Report on Predictive Modeling for Term Deposit Subscription

Project Title

Client Term Deposit Subscription Prediction Using Machine Learning

Objective

The main goal of this project is to develop a predictive model that determines the likelihood of a bank client subscribing to a term deposit product, using historical marketing data.

Step-by-Step Breakdown of the Notebook

1. Environment Setup

The notebook installs and configures necessary libraries like scikit-learn and imbalanced-learn. Due to compatibility issues, specific versions are enforced.

2. Data Loading

The dataset used is 'bank.csv'. The target variable is 'y', indicating subscription status.

3. Data Cleaning and Preprocessing

Whitespace issues are resolved in column names. Categorical variables are encoded and numeric features are scaled using a ColumnTransformer. This preprocessor is saved using pickle.

4. Handling Imbalanced Data

SMOTE (Synthetic Minority Oversampling Technique) is used to handle class imbalance, ensuring better model performance.

Report on Predictive Modeling for Term Deposit Subscription

5. Model Training and Evaluation

Multiple models including Logistic Regression, Decision Tree, Random Forest, and Gradient Boosting are evaluated. Gradient Boosting gives the best performance and is saved as 'best_model.pkl'.

6. API Deployment

A Flask API is built with two endpoints: '/' confirms the API is live, and '/predict' accepts JSON input and returns prediction probability. The app is designed for local or public deployment.

7. Common Issues Encountered

Key issues included feature mismatch due to different transformations during training and prediction, and invalid localhost URLs for public access.

Final Recommendations

1. Ensure consistent preprocessing at training and prediction stages.
2. Host your model publicly using services like Render or Heroku.
3. Use Postman or Python's requests to test APIs.

Files Used

bank.csv, preprocessor.pkl, best_model.pkl, Solution.ipynb

Outcome

A machine learning pipeline was built and deployed to predict term deposit subscriptions. With proper hosting and data handling, it can be used for real-time predictions.