Loan Eligibility Prediction: An Analysis of Feature Relationships and Regional Variations in Urban, Rural, and Semi-Urban Settings

Zhile Zhang *

Department of Mathematics, University of Manchester, England, UK

* Corresponding Author Email: zhile.zhang@student.manchester.ac.uk

Abstract. In the complex realm of banking and financial systems, the process of extending loans involves navigating through multifaceted criteria that reflect an individual's creditworthiness. As modern global economies evolve, they present distinct financial behaviors across urban, rural, and semi-urban geographies. This dynamic landscape prompts an essential inquiry: Are the determinants for loan approval consistent irrespective of these varying settings? Our research, set against the backdrop of this question, harnesses the power of machine learning to delve deep into this enigma.Utilizing a comprehensive dataset acquired from Kaggle[1], we embark on a rigorous journey of data preprocessing. This phase involves meticulous efforts like rectifying null values and sophisticated encoding techniques to prepare the data for machine learning models. Subsequent stages see the employment of multiple algorithms, including decision trees, random forests, logistic regression, XGBoost, among others, each rigorously trained and critically evaluated for their predictive capabilities. Of these, the Logistic Regression algorithm stood out, boasting an impressive accuracy of 83.78%.However, the essence of this research transcends mere algorithmic performance. The heart of our findings elucidates the subtle yet significant variations in factors dictating loan approvals across urban, rural, and semi-urban datasets. This nuanced understanding underscores a vital recommendation for financial institutions: the imperative to customize and refine their loan eligibility parameters in harmony with these regional intricacies, ensuring a more holistic and informed lending approach.

Keywords: Loan Eligibility Prediction, Loan Approval Determinants, Machine Learning.

1. Introduction

Banks serves the basic necessities of everyone. Loans and interest banks accrue are the heartbeat of the banking sector, as one of the main sources of income. While at the same time, people increasingly rely on the loan offered by banks and financial institutions for various needs such as a car or a house purchase. Banks offer different types of loans for different categories of applicant based on their income and other factors and applicants need to pay back the loan amount on installment basis along with interest in time. When customer applies for a loan, banks faces defaulters on repayment of the loan and it is significant for banks to validates the customer eligibility for loans. Predictions on loan eligibility helps banks make decisions and offer loan to those applicants with high chance and ability to repay, based on the proper studies of applicant status as well as completed data on applicants' information. However, inaccurate predictions on loan eligibility can cause severe problems like bankruptcy, as statistics reveals that number of banks are closed due to the huge loss caused by the failure in loan recovery. "Bad loans reduce banks' profitability and limit their ability to issue new credit. Large volumes of bad loans can cause banks problems with their capital adequacy and, at worst, can lead to default. Bad loans can lead to financial crisis." (Fredriksson,2019) [2]

While assessing loan applicants' ability to repay has always been a vital issue to banking sectors, traditional loan eligibility predictions methods which has been used for centuries, are now considered as outdated and are replaced or complemented by machine learning. This is because a variety of limitations have shown by the traditional assessment methods. They are found difficult in handling complex data patterns and making accurate predictions, as they are often simplistic and overlook crucial factors, leading to errors. Credit scoring model is a typical example of the traditional

assessment methods. In which it assigns a credit score based on the credit history, which is then used to determine loan eligibility. Although the credit history of an applicant is indeed a crucial factor influencing whether the loan is approved or not, it ignores other relevant factors like income status, leading to an incomplete picture of a borrower's creditworthiness.

Recent studies, such as those by Sharma. (2020) [3] and Gonzalez (2021) [4], underscore a growing consensus in the financial community about the limitations of traditional credit assessment methodologies. They advocate for a more evolved, data-driven approach, especially in an age burgeoning with information. Herein lies the promise of Machine Learning (ML). This computational method, which thrives on harnessing vast datasets, offers a more holistic lens to ascertain loan eligibility, discerning nuanced patterns and multifaceted relationships often overlooked previously. “The existing research demonstrates the potential of machine learning in loan eligibility prediction, offering more accurate data-driven decision-making in the lending industry.”(Kumar,2022)[5][6] This essay embarks on a comprehensive exploration of ML's prowess in loan eligibility prediction, meticulously analyzing the myriad determinants influencing loan approvals. Moreover, in recognizing the dynamic financial landscapes of urban, rural, and semi-urban regions, it offers insights into how regional nuances can shape creditworthiness assessments in the modern era.

2. Methodology

2.1. The Idea of the Entire Article

Central to this research is the utilization of data analysis methodologies to discern any pronounced variations in the relationships between loan approval factors characteristics across urban, rural, and semi-urban regions. A pivotal question posed here is whether identical feature analysis and filtering criteria are universally operable across these diverse geographical classifications. Through deploying machine learning models, we aim to verify if the data features discerned are transferable and effective when applied to datasets from cities, rural zones, and semi-urban areas.

2.2. Summary of Principles of Data Analysis Methods

To set the foundation, our primary task is data acquisition, sourcing it from reputable databases ensuring its integrity and relevance. This is followed by a thorough data preprocessing phase, a quintessential step to guarantee that subsequent analyses rest on a clean and well-structured dataset. Data cleaning, in this phase, addresses discrepancies, anomalies, and missing values, either by imputation or elimination based on the nature and significance of the data. Feature engineering, another pivotal step, involves the meticulous identification and crafting of variables. It's here that the core of our inquiry - understanding the regional distinctions in loan features - comes to the forefront. Every feature is scrutinized with a lens specific to its manifestation in urban, rural, and semi-urban settings.

With the dataset in place, Exploratory Data Analysis (EDA) is undertaken. Beyond the standard practices of deriving descriptive statistics and employing visualization tools, the EDA, in this study, places heightened emphasis on recognizing regional patterns or anomalies in loan attributes. Whether it's through scatter plots comparing loan amounts across regions or histograms detailing credit histories of applicants from different settings, the EDA is tailored to our geographic-centric inquiry.

2.3. Method Principle of Machine Learning Model

In our pursuit of understanding loan feature characteristics across distinct regions, various machine learning models present as viable solutions due to their distinct methodologies. Decision Trees, for instance, graphically represent possible solutions to a decision, differentiated by criteria such as Gini impurity, which measures the disorder in a set, and Entropy that gauges the randomness in information being processed. The objective here is to minimize these values to achieve efficient splits in the data.

Complementing Decision Trees are Random Forests, which, as ensembles of decision trees, take a democratic approach, voting for each prediction. This plurality often results in a more accurate representation than a singular tree might offer. The key to Random Forest's power is its introduction of randomness during model creation, thus ensuring diversified decision paths and a reduced risk of overfitting.

Logistic Regression, tailored for binary classification, thrives in quantifying relationships between categorical variables. By leveraging its inherent logistic function, it estimates probabilities, making it particularly suitable for our loan eligibility problem.

Further deepening our approach, the eXtreme Gradient Boosting or XGBoost combines predictions from multiple decision trees. Constructed sequentially, each tree in the series corrects the errors of its predecessor, enhancing accuracy. Similarly, AdaBoost, short for 'Adaptive Boosting', refines its predictions by adjusting weights for both classifiers and the data points. This adaptability ensures misclassified instances in previous iterations gain priority, improving overall model robustness.

Another intriguing method, K-Nearest Neighbors (KNN), is non-parametric. KNN classifies a data point based on its neighbors' classifications, ensuring context is always at the forefront of prediction. Moreover, Support Vector Classification (SVC), a pivotal subset of Support Vector Machines (SVMs), emerges as a force for binary classification. Its genius lies in determining the best hyperplane to segregate a dataset into distinct classes.

Collectively, these models promise to shed invaluable insights into the complex behaviors of loan features across cities, rural areas, and suburbs. The choice of model will be determined by the data's nature and the intricacies of the relationships we're aiming to unravel. Thus, with each model's unique approach, we're equipped to delve deeper, aiming for a thorough understanding, ensuring the methodology's robustness and relevance to our central research question. In deploying these models, we've ensured methodical training and validation phases. Each model is subjected to rigorous training, followed by hyperparameter tuning. This ensures that each model is optimally equipped to capture and reflect the regional distinctions in loan features. Evaluations are not mere formalities; they're exhaustive checks on the models' capabilities, leveraging metrics such as accuracy, precision, recall, and F1-score.

To encapsulate, our methodology is a harmonious blend of traditional data analysis techniques and modern machine learning approaches. All of it converges towards our goal: understanding the intricacies of loan feature characteristics across distinct geographical terrains and ensuring their accurate prediction.

3. Experiments

3.1. Data Acquisition and Preprocessing

In the complex realm of predicting loan approval, the bedrock is the dataset. Sourced from the comprehensive Kaggle platform, our dataset boasts a structure of 614x13, translating to 614 samples intricately described by 13 features. Each of these features, ranging from demographic details to credit history, provides a kaleidoscope view of the parameters influencing loan decisions. We can collect the key information from the dataset columns, seen as table 1, which enables us to understand the features that influence the loan approval on a basic level.

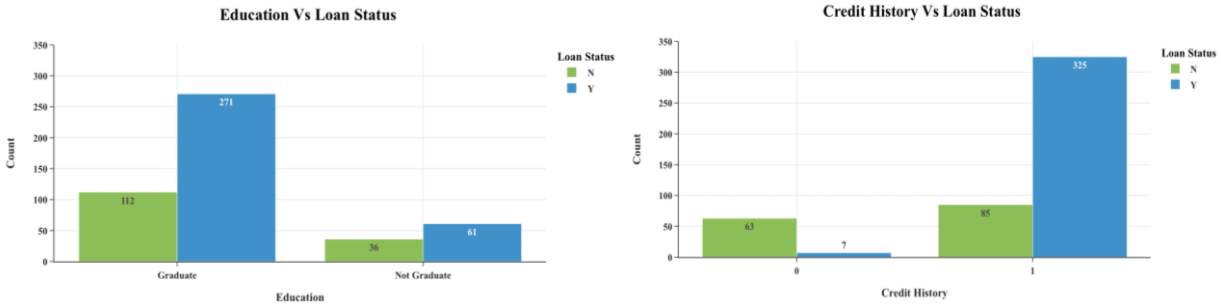
Table 1. Table of Dataset Key Information

Elements in columns of dataset	Loan ID	Married	Dependents	Education	Self_employed	Applicant Income	Loan Amount	Loan Amount_term	Credit History	Property area	Loan Status
Information collected and explanation	Unique Loan ID	Applicant marriage status (Yes/No)	Number of Dependents	Applicant Education Status (Graduate/ Under gratuate)	Self_employed (Yes/No)	Income of Applicant (Numerous)	Amount of loan in thousands	Time period of a loan in months	Applicant vredit history meets guidelines	Urban/Semi-Urban/Rural	Loan Approval (Yes/No)

Transitioning from acquisition to preprocessing, Python remains our tool of choice. Its versatility, combined with the power of libraries such as Pandas and NumPy, facilitates seamless data cleaning. This includes tasks like imputation of missing values, normalization to bring every feature to a comparable scale, and transformation to enhance data quality. Our dataset is further accentuated with regional divisions—urban, semi-urban, and rural—which add layers of granularity to our study. Visual aids like table 1 amplify our understanding, enabling us to discern patterns and anomalies in the data.

3.2. Data Analysis

At the heart of our project lies the intricate dance of data analysis. The primary objective is discerning how various characteristics sway the pendulum of loan approval. Before delving into more advanced analyses, an essential preliminary step involved using various graphs and data visualization tools during our Exploratory Data Analysis (EDA) phase. Tools like heatmaps were employed to discern foundational patterns within the data and to unearth relationships amongst features. This allowed us to grasp the overarching structure and potential intricacies of our dataset, setting the stage for subsequent, more detailed analyses. However, at this juncture, we limit our discussion to merely introducing the EDA process, reserving the outcomes and detailed conclusions for the subsequent sections. Graphical representations, shown as (a)(b) in figure 1, are not just images but are narratives. For instance, these visual aids elucidate that the parameter of educational attainment holds negligible influence over loan approval status, evidenced by the consistent numerical relationship pertaining to loan approvals irrespective of the education status. In stark contrast, the presence of a commendable credit history emerges as a significant determinant in loan sanctioning. This evaluative process was systematically reiterated to crystallize the core features for our model. Through such scrupulous feature selection, we judiciously excised variables of limited predictive value, such as gender and education status, thereby sharpening our focus on truly influential factors.



(a) (b)

Figure1. Loan Approval Factors. (a)Figure of Relationship Between Education and Loan Status; (b)Figure of Relationship Between Credit History and Loan Approval

A foundational aspect of our study is understanding the relationship between various influencing factors and the loan approval status. To achieve this, we adopt a ratio-based analysis. As an illustration, consider the credit history of an applicant. For those with a commendable credit history, we calculate a ratio: the number of loans approved to the number of loans not approved. A parallel ratio is derived for applicants with an unfavorable credit history.

Such ratios are immensely insightful. They offer a numeric testament to the predictive power of each factor. For instance, a stark difference in the ratios for those with good versus bad credit history could underscore the significant influence of credit history on loan approval decisions.

But our analysis doesn't halt at overarching insights. We delve deeper by dissecting these ratios across various geographical spectrums: urban, rural, and semi-urban datasets. The objective is to uncover regional nuances. Are the influential factors and their predictive power consistent across cities, villages, and suburbs? Or do we witness variances, suggesting regional specificities in loan approval determinants? By comparing these ratios, we aim to achieve two outcomes. First, to discern

the importance and predictive potency of each factor. And second, to fine-tune our feature selection, ensuring that our model's training data is both robust and regionally relevant.

3.3. Model Selection and Training

Transitioning from data analysis, we ventured into the intricate realm of model selection and training. This step holds paramount importance, as it determines the efficacy of our predictive algorithms.

Model Selection: Initially, a plethora of machine learning models were considered. From ensemble methods like Random Forest and AdaBoost to gradient boosting techniques like XGBoost, and not to forget the traditional ones like Logistic Regression, K-Nearest Neighbors (KNN), and Support Vector Classifier (SVC). The decision tree models were further subcategorized by their criterion, namely Gini impurity and entropy, to gauge which yields better predictive performance.

Data Splitting: To train and evaluate our models, the datasets for cities, rural areas, and suburbs were each split into training and testing sets. Typically, we earmarked about 70% for training and reserved 30% for testing, ensuring that the data distribution remains consistent. The training Process involves firstly feature scaling: Given the diverse nature of our features, ranging from credit scores to income levels, it was imperative to scale them. This ensures all features have equal influence on the model. We leveraged techniques like standardization and normalization for this purpose. Secondly hyperparameter Tuning: Prior to full-blown training, hyperparameters for each model were fine-tuned. Utilizing GridSearchCV and RandomizedSearchCV, optimal parameter combinations were identified to maximize model performance.[8]

After that is the model training, with hyperparameters in place, each model was trained on the training dataset. Python's Scikit-Learn library facilitated this, offering a seamless interface to train various models. Sebsequent steps would be cross-validation, as we need to further prevent overfitting and to ensure the robustness of our models, we employed K-Fold cross-validation. This provides a more holistic assessment of a model's performance, using different portions of the data for training and validation.

Eventually, Performance Metrics Selection, which is crucial not only to train a model but also to understand how well it's performing. Hence, metrics like accuracy, F1 score, precision, and recall were chosen to evaluate and compare model effectiveness.

In the culmination of this rigorous process, Logistic Regression emerged as the model best suited for our research objectives. As shown in the following tables (Table 2 and 3), showing the outcome of data training using different models, different performance metrics are selected for comparison as well. Its ability to quantify the relationship between our selected features and loan approval status, combined with its innate aptitude for binary classification, made it the most fitting choice for our dataset's nature and the research question at hand.

Table 2. Table of Model Metrics Evaluation

	Logistic Regression			KNNs			Naive-Bayes			SVM			Random Forests		
	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score	Precision	Recall	f1-score
N	0.47	0.89	0.62	0.00	0.00	0.00	0.43	0.85	0.57	0.00	0.00	0.00	0.41	0.78	0.54
Y	0.98	0.83	0.9	1.00	0.72	0.84	0.97	0.82	0.89	1.00	0.72	0.84	0.96	0.81	0.88
weighted average	0.9	0.84	0.86	1.00	0.72	0.84	0.89	0.82	0.84	1.00	0.72	0.84	0.88	0.81	0.83

Table 3. Table of Model Accuracy

Model	Logistic Regression	XGBoost	Decision Tree-Gini	Decision Tree-Entropy	Naive-Bayes	Adaboost-Entropy	Adaboost-Gini	Random Forest	Knns	SVC
Accuracy	83.78%	82.71%	82.70%	82.70%	82.16%	81.63%	81.62%	78.37%	72.44%	72.43%

However, it's paramount to approach these results with a modicum of circumspection. While the logistic regression model excelled, it also presented certain limitations. Notably, its accuracy exhibited a pronounced sensitivity to the quality of the dataset. “In scenarios where data might be riddled with anomalies, inconsistencies, or biases, the performance of logistic regression could potentially be compromised. Hence, ensuring the integrity and robustness of the dataset becomes imperative to harness the full predictive prowess of the model.” (Kaur, 2022) [7]

3.4. Model Verification

The pinnacle of our experiment is verification. With the crown bestowed upon logistic regression, the subsequent step is its rigorous validation. This involves a meticulous experiment across the trio of regional datasets: urban, semi-urban, and rural. The core aim is to unearth any disparities or consistencies in the model's performance across these regions. It's a multistage process—parameter initialization, region-specific training, and performance evaluation using a myriad of metrics. This structured approach ensures that the findings of our study remain robust and relevant across varied geographical settings.

4. Test Results

4.1. Data Analysis Results

Our exploration into the data's depths is evident in the various analytical tools we've harnessed. A noteworthy highlight is Figure 2 the heatmap which evolved from our initial Exploratory Data Analysis (EDA). This tool unfolds the pairwise correlations amongst our features in vivid detail. Each cell offers insight into the relationship between two variables, with the color gradients painting a story of strength and directionality. The intense shades, whether dark or bright, are telltale signs of strong correlations, be they positive or negative. On the contrary, softer hues communicate weaker relationships.

However, this heatmap goes beyond mere visual appeal. It's an integral facet of our investigative methodology. Its role in identifying correlations is paramount, especially when addressing complexities like multicollinearity in regression-based analyses. Such insights aren't just academic; they empower us to refine our models, ensuring that the predictors are rigorous. Moreover, these relationships hint at underlying causal narratives, enriching the depth and breadth of our analyses. Hence, Figure 2 is not just a visual representation; it's a lighthouse illuminating the intricate relationships embedded within our data, sculpting our further inquiries.

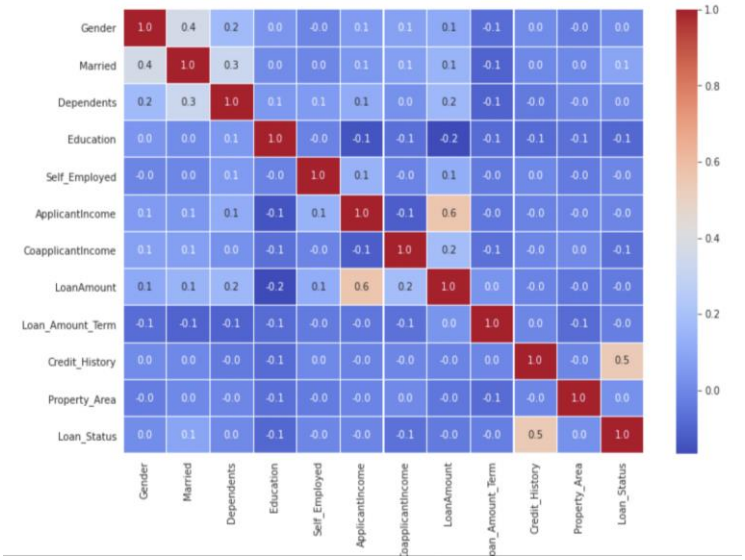


Figure 2. Heatmap of the factors

Our ratio analysis presents a fascinating exploration of factors like Gender, Education, Self Employed status, and Credit History in the context of Loan Status. These factors were meticulously analyzed across urban, rural, and semi-urban backdrops, revealing insights about regional nuances. Preliminary observations underscored that both Graduates and Non-Graduates witnessed near-similar patterns in loan approvals and rejections. Parallels were also observed for Gender and Self-Employed individuals. Surprisingly, these features appeared to have minimal bearing on the loan's status. In stark contrast, Credit History emerged as a game-changer. A higher frequency of loan approvals was noted amongst applicants boasting a favorable credit history, highlighting its significant influence.

To further our understanding, we ventured beyond ratios, employing tools such as boxplots and bar charts (figure 3 and 4). These tools divulged more regional patterns and showcased the impact of property areas on loan approvals. An insightful takeaway was the consistent pattern of loan approvals and rejections across the three property areas. This consistency underscores that the property area, in isolation, might not be a deciding factor for loan approvals.

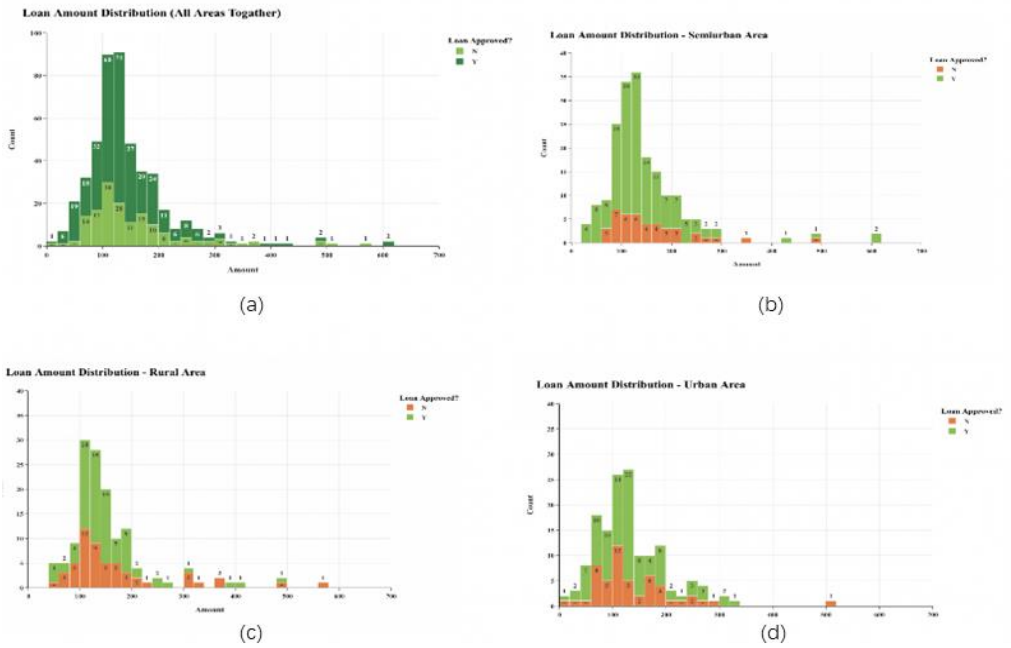


Figure 3. Bar chart of loan amount distribution in different property area. (a) Loan amount distribution in All areas; (b) in Semi-urban area;(c) in Rural area; (d) in urban area

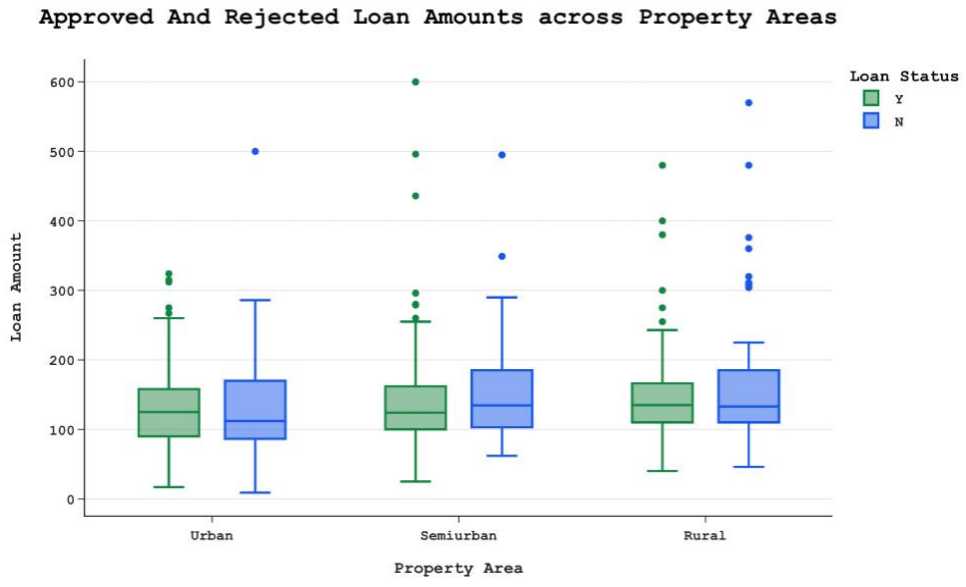


Figure 4. Box-plot diagram of Approved and Rejected loan amounts across property areas.

Lastly, our scatter plot (figure 5) accentuated how regional dynamics could influence the loan amount vis-a-vis the loan status. A salient observation was the proportional increase in approved loan amounts with the applicant's total income. In urban and semi-urban regions, this relationship appeared linear. However, in rural regions, the increase in approved loan amounts vis-à-vis total income was less predictable. A recurrent trend across all regions was the preponderance of approved loan amounts staying below 200.

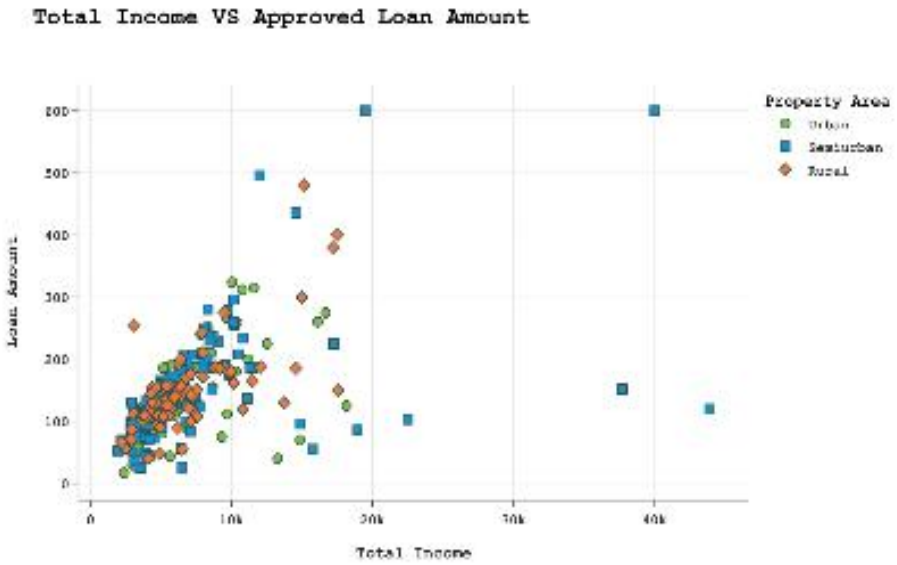


Figure 5. Scatter of Relationship of Total income and Loan Approval in different Property Area

This cumulative analysis, spanning various tools and methodologies, arms us with a holistic understanding of our dataset, paving the way for robust machine learning predictions in the subsequent phases.

The cogency of these key features across different geographies underlines their universality. It's an enticing proposition, as this commonality ensures the potential for machine learning models to make consistent and accurate predictions across regions. The foundation of a robust machine learning model, especially in the realm of loan approval prediction, is predicated heavily on the veracity and relevance of the data it is trained on. Therefore, by ensuring that our key features' extraction and representation remains consistent across rural, urban, and semi-urban sectors, we're not only ensuring the integrity of our model but also streamlining the loan prediction process.

4.2. Verification Results

Transitioning to the verification phase, the trained logistic regression model was rigorously tested. With the training data already primed, the test set, a melange of diverse instances, provided a comprehensive landscape to evaluate the model's real-world applicability. The logistic regression training achieved an accuracy of 83.76%, 83.78%, 84% for training data from rural, city and semi-urban area, showing a high accuracy, similar and applicability of machine learning model on loan eligibility prediction among different regions.

4.3. Result Analysis

Sifting through layers of data, training numerous machine learning models, and distilling critical insights, our expedition through the vast terrains of loan approval predictions has been illuminating. The patterns unearthed and the relationships deciphered provide a solid foundation upon which financial institutions can base their decisions. However, every silver lining has its cloud. Our methodology, while robust, does come with certain limitations for example, the scope of our data might not encompass all possible socio-economic variables affecting loan approvals.

Nonetheless, based on our current knowledge and the experiments conducted, it's discernible that key features, especially those consistent across regions, play a pivotal role in influencing loan

approvals. [9,10] Ensuring the accuracy and relevance of these features across datasets and geographies not only augments the practicality and precision of machine learning models but also ushers in a new paradigm for financial institutions, where data-driven decision-making becomes the gold standard.

5. Conclusion

Our comprehensive exploration into the intricate dynamics of loan approvals, underpinned by rigorous data analyses and machine learning, has offered profound insights into the factors that most influence loan decisions across diverse geographies. Leveraging tools such as heatmaps, we've unearthed pivotal relationships between variables, while our ratio analyses have demystified the significance of certain features across urban, semi-urban, and rural backdrops. The subsequent validation via logistic regression underscored the potential of data-driven approaches in revolutionizing the loan approval landscape.

The universality of certain features across regions not only promises a streamlined and consistent predictive model but also paves the way for enhanced accuracy in loan approval mechanisms, regardless of geographical variations. It's evident that the future of loan approvals will be significantly anchored in data, with machine learning models serving as indispensable tools in the arsenal of financial institutions.

Yet, as we peer into the horizon, it's crucial to acknowledge that this journey, though rewarding, is still in its nascent stages. While we've made substantial strides, the evolving nature of socio-economic dynamics mandates continuous refinement of our models and techniques. [11,12] The intersection of finance and technology offers a realm ripe with potential, and as we tread this path, the promise is not just of efficiency but of creating more transparent, fair, and data-informed financial ecosystems.

References

[1] Vikasukani, Ukase. “Loan Eligibility Prediction - Machine Learning.” Kaggle, 27 Dec. 2020, www.kaggle.com/code/vikasukani/loan-eligibility-prediction-machine-learning#5.-Traing-the-ML-Model.

[2] Fredriksson, Olle, and Niklas Frykström. “No 2 2019 Economic Commentary - Sveriges Riksbank.” Financial Stability Department of the Riksbank., www.riksbank.se/globalassets/media/rapporter/ekonomiska-kommentarer/engelska/2019/bad-loans-and-their-effects-on-banks-and-financial-stability.pdf. Accessed 12 Aug. 2023.

[3] Ioannis Antonopoulos, Valentin Robu, Benoit Couraud, Desen Kirli, Sonam Norbu, Aristides Kiprakis, David Flynn, Sergio Elizondo-Gonzalez, Steve Wattam,Artificial intelligence and machine learning approaches to energy demand-side response: A systematic review,Renewable and Sustainable Energy Reviews,Volume 130,2020,109899,ISSN 1364-0321,https://doi.org/10.1016/j.rser.2020.109899.

[4] Leo, M.; Sharma, S.; Maddulety, K. Machine Learning in Banking Risk Management: A Literature Review. Risks 2019, 7, 29. https://doi.org/10.3390/risks7010029

[5] C. Naveen Kumar, D. Keerthana, M. Kavitha and M. Kalyani, "Customer Loan Eligibility Prediction using Machine Learning Algorithms in Banking Sector," 2022 7th International Conference on Communication and Electronics Systems (ICCES), Coimbatore, India, 2022, pp. 1007-1012, doi: 10.1109/ICCES54183.2022.9835725.

[6] Meenaakumari, P. Jayasuriya, N. Dhanraj, S. Sharma, G. Manoharan and M. Tiwari, "Loan Eligibility Prediction using Machine Learning based on Personal Information," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 1383-1387, doi: 10.1109/IC3I56241.2022.10073318.

[7] P. Kaur, G. Panwar, N. Uppal, P. Singh, B. D. Shivahare and M. Diwakar, "A Review on Multi-Focus Image Fusion Techniques in Surveillance Applications for Image Quality Enhancement," 2022 5th International Conference on Contemporary Computing and Informatics (IC3I), Uttar Pradesh, India, 2022, pp. 7-11, doi: 10.1109/IC3I56241.2022.10073085.

[8] Guerra, P.; Castelli, M. Machine Learning Applied to Banking Supervision a Literature Review. Risks 2021, 9, 136. https://doi.org/10.3390/risks9070136

[9] Huang, J., Chai, J. & Cho, S. Deep learning in finance and banking: A literature review and classification. Front. Bus. Res. China 14, 13 (2020). https://doi.org/10.1186/s11782-020-00082-6

[10] Apley, D. W. (2016). Visualizing the effects of predictor variables in black box supervised learning models. arXiv preprint arXiv:1612.08468.

[11] Bergstra, J., Bardenet, R., Bengio, Y. & Kégl, B. (2011). Algorithms for hyper-parameter optimization. NIPS'11: Proceedings of the 24th International Conference on Neural Information Processing Systems. Granada, Spain.

[12] Binder, A., Montavon, G., Lapuschkin, S., Müller, K.-R. & Samek, W. (2016). Layer-wise relevance propagation for neural networks with local renormalization layers. International Conference on Artificial Neural Networks, (pp. 63-71).