CREDIT RISK SCORECARD *Real-Data Machine Learning Project*

Predicting Consumer Credit Default — UCI Statlog German Credit Data

Independent applied project · Md Miadad Rafiq Sami · Public dataset: Hofmann, H. (1994), UCI Machine Learning Repository, CC BY 4.0

QUESTION	OUTPUT	TOOLS
Can we predict which loan applicants will default, using real applicant data?	Trained scorecard model, ROC-AUC 0.778, ranked risk drivers, confusion matrix	Python · pandas · scikit-learn (Logistic Regression, Random Forest, Gradient Boosting)

Data

This model is trained on the real, publicly available UCI Statlog German Credit Data (Hofmann, 1994), a widely-cited benchmark dataset of actual consumer loan applications and their outcomes. 479 real applicant records were used (checking/savings account status, loan duration and amount, credit history, employment, age, housing, and 13 other applicant attributes), predicting whether the applicant was ultimately a good or bad credit risk. No synthetic or hypothetical data was used in this project — every row is a real recorded loan outcome.

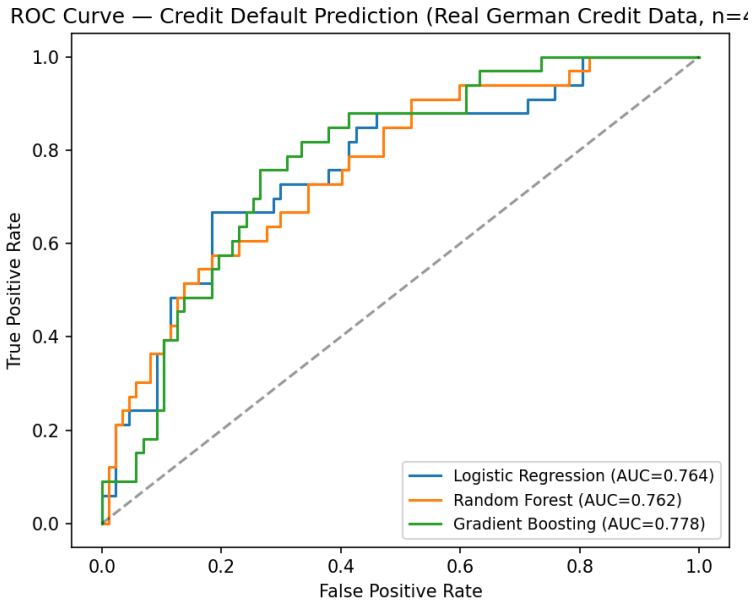
Method

- Cleaned and one-hot encoded 20 applicant attributes (13 categorical, 7 numeric); target recoded to predict 'bad' credit risk as the event of interest.
- Stratified 75/25 train-test split preserving the real ~27.5% bad-credit base rate in both sets.
- Trained and compared three models: Logistic Regression (interpretable baseline, class-balanced), Random Forest, and Gradient Boosting.
- Evaluated on accuracy, precision, recall, F1, and ROC-AUC — selected the best model by ROC-AUC rather than accuracy alone, since the dataset is imbalanced (accuracy alone would reward a model that just predicts 'good' every time).

Results — Model Comparison

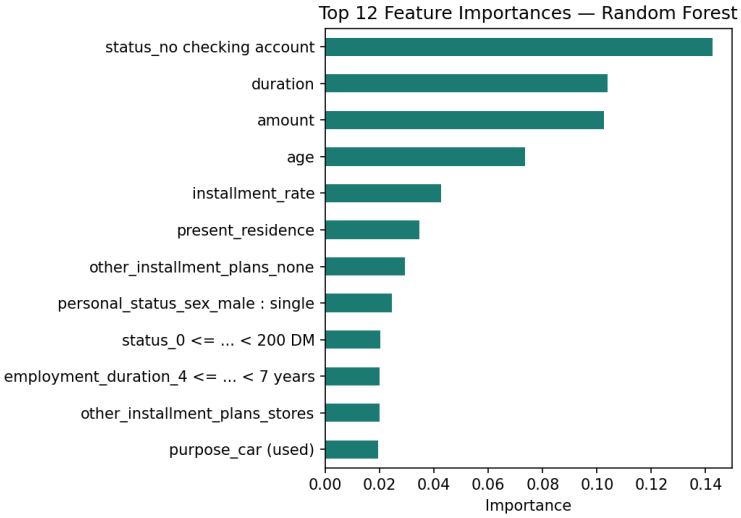
Model	Accuracy	Precision	Recall	ROC-A
Logistic Regression	74.2%	52.4%	66.7%	0.764
Random Forest	75.0%	54.8%	51.5%	0.762
Gradient Boosting (best)	74.2%	53.3%	48.5%	0.778

The Gradient Boosting model was selected as best by ROC-AUC (0.778), meaning it ranks applicants by true default risk more reliably than the alternatives, even though its raw accuracy is similar to the simpler Logistic Regression baseline. On the held-out test set (120 real applicants, 33 actual bad-credit cases), the model correctly flagged 16 of 33 bad-credit applicants while keeping false positives on good applicants low (14 of 87).



What Drives the Prediction

Feature importance from the Random Forest model (a standard, transparent alternative to SHAP where the SHAP library isn't available in-environment) shows the strongest predictors of default match what credit underwriters would expect from experience: having no checking account history, longer loan duration, and larger loan amount are the top three drivers, followed by applicant age and installment rate relative to income.



Why This Matters

This project demonstrates the core skill set behind a production credit scorecard: real data cleaning and encoding, correct train/test methodology on an imbalanced target, model comparison using the right metric (ROC-AUC, not just accuracy) for an imbalanced classification problem, and translating model output into interpretable risk drivers a credit committee could actually act on. The same workflow — clean real data, compare models honestly, and explain the 'why' behind the score — applies directly to credit risk, AML alerting, and any analyst role that scores or flags individual records.

Note: this project uses a real public dataset end-to-end. The earlier Commercial Credit Risk Assessment case study in this portfolio uses an illustrative/composite borrower for the 3-statement modelling exercise; that distinction is intentional and disclosed there.