Toronto Metropolitan University

ANALYZING CREDIT RISK ASSESSMENT USING PREDICTIVE ANALYTICS: IMPLICATIONS IN AN ERA OF ECONOMIC UNCERTAINTY

Brian Thomson (ID: 5012744327)
Supervisor: Ceni Babaoglu

Date of Submission: September 24, 2023

Abstract

In the current financial landscape, characterized by a confluence of economic challenges, including soaring inflation, escalating job losses, elevated interest rates, global uncertainties, and record-high levels of debt, encompassing student loans, mortgages, and line of credit, the topic of credit risk assessment has assumed unprecedented importance in the realm of financial governance, especially in the banking industry. In fact, credit risk assessment plays a pivotal role in financial institutions, influencing lending decisions and overall financial stability. Therefore, this capstone research project seeks to make a substantive contribution to this discourse by delving into the application of predictive analytics, with a specific focus on credit risk assessment.

The dataset selected for this research endeavor, drawn from Kaggle's "Credit Risk Dataset" provides a rich and extensive resource to explore and address the intricate challenges inherent in contemporary credit risk management. To access this dataset in its entirety, please follow this link: https://www.kaggle.com/datasets/laotse/credit-risk-dataset?resource=download

Given the compelling context, three interrelated research questions have been meticulously crafted to delve into the nuances of credit risk assessment:

- **Research question 1:** What are the pivotal determinants of credit risk within the dataset, and to what extent do these factors contribute to the development of a robust credit risk assessment model? An exploration of variables such as age, annual income, home ownership status, employment length, loan intent, loan grade, loan amount, interest rate, historical loan status, and credit history length will be undertaken.

- **Research question 2:** How can we identify key patterns and relationships within the credit risk data that offer insights into the development of effective credit risk management strategies, particularly in the context of the prevailing economic volatility? This inquiry will involve a comprehensive analysis of the dataset to uncover significant correlations, dependencies, and risk factors influencing credit outcomes.
- **Research question 3:** The decision tree algorithm will be used amongst others, how can it be effectively harnessed to uncover pivotal patterns and relationships within credit risk data, offering valuable insights for the development of robust credit risk management strategies, especially within the context of the ongoing financial turbulence? This exploration entails a thorough analysis of the dataset to reveal substantial correlations, dependencies, and influential factors shaping credit outcomes.

This research project will employ an array of predictive analytics techniques, encompassing pattern mining and causality assessment, to thoroughly investigate and address the posed research questions. The methodology will involve meticulous data preprocessing, judicious feature selection, and the development of predictive models that are finely attuned to the multifaceted challenges associated with contemporary credit risk assessment.

In this project, Python will be used along with essential packages like Pandas for data manipulation, Scikit-Learn for machine learning and modeling, NumPy for numerical operations, and Jupyter Notebooks for interactive analysis. This combination equips us to efficiently preprocess, analyze, model, and visualize credit risk data. All of this will be conducted within the overarching backdrop of the complex and volatile economic environment, which further underscores the relevance and urgency of this research in the field of financial governance.

Literature review

Wang, Y., Zhang, Y., Lu, Y., Yu, X. (2020) compared five machine learning classifiers for credit scoring, with the Random Forest classifier consistently outperforming others in precision, recall, AUC, and accuracy. It emphasizes the need for robust risk assessment models in finance, promoting machine learning for enhanced credit assessment, especially in online lending. The Random Forest classifier is highlighted for its proficiency in classification and regression tasks, handling missing and categorical data, and adapting to complex datasets. Overall, it underscores the importance of machine learning, particularly Random Forest, in improving credit scoring precision and efficiency, ideal for the big data era and online lending platforms.

The paper comparing machine learning classifiers for credit scoring, with a specific focus on the Random Forest classifier's superior performance, is highly pertinent to my research questions which offers several valuable insights for the study. Firstly, regarding the Research Question 1, the paper's findings emphasize the importance of precision and accuracy in credit risk assessment, which aligns with our goal of understanding pivotal determinants in the dataset. Secondly, for Research Question 2, the paper reinforces the need for robust risk assessment models, an aspect critical in the context of economic volatility. Thirdly, for Research Question 3, the paper showcases the effectiveness of the Random Forest classifier and its suitability for large-scale datasets, which can guide us in effectively harnessing the decision tree algorithm to uncover patterns and relationships within credit risk data. In essence, the paper provides valuable insights into machine learning's role in improving credit scoring precision and efficiency, making it a relevant resource for our study on credit risk assessment and management.

Xhumari, E., Haloci, S. (2023) had their study in the fintech industry which emphasized the shift to machine learning for more accurate credit scoring. It explores the comparison between regression analysis and machine learning in risk management. The paper's exploration of artificial neural networks (ANN) and convolutional neural networks (CNN) is particularly relevant to our research questions. The study provides insights into using machine learning effectively for credit risk assessment, especially in turbulent financial contexts. It also highlights the importance of maintaining predictor variable quality.

While it emphasizes the transition to machine learning for enhanced accuracy in credit assessments and compares regression analysis to machine learning in the context of risk management, the paper's exploration of artificial neural networks (ANN) and convolutional neural networks (CNN) is particularly relevant to our third research question, where we plan to use the decision tree algorithm. We can draw insights from this paper on how to harness machine learning techniques effectively for credit risk assessment, especially in the context of ongoing financial turbulence, as discussed in our research questions. It also underscores the need for maintaining the quality of predictor variables, which is pertinent to the first research question. This paper provides a valuable reference for improving our study's methodology and understanding the advantages of machine learning in credit risk management.

Crouhy, M., Galai, D., Mark, R. (2000) explored the 1998 capital requirements for market risks established by the Bank for International Settlements (BIS) and their implications for credit risk assessment models. It reviews different credit risk assessment methodologies, including credit migration, option pricing (structural approach), actuarial approach, and CreditPortfolioView, each focusing on various aspects of credit risk. The BIS regulations have led to a need for better internal models for specific risk, which is subject to interpretation by both banks and regulators.

The paper highlights the complexities of disentangling market risk and credit risk components in spread changes and the importance of integrating market and credit risk for a more comprehensive assessment. Various models are examined, and while they are suitable for straightforward bonds and loans, they may not fully address the complexities of derivative products. The future of credit risk models should consider stochastic interest rates and economic conditions. The paper notes that defaults have decreased in recent years during economic growth, impacting credit risk assessment.

The paper on credit risk assessment methodologies, while not directly addressing our specific research questions, offers valuable insights into the complexities of credit risk modeling and the importance of integrating various factors and market conditions. It highlights the need for a comprehensive approach that considers both market risk and credit risk, which could enhance the robustness of credit risk assessment models. Understanding how different models handle credit risk components and the challenges they face can inform our exploration of pivotal determinants (Research Question 1) and the identification of key patterns and relationships (Research Question 2) within our dataset. This knowledge may also guide our analysis of the decision tree algorithm's effectiveness (Research Question 3) in addressing credit risk within the context of ongoing financial turbulence.

Goyal, S. (2018) investigated the application of neural network algorithms in predicting credit default and assessing the creditworthiness of loan applicants. It primarily focuses on a small dataset of residential mortgages to develop a binary classifier to identify borrowers likely to default. The study employs a feed-forward neural network and backpropagation algorithm to train and validate models, comparing them to a linear regression model for accuracy. The results show that both neural network and linear regression models are highly effective, achieving around 97.68% accuracy, with similar mean square errors. While neural networks provide efficient credit default prediction, they are more challenging to interpret compared to linear regression models. This research demonstrates the effectiveness of artificial neural networks in predicting credit risk, suggesting their broad applications beyond residential mortgages, including bond ratings, currency ratings, and more.

In fact, this research is highly relevant to our research questions. It demonstrates the application of advanced machine learning techniques, such as neural networks, to assess and predict credit risk, which aligns with our aim to identify pivotal determinants of credit risk within the dataset. By comparing the neural network's performance with other methods like linear regression, it offers insights into the effectiveness of these algorithms in understanding credit risk factors (pertaining to the first research question). Moreover, this paper highlights the importance of data attributes and normalization in improving model performance, which could be valuable when exploring key patterns and relationships within credit risk data (related to the second question). The study's use of different algorithms, including decision trees, provides insights into their effectiveness, contributing to the exploration of robust credit risk management strategies (related to the third question). This paper can inform our research by showcasing the advantages and considerations of applying machine learning techniques to credit risk assessment and management.

Khemakhem, S., Boujelbènea, Y. (2015) focused on assessing credit risk using artificial neural networks (ANNs) as an alternative to traditional credit risk models, particularly discriminant analysis. The study examines financial ratios of 86 Tunisian companies over a specific period and concludes that ANNs offer more accurate predictability compared to discriminant analysis in terms of credit risk assessment. The research aims to improve decision support for bankers, highlighting the potential of ANNs in the context of credit risk prediction. To enhance our study, we can draw insights from the paper's findings, particularly the superiority of ANNs over discriminant analysis in credit risk prediction. It also emphasizes the need for considering a broader range of variables, both quantitative and qualitative, when assessing credit risk, which can inform our exploration of pivotal determinants of credit risk within our dataset (related to our first research question). Additionally, the paper's discussion of extending traditional models with techniques like genetic algorithms and large margin separators may offer valuable insights into enhancing credit risk assessment methods (related to our third research question).

It highlights the superiority of ANNs in credit risk prediction compared to traditional discriminant analysis, which is a relevant insight for our research questions. Specifically, for the first research question, it underlines the importance of considering advanced modeling techniques like ANNs to determine the pivotal determinants of credit risk. The paper's focus on assessing financial ratios of companies and comparing prediction accuracy offers a valuable reference for understanding and identifying pivotal factors. Additionally, the idea of improving credit risk assessment models aligns with the second and third research questions, where we aim to uncover key patterns, relationships, and influential factors within the credit risk data and employ decision tree algorithms for this purpose. We also can learn from this paper about the potential of advanced modeling techniques like ANNs and their applicability in the context of credit risk assessment, providing insights to enhance the study's methodology and results.

Zhou, J., Wang, C., Ren, F., Chen, G. (2021) introduced a comprehensive scheme for assessing online consumer credit risk, which has relevance to our study's research questions. It addresses the challenge of consumer risk profiling in the context of online consumer credit services, which aligns with our goal of understanding the determinants and patterns of credit risk. The paper's approach of augmenting consumer profiles with phone usage information to overcome the "thin file" challenge offers insights for enhancing credit risk assessment models. Additionally, its exploration of multi-staged consumer repayment timing and the impact on profits relates to our research questions about uncovering patterns and relationships in credit risk data and employing decision tree algorithms. We can learn from this paper's methodology and findings to improve our own study, particularly in terms of data augmentation, predictive modeling, and the multi-stage analysis of credit risk.

It offers insights into the development of credit risk assessment models, similar to our first question, by demonstrating the importance of augmenting consumer profiles with additional data sources. It also addresses the second question by emphasizing the need to understand multi-stage repayment behaviors and their impact on profits, which is analogous to uncovering key patterns and relationships in credit risk data. Moreover, this paper provides valuable insights into how to harness machine learning methods effectively, which aligns with our third question about using decision tree algorithms. We can learn from its approach to data augmentation, predictive modeling, and analysis of multi-stage credit risk to enhance our study's methodology and insights.

Most recently, Markov, A., Seleznyova, Z., Lapshin, V. (2022) provided a systematic review of credit scoring research, particularly focused on recent developments from 2016 to 2021. It highlights the significance of credit risk assessment for financial institutions and the impact of precise risk estimation on an organization's profitability, pricing, and even marketing strategies. The paper emphasizes the ongoing relevance of credit scoring and the need for a comprehensive understanding of best practices in this field. It touches upon various aspects of credit scoring, such as feature engineering, dataset considerations, imbalance issues, data preprocessing, model testing, and the use of both baseline models like logistic regression and more complex ensemble models. The review offers recommendations for researchers, including using multiple datasets, addressing data imbalances, and conducting more vigilant model testing. It also identifies the growing role of ensemble models and provides insights into the impact of COVID-19 on credit scoring research, which, as of June 2021, appears to have had limited direct influence. Researchers seeking to stay updated on recent credit scoring trends and best practices will find this paper informative and can use it as a reference for future research directions.

While it offers insights into the pivotal determinants of credit risk by discussing aspects such as feature engineering, data preprocessing, and model selection, this information can be valuable in enhancing our understanding of the determinants of credit risk within the dataset. Furthermore, the paper provides an overview of dataset considerations, model testing, and the role of ensemble models, which can inform our approach to identifying key patterns and relationships within credit risk data. As we plan to use the decision tree algorithm, this paper can serve as a reference for effectively harnessing decision trees to uncover pivotal patterns and relationships within credit risk data. It underscores the importance of thorough model testing and performance evaluation, aligning with our research goals of developing robust credit risk assessment models within the context of economic volatility.

Dataset

GitHub link: https://github.com/brianthomsoncad/TMU-Capstone-project.git

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

The dataset contains information related to individuals' credit and loan profiles, providing a comprehensive view of their financial backgrounds and loan-related characteristics. It encompasses various attributes, each contributing to the evaluation of an individual's creditworthiness and loan-related risk factors. These attributes include "person_age," which represents the age of the individual, "person_income" denoting their annual income, and "person_home_ownership" indicating whether they own their home or not. The dataset also captures details about employment, including "person_emp_length," the length of their employment in years. In the context of loan applications, it covers "loan_intent" and "loan_grade," reflecting the intention and grade associated with the loan request. Moreover, it includes "loan_amnt" as the loan amount, "loan_int_rate" as the interest rate, and "loan_status" to classify loans as non-default (0) or default (1). The "loan_percent_income" attribute calculates the percentage of an individual's income relative to the loan amount. Additionally, the dataset incorporates historical credit information through "cb_person_default_on_file" and "cb_preson_cred_hist_length" assessing whether a person has previously defaulted on loans and their credit history length, respectively. This dataset serves as a valuable resource for credit risk assessment, offering insights into the key factors and characteristics influencing loan approval and default prediction.

index	person_age	person_income	person_emp_length	loan_amnt	loan_int_rate	loan_status	loan_percent_income	cb_person_cred_hist_length
count	32581.0	32581.0	31686.0	32581.0	29465.0	32581.0	32581.0	32581.0
mean	27.73	66074.85	4.79	9589.37	11.01	0.23	0.17	5.80
std	6.35	61983.12	4.14	6322.09	3.24	0.41	0.11	4.06
min	20.0	4000.0	0.0	500.0	5.42	0.0	0.0	2.0
25%	23.0	38500.0	2.0	5000.0	7.9	0.0	0.09	3.0
50%	26.0	55000.0	4.0	8000.0	10.99	0.0	0.15	4.0
75%	30.0	79200.0	7.0	12200.0	13.47	0.0	0.23	8.0
max	144.0	6000000.0	123.0	35000.0	23.22	1.0	0.83	30.0

The statistic table provided contains summary statistics for several numerical variables:

- person_age: The average age is around 27.73 years, with a standard deviation of approximately 6.35. The age ranges from a minimum of 20 to a maximum of 144 years which is pretty rare.
- person_income: The average income is roughly $66,074.85, with a standard deviation of about $61,983.12. The income varies from a minimum of $4,000 to a maximum of $6,000,000.
- person_emp_length: The average employment length is approximately 4.79 years, with a standard deviation of about 4.14. The range is from a minimum of 0 to a maximum of 123 years which does not make sense.
- loan_amnt: The average loan amount is approximately $9,589.37, with a standard deviation of around $6,322.09. The loan amounts range from a minimum of $500 to a maximum of $35,000.
- loan_int_rate: The average loan interest rate is about 11.01%, with a standard deviation of roughly 3.24%. Rates vary from a minimum of 5.42% to a maximum of 23.22%.
- loan_status: This appears to be a binary variable with a mean of 0.218, indicating the proportion of "1" values in the dataset.
- loan_percent_income: On average, loans represent approximately 17% of a person's income, with a minimum of 0% and a maximum of 83%.
- cb_preson_cred_hist_length: The average credit history length is about 5.80 years, with a standard deviation of approximately 4.06. The length ranges from a minimum of 2 years to a maximum of 30 years.

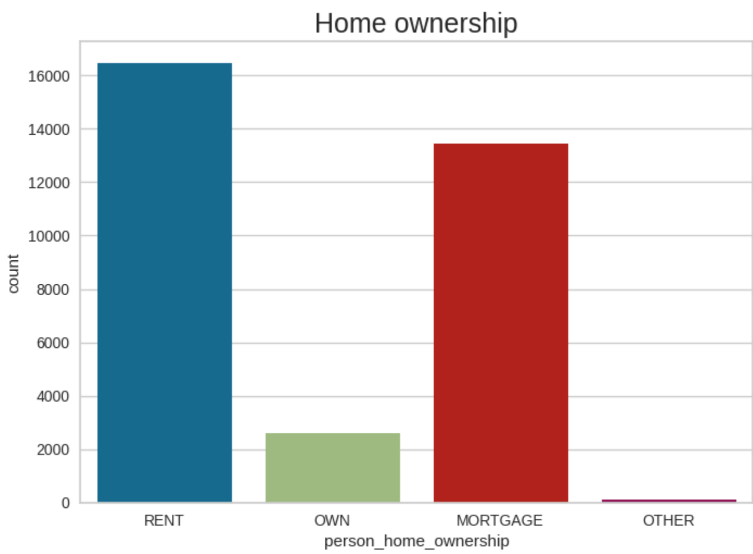
To handle those abnormal data in terms of age and length of employment, we need to replace those values with the max values.

```
df['person_age'].max()

#Assuming individuals with age > 90 to be errors
df = df.loc[df['person_age'] < 90]
df['person_emp_length'].max()

#Employment cannot be greater than the individual's age (accounting for childhood)
df = df.loc[df['person_emp_length'] < df['person_age'] - 10]
```

Plotting some numeric data



Descriptive dataset information

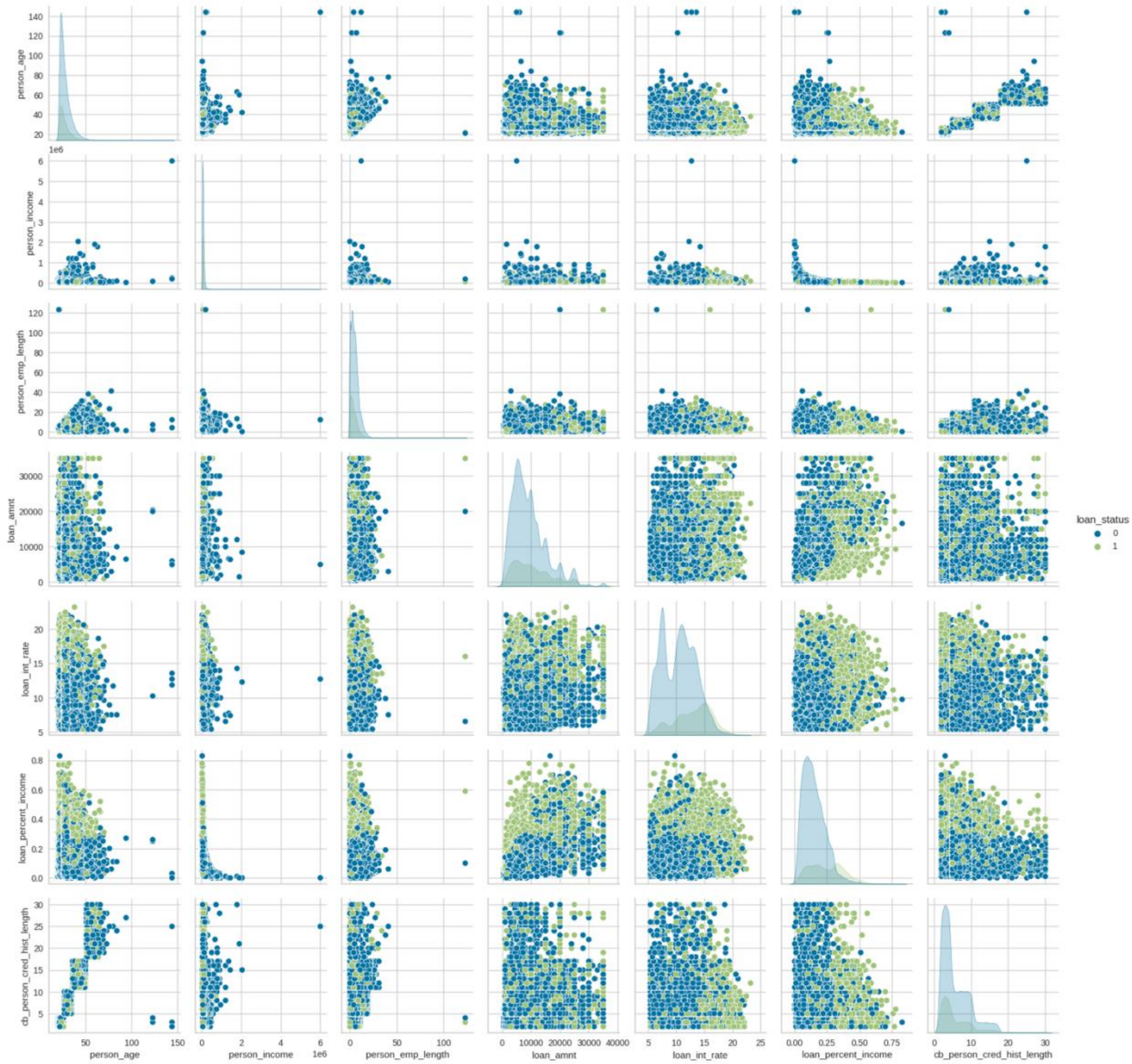
```
RangeIndex: 32581 entries, 0 to 32580
Data columns (total 12 columns):
 #   Column                      Non-Null Count  Dtype
---  ------                      --------------  -----
 0   person_age                  32581 non-null  int64
 1   person_income               32581 non-null  int64
 2   person_home_ownership       32581 non-null  object
 3   person_emp_length           31686 non-null  float64
 4   loan_intent                 32581 non-null  object
 5   loan_grade                  32581 non-null  object
 6   loan_amnt                   32581 non-null  int64
 7   loan_int_rate               29465 non-null  float64
 8   loan_status                 32581 non-null  int64
 9   loan_percent_income         32581 non-null  float64
 10  cb_person_default_on_file   32581 non-null  object
 11  cb_person_cred_hist_length  32581 non-null  int64
dtypes: float64(3), int64(5), object(4)
```

In this dataset, there are 32,581 entries or rows, with index values ranging from 0 to 32,580. It's essentially the row numbers or identifiers for the data points. Data columns (total 12 columns): This indicates that the dataset has a total of 12 columns or features. Each column represents a different attribute or variable in our dataset.

- person_age: This is a column representing the age of individuals. It contains 32,581 non-null (non-missing) values and is of data type int64 (integer).
- person_income: This column contains information about the income of individuals. It also has 32,581 non-null values and is of data type int64.
- person_home_ownership: This is a categorical column representing the type of home ownership. It is of data type object and also has 32,581 non-null values.

- person_emp_length: This column represents the length of employment for individuals. It contains 31,686 non-null values and is of data type float64 (floating-point numbers).
- loan_intent: This column indicates the intent or purpose of the loan. It's a categorical variable of data type object.
- loan_grade: This is another categorical column representing the grade of the loan. It's also of data type object.
- loan_amnt: This column contains information about the loan amount. It's of data type int64 and has 32,581 non-null values.
- loan_int_rate: This column represents the interest rate on the loan. It's a floating-point variable of data type float64 and has 29,465 non-null values.
- loan_status: This column indicates the loan status and is represented as integers. It contains 32,581 non-null values.
- loan_percent_income: This column represents the percentage of income the loan amount represents. It's of data type float64 and contains 32,581 non-null values.
- cb_person_default_on_file: This is a categorical column representing whether a person has a default on their credit record. It's of data type object.
- cb_person_cred_hist_length: This column indicates the length of the credit history of individuals. It's of data type int64 and contains 32,581 non-null values.

The pairplot was used to explore relationships between multiple numeric variables in our dataset, which can provide insights into relationships, patterns, and correlations in the data. The 'loan_status' column was used to color the data points on the scatterplots. By setting the 'hue' parameter, we can visually distinguish different classes of data points which used 'loan_status' to represents different loan statuses.



To examine missing values, the command missing_values = df.isna().sum() was used.

```
person_age                    0
person_income                 0
person_home_ownership         0
person_emp_length           895
loan_intent                   0
loan_grade                    0
loan_amnt                     0
loan_int_rate              3116
loan_status                   0
loan_percent_income           0
cb_person_default_on_file     0
cb_person_cred_hist_length    0
dtype: int64
```

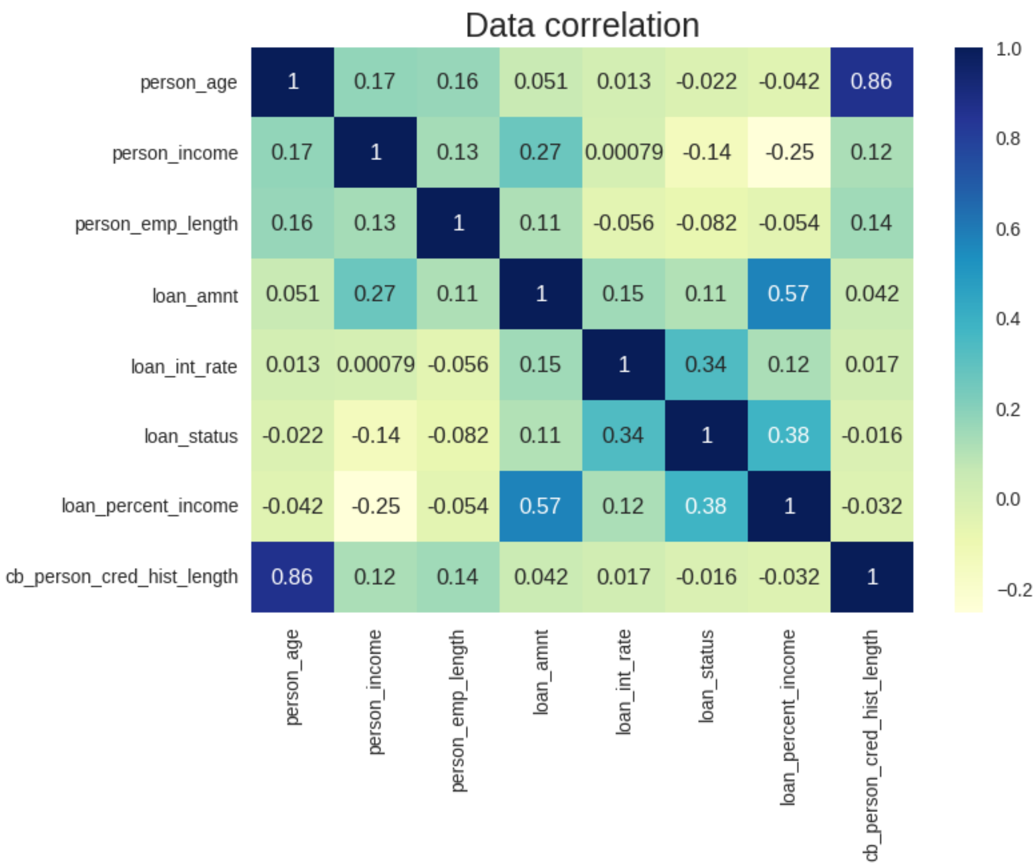
The table displays the count of missing values in each column of the dataset. It shows that the "person_emp_length" column has 895 missing values, while the "loan_int_rate" column has 3,116 missing values. All other columns, including "person_age," "person_income," "person_home_ownership," "loan_intent," "loan_grade," "loan_amnt," "loan_status," "loan_percent_income," and "cb_person_default_on_file," have no missing values, indicating that these columns contain complete data. Missing value assessment is essential for data quality assurance and informs subsequent data handling processes.

To handle missing values by replacing those value with the median of that features' values

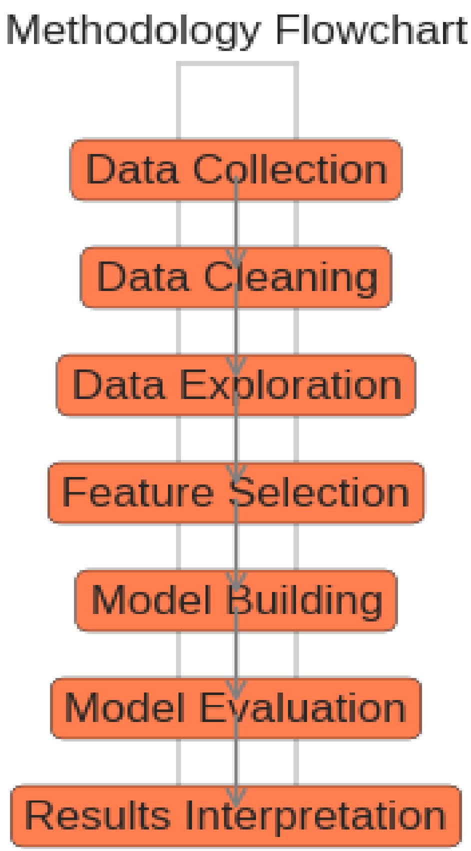
```
df.isnull().sum()
#Filling missing values with mean:
df.loc[df['loan_int_rate'].isnull(), 'loan_int_rate'] = df['loan_int_rate'].median()
df.loc[df['person_emp_length'].isnull(), 'person_emp_length'] = df['person_emp_length'].median()
df.isnull().sum()
```

```
person_age                    0
person_income                 0
person_home_ownership         0
person_emp_length             0
loan_intent                   0
loan_grade                    0
loan_amnt                     0
loan_int_rate                 0
loan_status                   0
loan_percent_income           0
cb_person_default_on_file     0
cb_person_cred_hist_length    0
dtype: int64
```

The correlation heatmap illustrates the relationships between various financial and personal attributes in the dataset. The values range from -1 to 1, with 1 indicating a perfect positive correlation and -1 indicating a perfect negative correlation. Here, we observe several notable correlations. Firstly, there is a strong positive correlation of approximately 0.86 between "person_age" and "cb_person_cred_hist_length," suggesting that older individuals tend to have longer credit histories. Secondly, "person_income" and "loan_amnt" exhibit a positive correlation of around 0.27, indicating that as income increases, loan amounts tend to be higher. However, it's noteworthy that "loan_status" has a negative correlation of about -0.14 with "person_income," indicating that individuals with lower incomes might have a higher likelihood of loan default. Additionally, "loan_int_rate" and "loan_status" have a positive correlation of approximately 0.34, suggesting that loans with higher interest rates might be associated with a higher likelihood of loan default. Overall, the heatmap helps identify potential relationships and dependencies among the variables in the dataset, which can be valuable for data analysis and modeling.



Approach



Data Collection:

The dataset was retrieved from Kaggle which can be access from https://www.kaggle.com/datasets/laotse/credit-risk-dataset?resource=download

Data Exploration:

The dataset was analyzed to gain an understanding of its contents by employing the summary statistics (mean, median, standard deviation), using histograms, scatter plots, and box plots, to visualize the data distribution, relationships, and potential outliers. We also checked for missing values, abnormal data and determining the distribution of variables.

Data Preparation:

Data Preparation is crucial in ensuring that the data is ready for analysis. Missing values were handled by using its median values.

Modeling Techniques:

In this phase, a variety of modeling techniques will be employed including random forest, decision tree, regression. After that, the choice of methods and algorithms will be adjusted according to the predictive analytics and machine learning which are relevant to credit risk assessment.

Validation and Evaluation:

The study will use evaluation metrics including accuracy, precision, recall, F1-score to evaluate models from the previous step. To do that, the dataset will be split into training and testing sets to ensure robust model evaluation.

Results Interpretation:

The output of our predictive models and the insights will be presented in relevant to our research questions to compare our findings back to the original context, which is the credit risk assessment in the current economic landscape. Also, the significance of our results in the context of financial governance and the broader economic environment will be highlighted.

References

Crouhy, M., Galai, D., Mark, R. (2000). A comparative analysis of current credit risk models. Journal of Banking & Finance. Volume 24, Issues 1–2, January 2000, Pages 59-117 (Link: https://www.sciencedirect.com/science/article/abs/pii/S0378426699000539)

Goyal, S. (2018). Credit Risk Prediction Using Artificial Neural Network Algorithm (Link: https://www.datasciencecentral.com/credit-risk-prediction-using-artificial-neural-network-algorithm/)

Khemakhem, S., Boujelbènea, Y. (2015). Credit risk prediction: A comparative study between discriminant analysis and the neural network approach. Accounting and Management Information Systems. Vol. 14, No. 1, pp. 60-78, 2015 (Link: https://www.researchgate.net/profile/Sihem-Khemakhem/publication/323560727_Credit_risk_prediction_A_comparative_study_between_discriminant_analysis_and_the_neural_network_approach/links/5a9d8419aca272cd09c2195c/Credit-risk-prediction-A-comparative-study-between-discriminant-analysis-and-the-neural-network-approach.pdf)

Markov, A., Seleznyova, Z., Lapshin, V. (2022) Credit Scoring Methods: Latest Trends and Points to Consider. The Journal of Finance and Data Science, Volume 8, Pages 180-201 (Link: https://www.sciencedirect.com/science/article/pii/S2405918822000095)

Xhumari, E., Haloci, S. (2023). A comparative study of Credit Scoring and Risk Management Techniques in Fintech: Machine Learning vs. Regression Analysis. CEUR Workshop Proceedings RTA-CSIT 2023, April 26–27, 2023 Tirana, Albania (Link: https://ceur-ws.org/Vol-3402/paper02.pdf)

Wang, Y., Zhang, Y., Lu, Y., Yu, X. (2020). A Comparative Assessment of Credit Risk Model Based on Machine Learning - a Case Study of bank Loan Data. 2019 International Conference on Identification, Information and Knowledge in the Internet of Things. Procedia Computer Science Volume 174, 2020, Pages 141-149 (Link: https://www.sciencedirect.com/science/article/pii/S1877050920315830)

Zhou, J., Wang, C., Ren, F., Chen, G. (2021). Inferring Multi-stage Risk for Online Consumer Credit Services: An Integrated Scheme Using Data Augmentation and Model Enhancement. Decision Support Systems, Volume 149, ID: 113611 (Link: https://www.sciencedirect.com/science/article/abs/pii/S0167923621001214)