback on the report content.
- Iterative revisions of drafts to ensure quality content.

Learning Outcomes:

- Enhanced skills in writing reports and the ability to clearly present findings.
- Learned how to relate project deliverables to business impacts.

Weekly Summary

Total Time Spent: 30+ hours

Key Accomplishments:

- Cleaned and transformed categorical data in the marketing dataset.
- Conducted exploratory data analysis with visualizations to derive insights from client behavior and financial data.
- Saved the processed marketing data into a SQLite database named project.db for better storage and retrieval.
- Serialized the trained classifier and encoders using pickle to be integrated into a Flask application for deployment.