ANALYZING CREDIT RISK

ASSESSMENT USING PREDICTIVE

ANALYTICS: IMPLICATIONS IN AN

ERA OF ECONOMIC UNCERTAINTY

Brian Thomson
ID: 5012744327

Toronto
Metropolitan
University

Objectives

- In today's financial landscape, effective credit risk assessment is crucial for ensuring the stability of lending institutions.

- My project aims to leverage machine learning for accurate credit risk prediction, contributing to the efficiency of financial decision-making.

Research Questions

- To what extent do income group, loan grade, and credit history length influence credit risk, and how can these factors effectively contribute to a robust credit risk assessment model?

- How do income group, loan grade, and credit history length relate to credit risk, and what insights do these correlations offer for developing effective credit risk management strategies?

- Through decision tree algorithms and other methods, what key patterns and relationships, especially regarding income group and loan grade, can be revealed to enhance our understanding of credit risk and contribute to robust risk management?

Literature Review

- Key findings

   → Existing literature emphasizes the importance of accurate credit risk assessment.

   → Identified gaps led us to explore the applicability of machine learning models.

- Limitations

   → Some studies lacked comprehensive evaluations of various models, motivating our in-depth comparative analysis

Data and Methodology

- Utilized a diverse dataset

- Performed label encoding, one-hot encoding, and standard scaling for data preparation

- Applied Naive Bayes, Decision Tree, Random Forest, K-Nearest Neighbors, Logistic Regression, and Support Vector Machine.

Feature Name	Description
person_age	Age
person_income	Annual Income
person_home_ownership	Home ownership
person_emp_length	Employment length (in years)
loan_intent	Loan intent
loan_grade	Loan grade
loan_amnt	Loan amount
loan_int_rate	Interest rate
loan_status	Loan status (0 is non default 1 is default)
loan_percent_income	Percent income
cb_person_default_on_file	Historical default
cb_preson_cred_hist_length	Credit history length

Data correlation
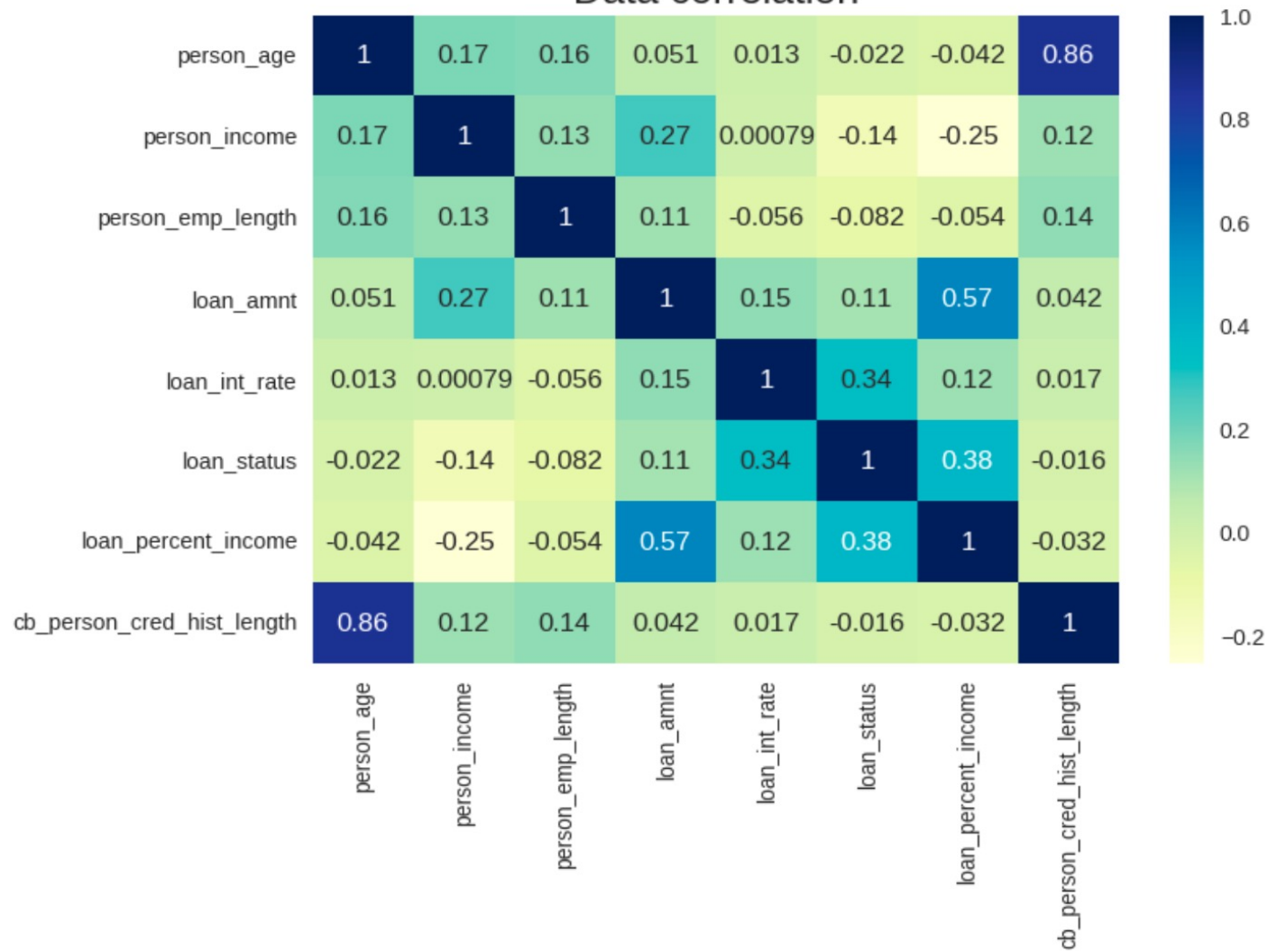
Model Performance

Model	Time to Train	Time to Test	Mean Accuracy	Precision	Recall	F1-Score
GaussianNB	0:00:00	0:00:00	0.8168	0.74	0.25	0.38
DecisionTreeClassifier	0:00:00	0:00:00	0.8891	0.74	0.76	0.75
Random-forest-classifier	0:00:07	0:00:00	0.9321	0.97	0.72	0.82
KNeighborsClassifier	0:00:00	0:00:01	0.8847	0.83	0.61	0.7
LogisticRegression	0:00:00	0:00:00	0.8697	0.75	0.58	0.65
SVM	0:00:20	0:00:03	0.913	0.9	0.66	0.76

→ Random Forest emerges as the top performer with the highest accuracy and F1-Score

Interpretability and Insights

- Examined feature importance in Random Forest for clearer insights into the model's decision-making process and key predictors of credit risk.

- Key Insights

   → Identified income, loan amount, and loan-to-income ratio as crucial factors influencing credit risk.

Limitations, Challenges and Future Research

- The dataset might lack certain features affecting credit risk.

- Interpretability challenges in complex models like Random Forest

- Project Continuity

    → Future research could focus on enhancing interpretability and addressing dataset limitations.

    → Critical insights gained during the study can shape future research in credit risk assessment.

CONCLUSION

→ Random Forest outperforms other models, providing a robust solution for credit risk assessment.

→ Our project contributes by showcasing the effectiveness of machine learning in credit risk prediction.

Thank you for your attention!

Q&A