Credit Risk Scoring: A Stacking Generalization Approach

Bernardo Dias Raimundo

Prof. Dr. Jorge Miguel Ventura Bravo

Nova Information Management School

World Cist 2023

1

Introduction

A momentary overview

01

Credit Risk
A significant concern for financial institutions

02

Banking Regulations
The revised Basel accords and it's contributions

03

Credit Scoring Methods
Regression models and machine learning applications

04

Ensemble Learning
Homogeneous and heterogenous ensembles

1

Introduction

Research questions

01 **How does an ensemble-based approach compare to single based classifiers when predicting the probability of default?**

02 **Given the ensemble-based approach, how sensitive is the outcome of the ensemble to different model combinations?**

2 Literature Review

How is the current research in credit risk modeling?

01 Altman, E. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4): 589–609.

02 Wolpert, D. (1992). Stacked Generalization. Neural Networks. 5. 241-259.

03 Dietterich, T.G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems: First International Workshop, MCS 2000, Lecture Notes in Computer Science. 1-15.

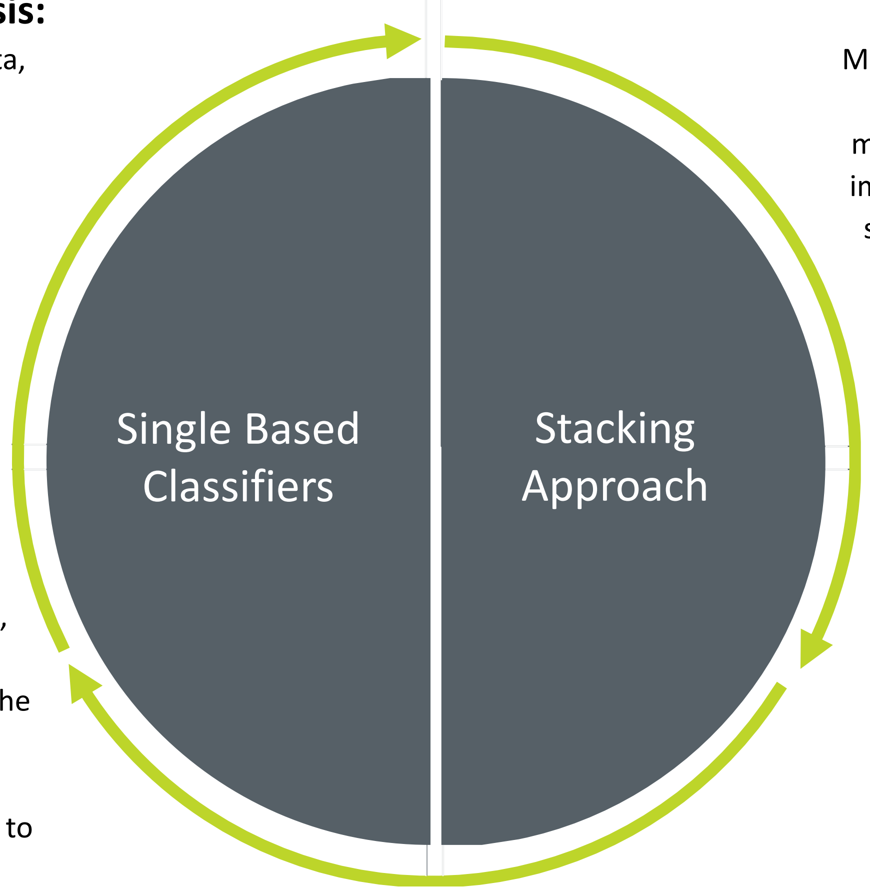
04 Steel, M. F. J. (2020). Model Averaging and Its Use in Economics. Journal of Economic Literature, 58, 644–719.

05 Ashofteh, A., & Bravo, J. M. (2021). A conservative approach for online credit scoring. Expert Systems with Applications, 176, 1-16.

3 Methodology

Exploratory Data Analysis:
Deep dive to understand the data, uncover interesting and hidden patterns and data visualization.

Data Preprocessing:
Machine learning pipelines will be used to split the process into a multistep workflow. This includes imputation, categorical encoding, scaling and principal component analysis.



Modeling:
Four different algorithms will be considered: K nearest neighbors, decision tree, support vector machine as level 0 models and the logistic regression as the meta model. In addition to this, a sampling technique was applied to deal with class imbalance.

Performance Metrics:
Performance metrics such as accuracy, precision, recall and f1 score will be used to compare the performance of both approaches.

Machine Learning application to online credit scoring

Exploratory data analysis

Summarized details of the dataset

Population amount

The dataset was provided by Lending Club. Initially it contained 2.26 million rows and 151 different features.

Features description

Personal details (e.g., address, employment status, home ownership)

Credit history (e.g., the balance of accounts and revolving and current past due accounts)

Loan characteristics (e.g., purpose, grade, term)

Dataset transformation

Population amount

Removing features that contained more than 25% of missing data

Only loans with a final "status" and features available at the time of application were considered

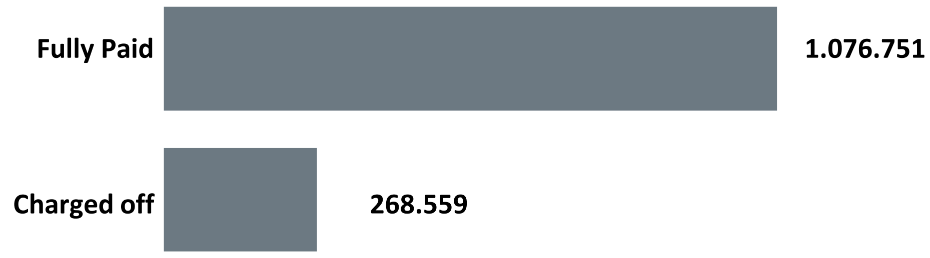
The final dataset contains 1,345 rows and 33 features

4

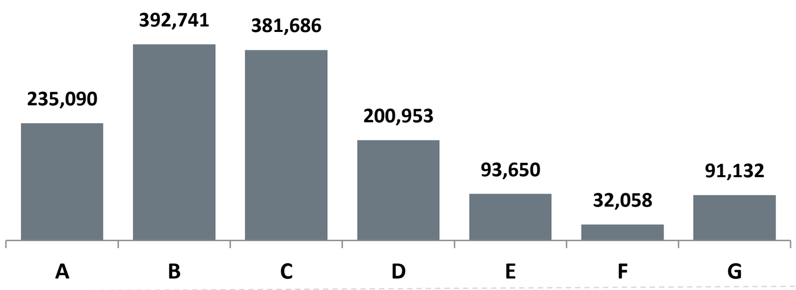
Machine Learning application to online credit scoring

Exploratory data analysis

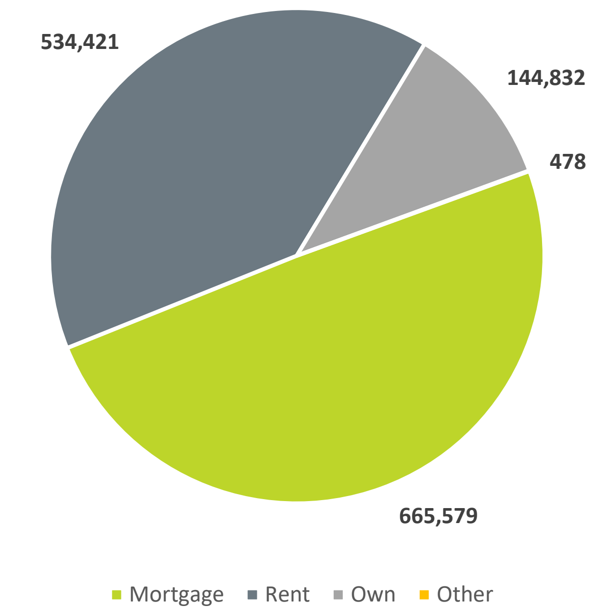
Loan status distribution



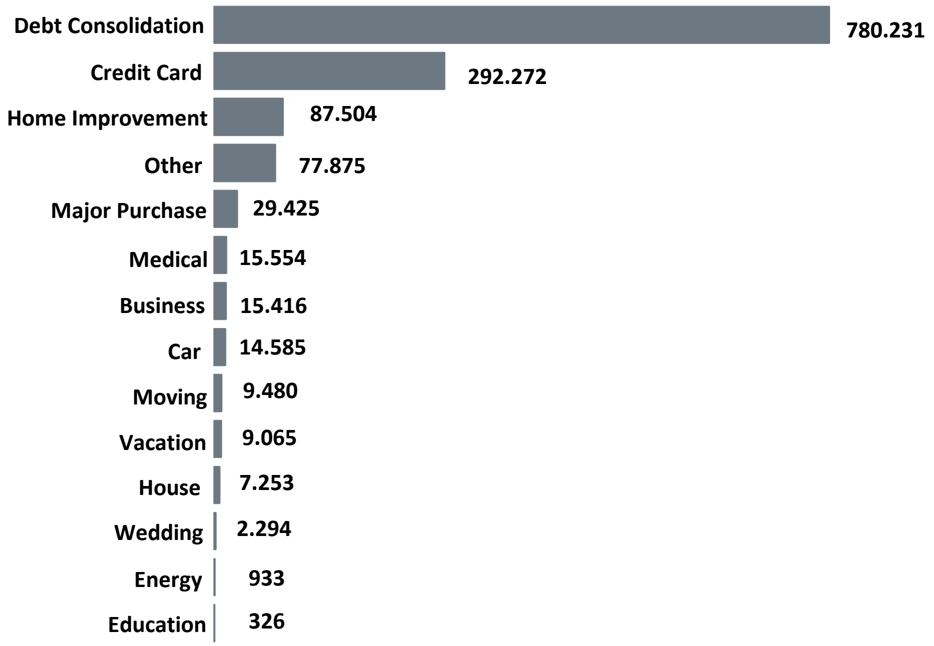
Grades



Home Ownership

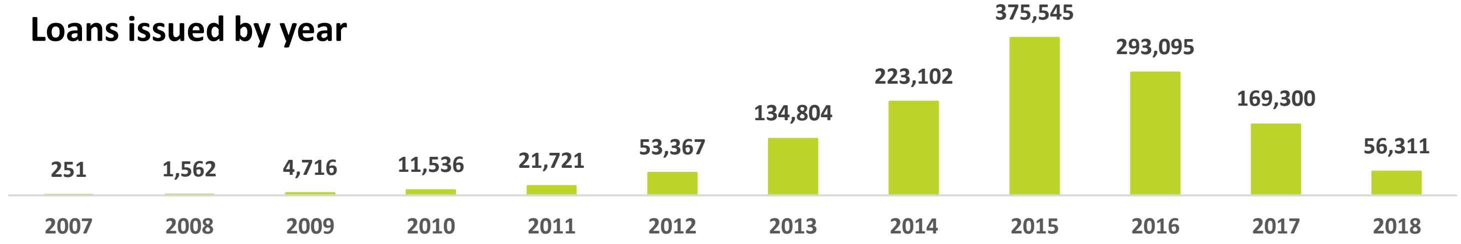


Purpose Occurance



Loan term

Loans issued by year

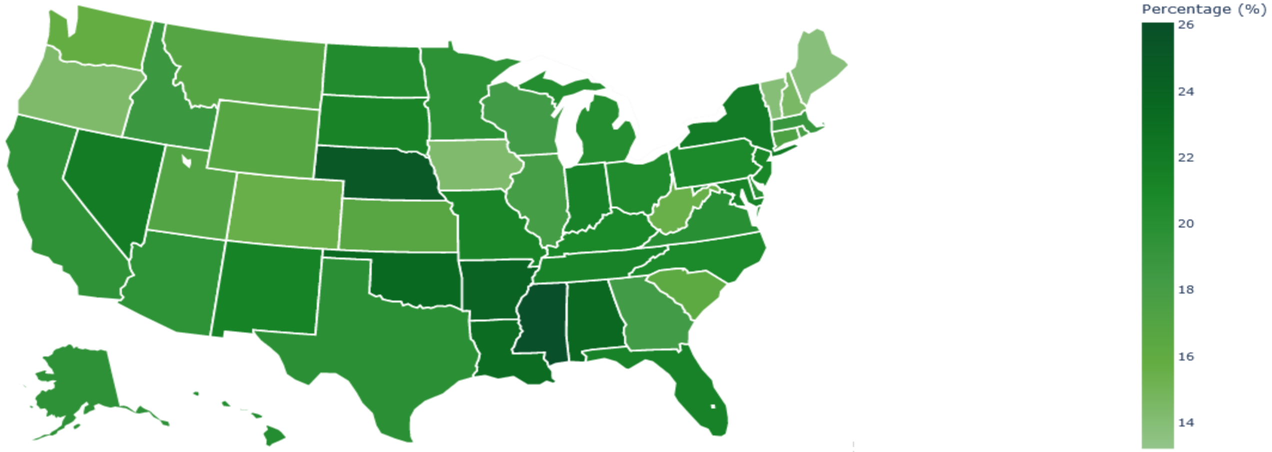


4

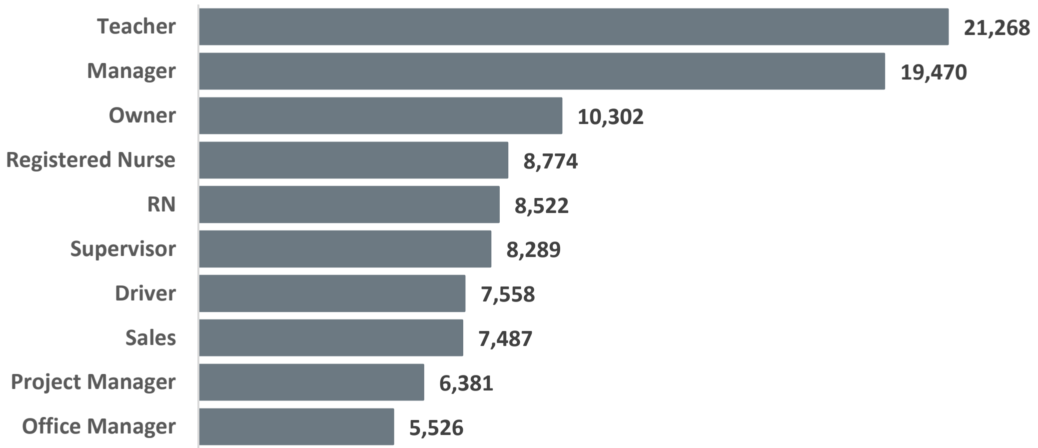
Machine Learning application to online credit scoring

Exploratory data analysis

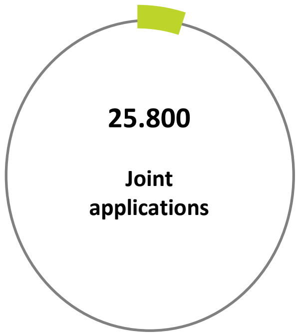
Percentage of charged off loans



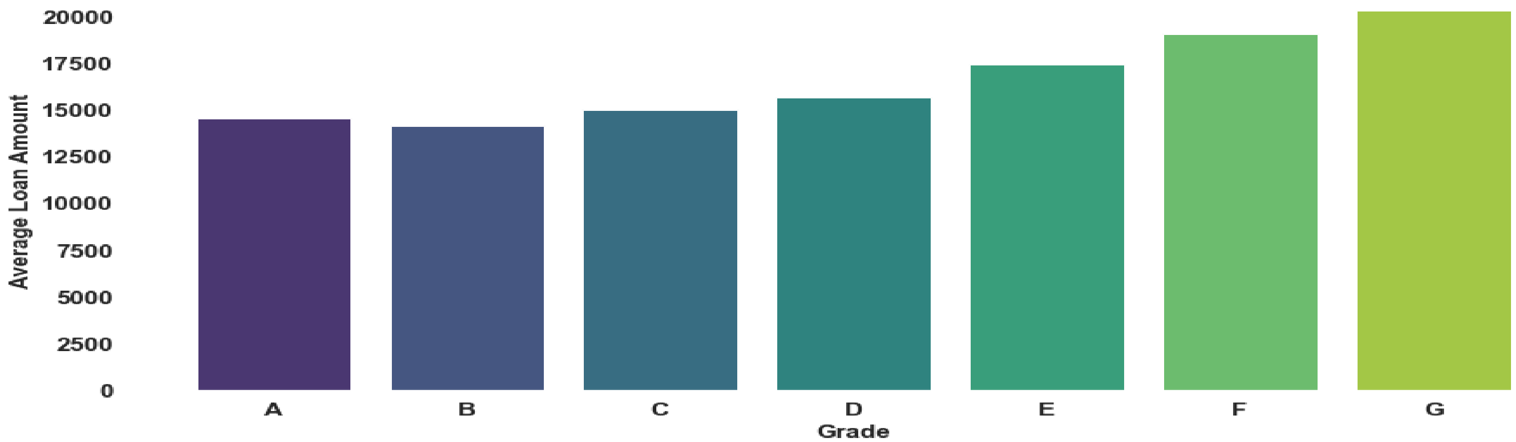
Employment title (Top 10 records)



Type of application



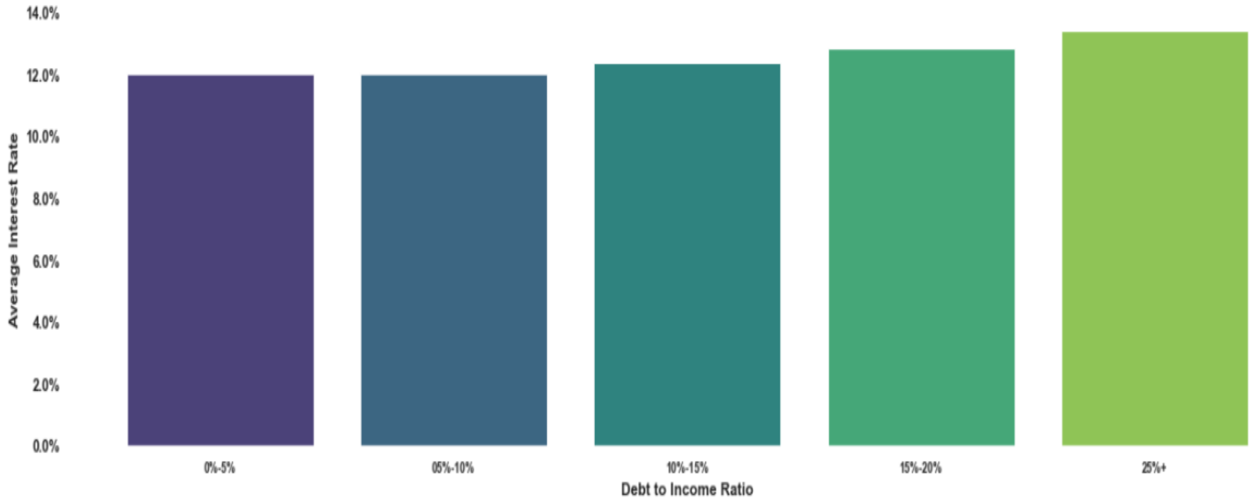
Relationship between grade and loan amount



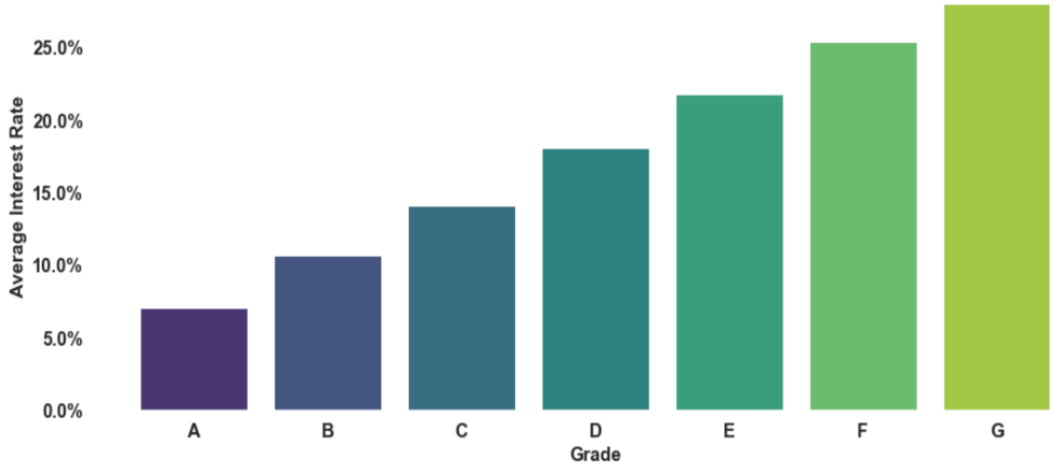
4 Machine Learning application to online credit scoring

Exploratory data analysis

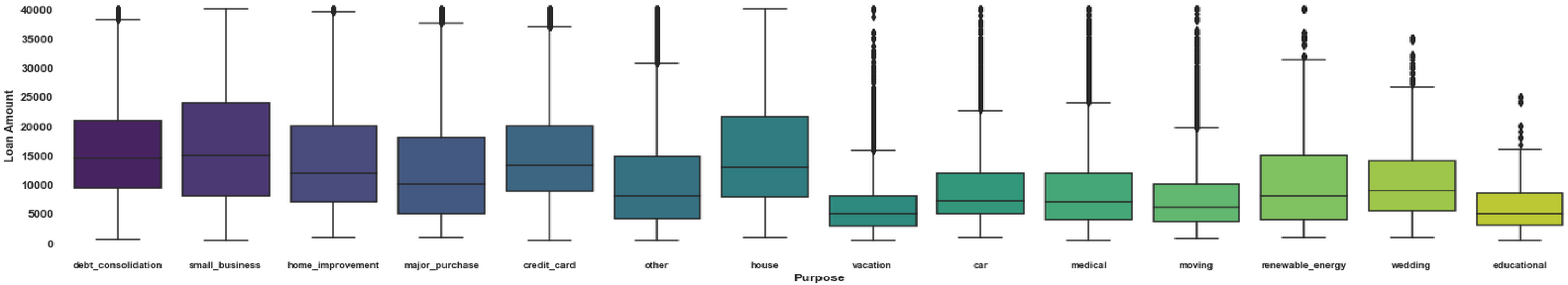
Relationship between interest rate and debt to income ratio



Relationship between interest rate and grade



Relationship between purpose and loan amount



Machine Learning application to online credit scoring

Exploratory data analysis

Selected features

Characteristics	Selected features
Borrower	Annual income Homeownership Verification status Address State Last payment amount.
Credit	Debt-to-Income-Ratio Delinquency status in the last 2 years Number of inquiries in the last 6 months The number of open credit lines in the borrower's credit file The number of current open credit lines Number of derogatory public records Total credit revolving balance Revolving line utilization rate Number of mortgage accounts FICO Score The number of public record bankruptcies.
Loans	Loan amount Loan term Loan interest rate LC assigned loan grade Purpose

4 Machine Learning application to online credit scoring

Preprocessing pipeline

Outlier Treatment

Outliers were appropriately treated to avoid hampering model performance

Categorization

Two types of categorization was performed: Ordinal and One Hot Encoding

Principal Component Analysis

PCA was applied to remove irrelevant information increasing performance

Deal with Missing Values

Five features have missing values that represent less than 10% of the total data. These were median imputed

Proper Scaling

Numerical features were standardized by removing the mean and scaled to unit variance

5 Results and discussion

Single based classifiers

Before and after tunning

Models	Accuracy		Recall		Precision		F1 Score		AUC Score	
	Before Tunning	After Tunning	Before Tunning	After Tunning	Before Tunning	After Tunning	Before Tunning	After Tunning	Before Tunning	After Tunning
Logistic Regression	0.7438	0.7489	0.7676	0.7607	0.4221	0.4276	0.5446	0.5474	0.8325	0.8329
Support Vector Machine	0.7374	0.7454	0.7766	0.7667	0.4156	0.4238	0.5415	0.5459	0.7528	0.7564
K Nearest Neighbors	0.8315	0.8396	0.4639	0.4566	0.6012	0.6372	0.5237	0.5320	0.7845	0.8453
Decision Tree	0.9135	0.9258	0.7757	0.9433	0.7876	0.7492	0.7816	0.8351	0.5768	0.9764
Average	0.8066	0.8149	0.6960	0.7318	0.5566	0.5595	0.5979	0.6151	0.7365	0.8528

5

Results and discussion

Single based classifiers

After Tunning Vs Sampling

Models	Accuracy		Recall		Precision		F1 Score		AUC Score	
	After Tunning	Sampling	After Tunning	Sampling	After Tunning	Sampling	After Tunning	Sampling	After Tunning	Sampling
Logistic Regression	0.7489	0.8387	0.7607	0.8514	0.4276	0.5636	0.5474	0.6783	0.8329	0.9185
Support Vector Machine	0.7454	0.8379	0.7667	0.8584	0.4238	0.5618	0.5459	0.6782	0.7564	0.8450
K Nearest Neighbors	0.8396	0.8607	0.4566	0.8829	0.6372	0.6003	0.5320	0.7167	0.8453	0.9265
Decision Tree	0.9258	0.9262	0.9433	0.9422	0.7492	0.7514	0.8351	0.8361	0.9764	0.9732
Average	0.8149	0.8659	0.7318	0.8837	0.5595	0.6193	0.6151	0.7272	0.8528	0.9162

5

Results and discussion

Single based classifiers with sampling applied

NOVA IMS Information Management School

Single Based Classifiers vs Stacking Approach

Stacking Combination	Models	Accuracy	Recall	Precision	F1 Score	AUC Score
SC1	KNN and SVM (level 0) LR (level 1)	0.8623	0.8790	0.6072	0.7182	0.9335
SC2	DT and SVM (level 0) LR (level 1)	0.9265	0.9415	0.7525	0.8365	0.9726
SC3	DT and KNN (level 0) LR (level 1)	0.9284	0.9404	0.7587	0.8399	0.9739
SC4	DT, KNN and SVM (level 0) LR (level 1)	0.9277	0.9409	0.7564	0.8386	0.9731

6 Final Conclusions

Main Findings

1. Conclusion #1

An ensemble-based approach can contribute to improving the predictive accuracy of traditional classifiers in credit risk scoring.

2. Conclusion #2

Despite the overall better performance of the ensemble model combination, the usage of single based classifiers may be on par or even a better alternative. Often, the best solution may be the simplest one.

3. Conclusion #3

Due to its computational power and complexity, as of right now an ensemble model combination may not be the best application in realistic scenarios. Solving complex problems is an arduous task and if even the best solution is achieved it won't be of much use if it lacks interpretability.

6

Final Conclusions

Recommendations for future work

Recommendation 1

Explore other state-of-the-art machine learning models combinations and different sampling techniques (e.g. SMOTE)

Recommendation 2

A cost sensitive learning approach could serve as an alternative to both undersampling and oversampling

There is always room for improvement, new promising methods, and techniques that may improve results achieved by the proposed solutions are constantly arising permitting continuous learning.

7 References

01 Altman, E. (1968). Financial Ratios, Discriminant Analysis, and the Prediction of Corporate Bankruptcy. The Journal of Finance, 23(4): 589–609.

02 Wolpert, D. (1992). Stacked Generalization. Neural Networks. 5. 241-259.

03 Dietterich, T.G. (2000). Ensemble methods in machine learning. Multiple Classifier Systems: First International Workshop, MCS 2000, Lecture Notes in Computer Science. 1-15.

04 Steel, M. F. J. (2020). Model Averaging and Its Use in Economics. Journal of Economic Literature, 58, 644–719.

05 Ashofteh, A., & Bravo, J. M. (2021). A conservative approach for online credit scoring. Expert Systems with Applications, 176, 1-16.