Treeleaf Technologies Pvt. Ltd.
MACHINE LEARNING INTERNSHIP
QUALIFICATION TASK COMPLETION

Kushal Dahal

May 18 2024

Contents

1 Introduction

1.1 Purpose

The purpose of this project is to develop an interactive chatbot within a Jupyter Notebook that can engage users in a friendly conversation and provide predictions based on a trained machine learning model. Additionally, the chatbot can respond to statistical queries about the dataset used for training.

1.2 Scope

Implementing a basic interactive chatbot using ipywidgets,IPython.display. Training a machine learning model for prediction. Saving and loading model using joblib Enhancing the chatbot to provide statistical insights from the dataset.

2 Methodology

2.1 Data Preparation

Dataset: The dataset used for this task is an Excel file (xlsx format) containing data relevant to the project. The data includes various features that will be used to train and evaluate the machine learning model.

Data Loading: The dataset is loaded into the environment using appropriate libraries such as pandas, which facilitates the handling of Excel files and data manipulation.

Data Cleaning: Before splitting the data, it undergoes a cleaning process to ensure its quality and relevance. This step involves:

Handling missing values by either imputing with mean/median or removing the rows/columns with missing data. Removing duplicates to ensure the dataset's integrity. Filtering out irrelevant or outlier data points that could skew the model's performance.

Feature Engineering: New features are created from the existing data to improve the model's predictive power. This may include:

Encoding categorical variables using techniques such as ordinal encoding. Creating interaction terms or polynomial features if necessary.

Data Preprocessing: After cleaning and feature engineering, the dataset is prepared for model training. This involves:

Splitting the data into features (X) and target (Y). The features (X) represent the independent variables, while the target (Y) represents the dependent variable we aim to predict.

2.2 Model Training

Model Used: Random Forest Classifier.
Training: The model is trained on the training dataset.
Evaluation: Model accuracy is assessed using the test split of dataset

2.3 Chatbot Implementaion

Libraries Used: ipywidgets, pandas, numpy, scikit-learn.
Functionality:
Greeting the user.
Making predictions based on predefined input features. Providing statistical insights (mean, median, mode) of specific columns in the dataset.

3 Results

3.1 Model Performance

Accuracy: The model achieved an accuracy of 97.90% on the test set.

3.2 Chatbot interactions

Example Interactions:
Greeting:
User: "Hello"
Chatbot: "Hi there! How can I help you today?"
User: "Predict"
Chatbot: {input fields to enter data} and "Congrats! Personal Loan can be credited to You!" or "Sorry, Personal Loan cannot be credited to You."
Statistical Query:
User: "What is the mean of column_name ?"
Chatbot: "The mean of column_name is mean_value "

4 Conclusion

4.1 Summary

The project successfully demonstrates an interactive chatbot that can make predictions using a machine learning model and provide statistical insights from the dataset.

4.2 Future Work

Future enhancements could include:

Improving natural language processing to better understand user queries
Larger dataset for better model evaluation