Evaluation of a Logistic Regression Model for Loan Approval Decisions

Anthony Cade Streich

CSCI 396: AI Topics
Instructor: Dr. Sherine Antoun
Colorado Mesa University

May 2025

1. Introduction

This project aims to develop and evaluate an algorithmic decision-making system for loan approval using the German Credit dataset through the use of the sklearn library [5]. The main objective is to predict the likelihood of loan default, allowing financial institutions to make informed lending decisions. Logistic Regression was selected because of it's usefulness as a baseline model in machine learning pipelines. Logistic Regression also provides simplicity and effectiveness on smaller datasets [1]. Although real life loan approval models are going to be larger and more complex, this model will provide insight to the challenges one will face when implementing a similar model. In addition to assessing model performance, this project explores fairness, limitations, and ethical concerns to ensure the system aligns with responsible AI practices.

2. Dataset Overview

The dataset is from German Credit Data, a widely recognized benchmarking company in credit risk modeling. It contains 1000 data rows and 20 input columns, including demographic, financial, and employment-related attributes. Although a small data set, the aim of this project is to explore the implications and impacts of a loan approval model, and 1000 data entries will be enough accomplish our goals [2]. The target variable, `CreditRisk`, is binary: originally coded as 1 (good) and 2 (bad), it was remapped to 0 (good credit risk) and 1 (bad credit risk) for compatibility with scikit-learn conventions.

Categorical features include variables like `StatusCheckingAcc`, `CreditHistory`, `Savings`, and `ForeignWorker`, while numerical features include `Duration`, `CreditAmount`, and `Age`. The preprocessing pipeline handled categorical features using one-hot encoding, and numerical features were passed through without transformation.

3. Modeling Approach

Although we typically used R-script for exploratory data analysis, this data set has already been preprocessed [1]. Because of this, we can use EDA through python and the scikit-learn library. After this, a Logistic Regression model was implemented, also using the sklearn pipeline [5]. The steps included:

- **Preprocessing:** Categorical columns were one-hot encoded with `handle_unknown='ignore'`.
- **Data splitting:** standard practice splitting with 80% training and 20% testing, stratified by target variable.
- **Model training:** Logistic Regression with `liblinear` solver and `max_iter=1000`.

The model outputs probabilities for each class, which are then converted into binary predictions for evaluation.

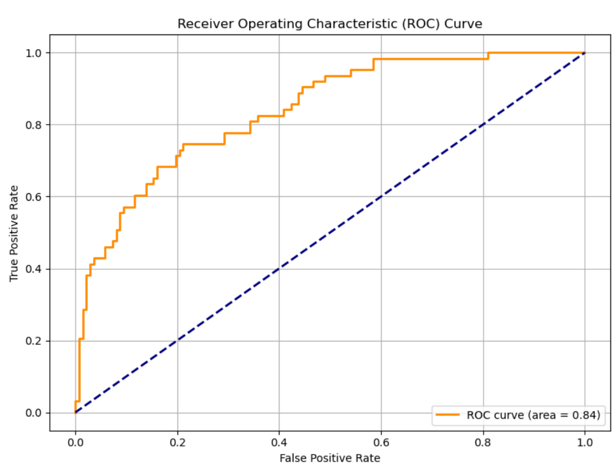
4. Model Performance

Although this model did not yield the most promising results, these were the highest scores I was able to achieve all semester. The classification report on the test set yielded the following results:

	precision	recall	f1-score	support
0	0.79	0.93	0.86	137
1	0.76	0.46	0.57	63
accuracy			0.79	200
macro avg	0.78	0.70	0.72	200
weighted avg	0.78	0.79	0.77	200

- **Accuracy:** 79%
- **Precision (Good Credit Risk):** 79%
- **Precision (Bad Credit Risk):** 76%
- **Recall (Good Credit Risk):** 93%
- **Recall (Bad Credit Risk):** 46%
- **F1-Score (Good):** 86%
- **F1-Score (Bad):** 57%

These results indicate that the model is better at identifying good credit risks than bad ones. The relatively low recall for bad credit risks suggests potential risk in misclassifying risky applicants. This is especially evident after viewing the ROC Curve Graph:



5. Feature Importance and Interpretation

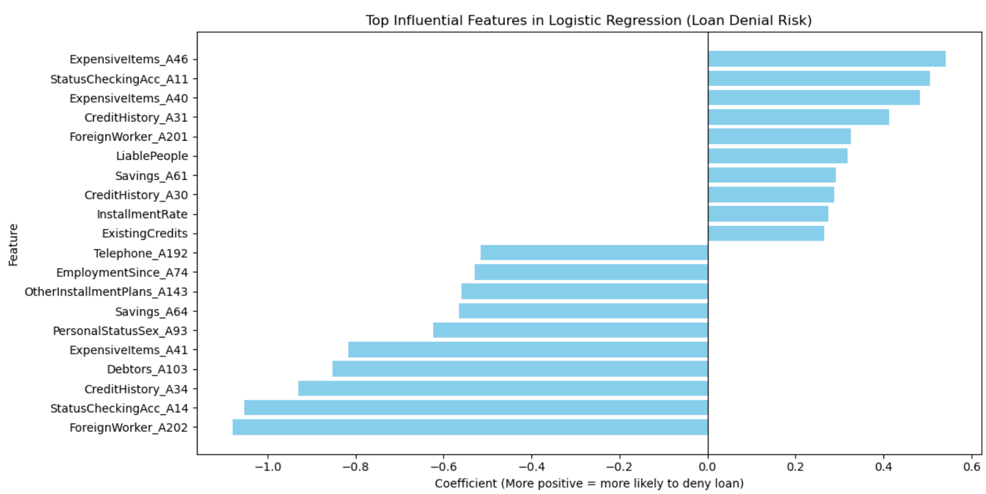
Logistic regression coefficients were extracted and sorted. Positive coefficients indicate a higher likelihood of loan denial; negative coefficients indicate higher likelihood of approval.

Top positive influencers (more likely to deny):

- `ForeignWorker = Yes`
- `StatusCheckingAcc = None`
- `CreditHistory = Critical`
- `Job = Unskilled`
- `PersonalStatusSex = Male, Single`

Top negative influencers (more likely to approve):

- `ExpensiveItem = Owns new car`
- `Savings = Rich`
- `EmploymentSince =>7 years`
- `Property = Owns Real Estate`
- `Housing = Own`
- `Telephone = Yes`



This visualization reveals an important imbalance: there are more factors that increase the likelihood of loan denial than factors that improve approval odds. While positive indicators (like maintaining a rich savings account or having over seven years of continuous employment) do reduce the risk score, their influence is limited compared to the number and strength of negative risk-related features. This suggests that the model places greater weight on penalizing risk factors rather than rewarding "positive" attributes. As a result, individuals from less privileged backgrounds, who may lack long employment history or significant financial assets, are continuously disadvantaged, reinforcing wealth inequality [3].

6. Fairness Evaluation

Several features in the dataset served as proxies for "combined" attributes:

- `PersonalStatusSex` encoded gender and marital status.
- `ForeignWorker` encoded nationality.
- `Age` can indirectly disadvantage certain age groups.

The high positive coefficient for `ForeignWorker = Yes` raises fairness issues, with similar concerns applying to gender-encoded features. Mitigation strategies include removing sensitive features, applying fairness-aware algorithms, or post-processing outputs.

7. Limitations

- **Linearity Assumption:** Logistic Regression assumes linearity. This means that the model doesn't automatically map complex interactions or non-linear effects (like diminishing returns or thresholds).
- **High Dimensionality:** One-hot encoding increases dimensionality, risking memory constraints, but this only becomes an issue with larger data sets.

8. Ethical and Societal Implications

Algorithmic loan approval offers efficiency but also risks:

- **Discrimination:** Sensitive features may cause disparate treatment.
- **Transparency:** Logistic Regression is explainable, but excessive dummy variables reduce clarity.
- **Accountability:** Models must be audited regularly, preferably by an unbiased third party [4].

9. Conclusion and Future Directions

The model performs moderately well but presents fairness and ethical challenges. While interpretable, it relies on features that may bias results.

Improvements include:

- Excluding sensitive features.

- Applying fairness constraints.

Future research should incorporate fairness metrics and real-world validation to support the responsible deployment of AI models. Without proper safeguards, deploying this model in a real-world setting could reinforce existing societal biases embedded in the training data, ultimately worsening inequality. According to IBM (2023), over 42% of enterprises have already implemented AI technologies. This trend is concerning, as public awareness of AI's negative capabilities remains limited.

10. Discussion Questions

Model Fairness and Bias

1. Should models be allowed to use features like nationality, gender, or age—even if they improve accuracy?
2. How can we ensure fairness if excluding sensitive features reduces model performance?

Ethical and Societal Implications

1. Who should be held accountable when an algorithm denies a loan unfairly—data scientists, institutions, or the model itself?
2. Do the benefits of automated decision-making outweigh the risks in high-stakes applications like lending?
3. Should applicants be entitled to a full explanation of their denial—even if the model is complex?

Broader Impact

1. How might deployment of this model impact financial inclusion for marginalized communities?
2. If this model were deployed globally, what cultural or regulatory issues would need to be considered?

11. Sources

1. Dua, D., & Graff, C. (n.d.). *Statlog (German Credit Data) Data Set*. UCI Machine Learning Repository. https://archive.ics.uci.edu/dataset/144/statlog+german+credit+data

2. Feldman, M., Friedler, S. A., Moeller, J., Scheidegger, C., & Venkatasubramanian, S. (2015, August). *Certifying and removing disparate impact*. In Proceedings of the 21st ACM SIGKDD International Conference on Knowledge Discovery and Data Mining (pp. 259–268).

3. IBM. (n.d.). *AI decision-making: Where do businesses draw the line?* IBM. https://www.ibm.com/think/insights/ai-decision-making-where-do-businesses-draw-the-line

4. IEEE Standards Association. (n.d.). *IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems*. https://standards.ieee.org/industry-connections/activities/ieee-global-initiative/

5. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). *Scikit-learn: Machine learning in Python*. Journal of Machine Learning Research, 12, 2825–2830.