Predictive analysis of credit card approval by machine learning algorithms.

University of Hertfordshire UH

By : Bubly Babu (22031115)

Supervisor : Dr Stephen Kane

Introduction :

The traditional credit card approval processes are coming under increasing scrutiny in the modern financial world because of their insufficient capacity to comprehend modern financial habits. This study aims to evaluate the predictive power of several models for credit card acceptance outcomes by employing advanced data science approaches, particularly machine learning. Through examination and comparison of these models, the study seeks to identify ways to enhance the accuracy and efficacy of credit assessment processes, consequently promoting improved decision-making in financial institutions.

Objective:

- Utilize advanced data science techniques to enhance credit card approval procedures.
- Assess the predictive capabilities of three machine learning algorithms for credit card approval.
- Improve decision-making frameworks in financial institutions for more dependable credit card approvals.

Aim:

To identify the most effective machine learning model for predicting the likelihood that new credit card applicants will be approved or denied by banks.

Dataset:

The study's dataset, which comes from the UCI Machine Learning Repository, includes a variety of factors relevant to credit card applications. It provides a substantial dataset for researching credit card approval procedures with 690 instances and 15 features. Notably, attribute names and values have been anonymized to protect data privacy. Although the names of the attributes have been anonymized, the blog link provided a concise summary of each feature. The project incorporates ethical considerations throughout.

Algorithm
- Logistic Regression

Data Split
- Train **70%**
- Test **30%**

Metrics
- Accuracy
- Precision
- Recall
- F1 – score
- Area under ROC curve (AUC)

Result
- Training Accuracy: **86.34%**
- Testing Accuracy: **85.02%**

Litreature Review:

Title	Author and Year	Models used	Results
Credit Card Approval Predictions Using Logistic Regression, Linear SVM and Naïve Bayes Classifier	(Zhao, 2022)	Logistic Regression (L1 Penalty) Logistic Regression (L1 Penalty) Logistic Regression (elasticnet penalty) Linear SVM Navie Bayes Classifier	Logistic Regression (L1 Penalty) – 87.27% Logistic Regression (L1 Penalty) – 86.06% Logistic Regression (elasticnet penalty) – 86.67% Linear SVM - 88.48% Navie Bayes Classifier – 84.24%
Credit Card Assent Using Supervised Learning	(Therese et al., 2022)	KNN Decision Tree Naive Bayes Classifier Logistic Regression	KNN – 98.50% Decision Tree -98.13% Naive Bayes Classifier – 99.62 % Logistic Regression – 84.32%
Performance Evaluation and Practical Use of Supervised Data Mining Algorithms for Credit Card Approval	(Duan, 2020)	Logistic Regression Decision Tree(Gini-index based) Decision Tree(information based) KNN Neural Network	Logistic Regression – 75.73% Decision Tree(Gini- index based) – 73.3% Decision Tree(information based) – 75.73% KNN – 67.48% Neural Network -76.7%
Comparison of Different Supervised Machine Learning Classifiers to Predict Credit Card Approvals	(Bansal and Punjabi, 2021)	Logistic Regression Random Forest Decision Tree Neural Network Support Vector Gradient Boost XGBoost	Logistic Regression – 85.362% Random Forest – 88.754% Decision Tree – 85.942% Neural Network – 88.592% Support Vector – 85.507% Gradient Boost – 85.507% XGBoost – 92.753%
Credit Card Default Prediction using Machine Learning Techniques	(Sayjadah et al., 2018)	Logistic Regression Rpart Decision Tree Random Forest	Logistic Regression – 82% Rpart Decision Tree -82.06% Random Forest – 81.81%

Data Collection → **Exploratory Data Analysis** → **Data Preprocessing** → **Model Implementation** → **Evaluation Metrics**

Methodology:

Initially, the study conducted thorough exploratory analysis on the credit approval dataset from the UCI Machine Learning Repository. This included metadata examination, variable analysis, and dataset structure assessment. Univariate and bivariate analyses were performed to understand variable distributions and relationships. Additionally, a correlation matrix was utilized to identify features correlated with the target variable.

Careful preprocessing was carried out after EDA to guarantee data integrity and model readiness. This involved renaming columns, concatenating features with the target variable, and rearranging columns. Outliers were handled with the proper methods, missing values were imputed, and categorical variables were encoded. Furthermore, the data was split into training and testing sets to facilitate model training and evaluation.

With preprocessed data, the study applied the Logistic Regression model to predict credit card approval decisions. Hyperparameter tuning was then conducted to optimize model performance, enhancing its predictive capabilities. The model's performance was evaluated using various metrics. Comparisons between training and testing accuracies were made to assess overfitting or underfitting tendencies.

What's Next?

- Proceed with implementing decision tree and random forest models for credit card approval prediction.
- Conduct a comparative analysis to assess the performance of logistic regression, decision tree, and random forest models.
- Determine the optimal model for credit card approval prediction based on the comparative evaluation.

Comparison of Evaluation Metrics (Train vs. Test)

Metric	Training Data	Testing Data
Accuracy	0.863	0.850
Precision	0.894	0.841
Recall	0.862	0.880
F1-score	0.878	0.860
ROC AUC Score	0.864	0.849

Table 1 : Evaluation Metrics

Dataset Link :
https://archive.ics.uci.edu/dataset/27/credit+approval

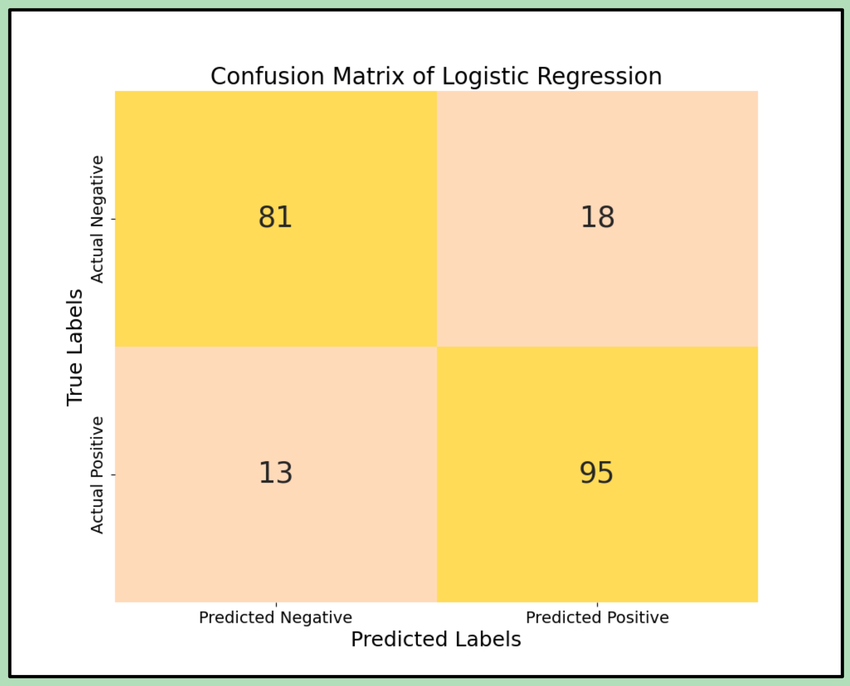

Figure 1 : Confusion Matrix

Reference:

Bansal, S. and Punjabi, T. (2021). *Comparison of Different Supervised Machine Learning Classifiers to Predict Credit Card Approvals*. *International Research Journal of Engineering and Technology*.

Duan, L. (2020). Performance Evaluation and Practical Use of Supervised Data Mining Algorithms for Credit Card Approval. *2020 International Conference on Computing and Data Science (CDS)*. doi:https://doi.org/10.1109/cds49703.2020.00057.

rstudio-pubs-static.s3.amazonaws.com. (n.d.). *Analysis of Credit Approval Data*. [online] Available at: https://rstudio-pubs-static.s3.amazonaws.com/73039_9946de135c0a49daa7a0a9eda4a67a72.html [Accessed 29 Feb. 2024].

Sayjadah, Y., Hashem, I.A.T., Alotaibi, F. and Kasmiran, K.A. (2018). Credit Card Default Prediction using Machine Learning Techniques. *2018 Fourth International Conference on Advances in Computing, Communication & Automation (ICACCA)*. doi:https://doi.org/10.1109/icaccaf.2018.8776802.

Therese, M.J., Devi, A., Gurulakshmi, R., Sandhya, R. and Dharanyadevi, P. (2022). Credit Card Assent Using Supervised Learning. *2022 International Conference on Smart Technologies and Systems for Next Generation Computing (ICSTSN)*. doi:https://doi.org/10.1109/icstsn53084.2022.9761307.

Zhao, Y. (2022). Credit Card Approval Predictions Using Logistic Regression, Linear SVM and Naïve Bayes Classifier. *2022 International Conference on Machine Learning and Knowledge Engineering (MLKE)*. doi:https://doi.org/10.1109/mlke55170.2022.00047.