

Project Report: Credit Card Approval Prediction

Introduction

The goal of this project is to predict credit card approval using a dataset containing various applicant features. The dataset is analyzed using machine learning techniques, specifically logistic regression, to determine the likelihood of approval based on the input features.

Dataset

The dataset used in this project is **cc_approvals.csv**, which contains 690 entries and 16 features. The features include various applicant attributes, such as income, age, and other categorical data.

Data Inspection

- The first few rows of the dataset were inspected using **head()**, revealing the structure and initial values of the dataset.
- Descriptive statistics were generated using **describe()** to summarize the central tendency, dispersion, and shape of the dataset's distribution.

Data Summary

- The dataset consists of 690 entries with a mix of data types: 3 numerical (float64, int64) and 12 categorical (object).
- The descriptive statistics show the mean, standard deviation, and range of the numerical features.

Data Preprocessing

- Missing values were identified and handled appropriately to ensure the dataset is clean before model training.
- The features were rescaled using a scaler to improve the performance of the logistic regression model.

Model Selection

Logistic Regression

- A logistic regression model was instantiated using default parameters.
- The model was fitted to the training dataset, and predictions were made on the test set.

Model Evaluation

- The accuracy of the logistic regression classifier was calculated, achieving an accuracy of approximately **84.21%**.
- A confusion matrix was generated to visualize the performance of the model in terms of true positives, true negatives, false positives, and false negatives.

Hyperparameter Tuning

- A grid search was performed using **GridSearchCV** to optimize the hyperparameters of the logistic regression model.
- The parameters tuned included **tol** (tolerance) and **max_iter** (maximum iterations).
- The best model achieved an accuracy of approximately **85.36%** using **max_iter=100** and **tol=0.01**.

Conclusion

The logistic regression model demonstrated a good performance in predicting credit card approvals with an accuracy of 84.21% initially, which improved to 85.36% after hyperparameter tuning. Future work could involve exploring more complex models or incorporating additional features to further enhance prediction accuracy.

Recommendations

- Investigate the impact of additional features on model performance.
- Consider using more advanced models such as Random Forests or Gradient Boosting for potentially better accuracy.
- Implement cross-validation to ensure the model's robustness.