DefaultShield: Business Loan Default Prediction

By: Dhevina Tailor

Predictive Business Loan Default Analysis

Problem Statement: In the financial sector, predicting business loan defaults is crucial. Inaccurate predictions can lead to significant financial losses.

Proposed Solution: The project introduces a data-driven machine learning approach to accurately predict loan defaults, utilizing historical loan data.

Potential Impact: By improving prediction accuracy, financial institutions can reduce risks, optimize loan approval processes, and support more informed decision-making.

Overview of Dataset and Preprocessing

Dataset Overview: The analysis is based on a dataset of over four hundred thousand business loans, including features like business type, industry type, and revolving line of credit.

Preprocessing Steps: Performed data cleaning, handled missing values, conducted feature engineering for categorical variables, and applied SMOTE for addressing class imbalance.

```
# Scaling the features
scaler = StandardScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_validation_scaled = scaler.transform(X_validation)
X_test_scaled = scaler.transform(X_test)
```

```
# Oversampling the minority class using SMOTE
smote = SMOTE(random_state=1)
X_train_smote, y_train_smote = smote.fit_resample(X_train_scaled, y_train)
```

Exploratory Data Analysis Highlights

Key Findings: EDA revealed significant insights: Industry type strongly correlates with default rates, and larger loan amounts are more prone to default.

Modeling and Performance Metrics

Baseline Models: Evaluated several models including Logistic Regression, Decision Tree, and KNN.

Evaluation Approach: The models were assessed based on accuracy, precision, recall, and F1-score to gauge their effectiveness in predicting defaults.

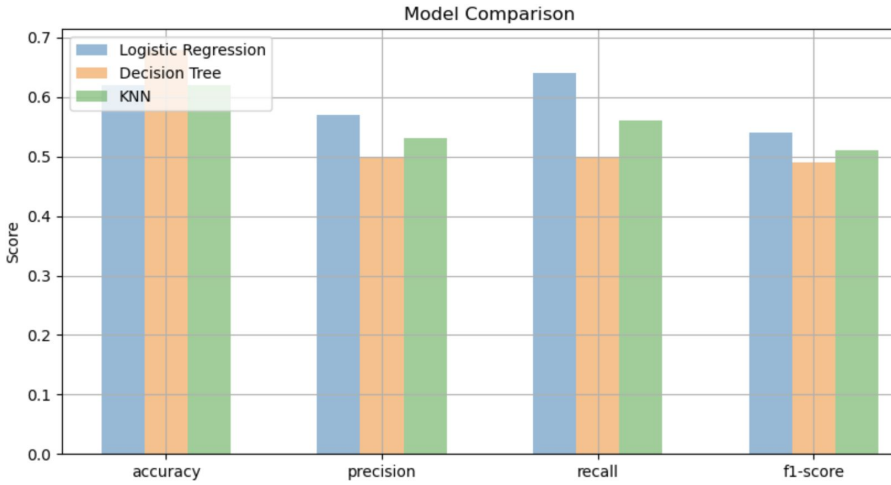
Model Performance: Logistic Regression showed promising results, especially in terms of precision and recall balance.

Classification Report on Validation Set:

	precision	recall	f1-score
0	0.24	0.67	0.36
1	0.91	0.61	0.73
accuracy			0.62
macro avg	0.57	0.64	0.54
weighted avg	0.80	0.62	0.67

Confusion Matrix on Validation Set:

```
[[11317  5594]
 [35467 54560]]
```



Next Steps for Advanced Modeling and Implementation

Advanced Modeling Plans: Moving forward, I plan to explore deep learning models and ensemble methods for improved accuracy.

Productization Strategy: The ultimate goal is to integrate this predictive model into a financial risk assessment tool for lenders.

Final Thought: The work paves the way for more reliable and data-driven financial decision-making.

Thank You!