ENHANCING CREDIT APPROVAL PROCESSES THROUGH ADVANCED PREDICTIVE MODELING: A COMPARATIVE ANALYSIS

Group 4: Sreeja Macha, Colin Chen, Jessica Yang

2024-3-23

TABLE OF CONTENTS

1. Executive Summary

Business Problem

Financial institutions often face the challenge of credit risk assessment – a critical factor in their stability and profitability. As consultants specializing in data analytics, our primary objective was to enhance the accuracy and efficiency of the credit approval processes within the financial sector. By leveraging predictive analytics, we aimed to develop a robust model which is capable of accurately assessing an applicant's creditworthiness, thus streamlining the decision-making process and minimizing the rate of defaults.

Data Analysis and Methodology

We performed a comprehensive analysis using two datasets that included a range of applicant attributes, such as demographics, financial history, and credit records. We constructed nine distinct models on this data using advanced machine learning techniques such as RandomForest, XGBoost, and LightGBM algorithms. Our preliminary analysis of the data suggested that the dataset was heavily imbalanced where instances of good credit significantly outnumber those of bad credit. We employed three oversampling techniques: SMOTE, ADASYN, and Borderline-SMOTE, to achieve a balanced dataset for training.

Key Findings

- Precision in Prediction: Our models exhibited high accuracy levels in predicting credit approval, with precision-recall metrics providing essential insights into each model's ability to effectively distinguish between creditworthy and non-creditworthy applicants.
- Challenges: Our models struggled with overfitting and imbalanced data.
- Importance of Features: Our analysis highlighted that features such as employment stability, income, demographics, education, and asset ownership play pivotal roles in

predicting creditworthiness across various models and resampling techniques. While specific feature importance may slightly vary, the consistent emphasis on these factors aligns with traditional credit risk assessment.

- Impact of Resampling Techniques: The comparison between models trained on the newly balanced datasets revealed significant implications for model performance, particularly in predicting the category of bad credit.

Potential Impact on Credit Approval Decisions

Implementing these findings could significantly change the credit approval process. The financial institutions can reduce default rates by accurately identifying risky applicants. Conversely, they can ensure that creditworthy individuals are not unjustly denied. This inturn helps boost their income and attract more customers.

2. Background, Context, and Domain Knowledge

Business Context

In the competitive landscape of the financial sector, credit approval processes stand as a cornerstone of operational success. These processes are pivotal for institutions offering a variety of products and services, including loans, credit cards, and mortgages. The ability to accurately assess the credit risk of applicants not only mitigates potential defaults but also enables institutions to tailor their offerings to meet diverse customer needs effectively. In this context, the industry sees a constant flux of competition, with entities striving to innovate and leverage data to gain an edge in both risk management and customer service.

Given this scenario, our project's primary focus was to enhance the credit approval framework for a client within this sector. Our client, like many in the industry, faces the dual challenge of minimizing credit risk while maximizing potential revenue through customer acquisition and retention. This requires a nuanced understanding of creditworthiness of applicants, beyond what traditional assessment methods can provide.

Evolution from Traditional to Analytical Methods

This assessment has historically placed a great deal of reliance on credit scores, such as the FICO score in the United States, which turns a person's credit history into a number. But these ratings could not fully reflect an applicant's financial stability, particularly if they have little or no credit history. Financial firms have also depended on manual review procedures, in which loan personnel assess applications based on broader criteria, in addition to credit scores. However, because of individual bias, it takes a lot of time and might lead to inconsistent results. Logistic models later gained popularity as a credit scoring technique. They have provided a more thorough examination of an applicant's financial situation by taking into account a number of variables, including income, outstanding debts, and credit history.

3. Traditional Problem Solving in the Industry

Harnessing Advanced Analytics for Enhanced Credit Risk Assessment

The shortcomings of the traditional techniques and logistic regression have led the industry to explore more sophisticated predictive methods, such as machine learning algorithms like Random Forest, XGBoost, and LightGBM. These techniques address the complex dynamics of credit risk in the current market and outperform the standard models in terms of accuracy. But they also present issues with explaining decisions, which is important for upholding transparency with customers and regulatory compliance.

Alignment with Business Model

This shift towards advanced analytics reflects the industry's response to changing consumer preferences and market conditions. Financial institutions hope to increase customer satisfaction and loyalty by incorporating advanced prediction algorithms, which will also help them assess risk better. This strategic alignment demonstrates the industry's dedication to using technology to gain a competitive edge and sustain growth.

Our approach of incorporating cutting-edge machine learning algorithms, such as XGBoost, Random Forest, and LightGBM, improves productivity, mitigates risk, and increases customer happiness.

- Efficiency: Automating the credit approval process reduces time and cost. This allows for more applications to be processed without extra resources.
- Risk Management: These algorithms improve the accuracy of identifying applicants likely to default, protecting the institution's finances.
- Customer Satisfaction: Faster and more accurate credit approvals lead to a better customer experience, increasing loyalty and potentially attracting more customers.

4.1 Data Exploration

For our project, we used two datasets - application and credit records. They are sourced from Kaggle. These include a comprehensive range of applicant attributes including demographics, financial history and credit history.

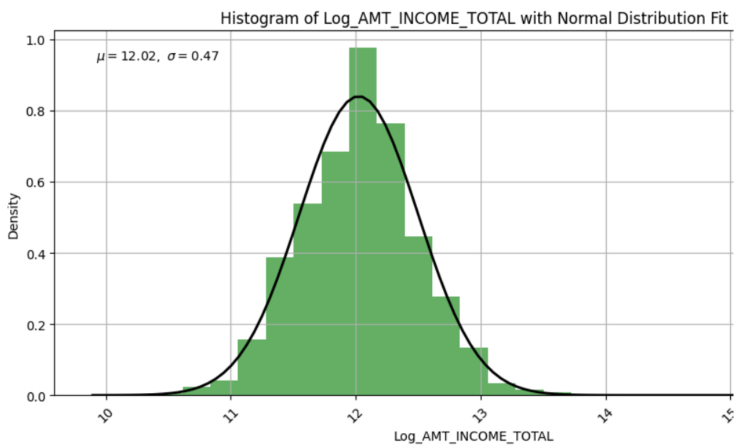
Our exploratory phase revealed key insights into applicant profiles, which could influence their creditworthiness:

Demographic Insights:

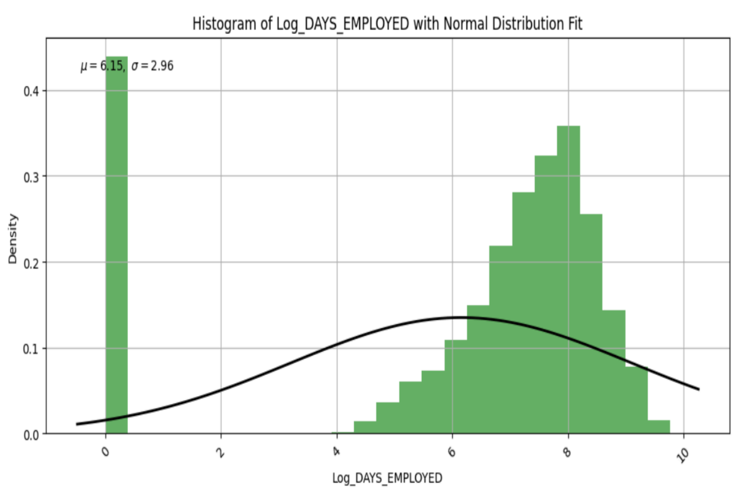
- Gender, car ownership, and real estate status are distributed as one might expect in a general population, with no extreme imbalances.
- Most applicants are employed, with a smaller proportion being commercial associates, pensioners, state servants, and students.
- A significant majority have completed secondary education, with fewer achieving higher education levels or having an academic degree.

- Married applicants constitute the majority, with other family statuses represented to a lesser extent.
- Most applicants live in a house or apartment, which could be indicative of financial stability.

Financial Capacity and Stability:

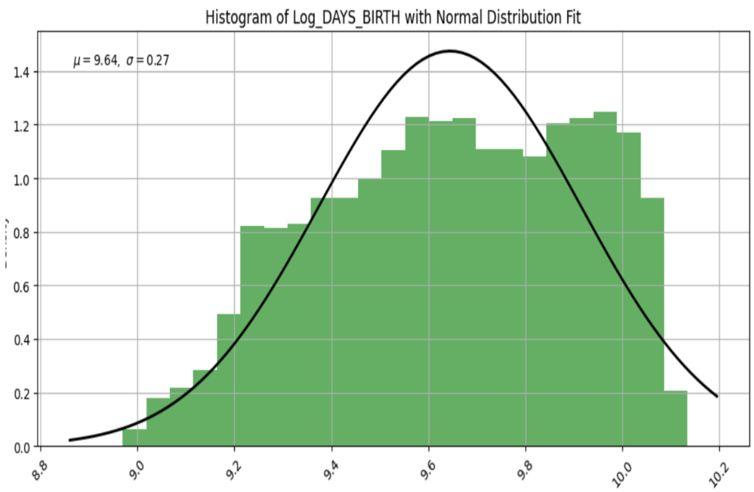


Income levels are log-normally distributed. The majority of applicants have income levels around the mean. However, there are some high end earners.



The log(days_employed) tells us that most applicants have been in their jobs for a decent amount of time. This is a good sign of financial stability and their ability to repay loans. However, there are also some who haven't been working for as long or haven't worked at all.

Risk Profile Indications:



The ages of the applicants are mostly on the younger to middle-aged side.This is a positive sign as applicants will be working for a while longer. This might mean there's less risk involved with them paying back loans over time.

4.2 Data Preparation

For the credit records dataset, we added two new columns:

WINDOW: The duration in months from the first recorded month to the current record month.

MONTHS_ON_BOOK: The # of active months on the record for each account.

Purpose of Window: In credit risk assessment, the 'WINDOW' refers to the timeframe we examine to keep an eye on a client's financial behavior. We look at things like if they pay their bills on time, how much they use credit, and if their credit score changes.

To determine this timeframe, we computed the cumulative % of bad customers based on different past due periods (STATUS).

Low Percentage of Bad Debt: The purple line that's close to the bottom of the graph indicates the percentage of loans categorized as bad debt (>150 days past due). This is very low relative to the total loan portfolio.

Stable After Initial Period: Initially, there is an increase in the cumulative percentage as the months on books increase. But this line flattens out after about 15 months. This means that most of the loans that become bad debts do so around this period, and after a certain point the likelihood of loans becoming bad debt stabilizes and doesn't increase much.

Comparative Performance: Compared to loans that are past due more than 30, 60, 90, and 120 days, the loans that are past due more than 150 days represent a much smaller fraction.

This indicates that most of the loans do not reach the threshold of >150 days past due (bad debt or write-offs).

To conclude: A 20-months observe window covers most bad customers. So, this will be our observation window for further analysis.

2. Binary Encoding of STATUS: Encoded categories [0, 1, 2, 3, 4, C, X] as '1' for good debts, and category [5] to '0' for bad debts or write-offs.

3. Filtering: Filtered for credit records with WINDOW > 20.

4. Merging Data: Dropped irrelevant columns related to contact methods from application records and MONTHS_BALANCE and WINDOW columns from client records and then merged both the datasets via ClientID

5. Missing Values: We dropped rows with missing values in the "OCCUPATION_TYPE" column (190,510 records approx.), to improve the quality of our data.

6. One-hot Encoding: Other categorical variables in both the datasets were one-hot encoded.

7. Renaming Columns: Renamed the columns for better readability.

8. Oversampling Techniques: To address the imbalance between good and bad debts to accurately identify creditworthy and non-creditworthy applicants, we employed the following three oversampling techniques:

- SMOTE (Synthetic Minority Over-sampling Technique)
- ADASYN (Adaptive Synthetic Sampling)
- Borderline-SMOTE

4.3 Comparative Analysis of Credit Approval Prediction Models

Objective: To evaluate and compare the performance of nine different predictive models on credit approval data while addressing class imbalance using three sampling techniques.

Model Summaries:

RF: Random Forest; XGB: XGBoost; LGB: LightGBM; Class 0: Bad Debt; Class 1: Good Debt

Model	Sampling Technique	Accuracy	Precision (Class 0)	Recall (Class 0)	FPR (Class 0)	FNR (Class 0)	MSE	R^2	AUC-PR (Class 0)
RF 1	SMOTE	0.9985	0.66	0.68	0.0008	0.32	0.0015	0.32	0.73
XGB 1	SMOTE	0.9981	0.55	0.89	0.0017	0.11	0.0019	0.15	0.78
LGB 1	SMOTE	0.9985	0.62	0.84	0.0012	0.16	0.0015	0.32	0.78
RF 2	ADASYN	0.9985	0.67	0.68	0.0008	0.32	0.0015	0.34	0.65
XGB 2	ADASYN	0.9981	0.54	0.90	0.0018	0.10	0.0019	0.15	0.65
LGB 2	ADASYN	0.9982	0.58	0.82	0.0014	0.17	0.0018	0.22	0.67
RF 3	Borderline-SMOTE	0.9985	0.67	0.69	0.0008	0.31	0.0015	0.35	0.66
XGB 3	Borderline-SMOTE	0.9979	0.52	0.88	0.0019	0.12	0.0021	0.07	0.66
LGB 3	Borderline-SMOTE	0.9982	0.58	0.79	0.0013	0.21	0.0018	0.21	0.66

Feature Importances:

Model	Sampling Technique	1st Important Feature	2nd Important Feature	3rd Important Feature	4th Important Feature	5th Important Feature
RF 1	SMOTE	Employment Start (0.162)	Days of Birth (0.149)	Months on Book (0.128)	Annual Income (0.080)	Family Size (0.038)
XGB 1	SMOTE	Incomplete Higher Education (0.105)	Commercial Associate (0.081)	Private Service Staff (0.073)	Core Staff (0.052)	Drivers (0.048)
RF 2	ADASYN	Days of Birth (0.151)	Employment Start (0.143)	Months on Book (0.115)	Annual Income (0.084)	Family Size (0.040)
XGB 2	ADASYN	Working (0.115)	Car Yes (0.066)	With Parents (0.059)	Pensioner (0.057)	Family Size (0.051)
RF 3	Borderline-SMOTE	Employment Start (0.160)	Days of Birth (0.141)	Months on Book (0.126)	Annual Income (0.079)	Family Size (0.041)
XGB 3	Borderline-SMOTE	Working (0.109)	Gender Male (0.057)	Family Size (0.057)	Property Yes (0.049)	Municipal Apartment (0.044)

Comparison of Statistical Performance:

Accuracy

1. RF with SMOTE: Highest among RF models, but the model overfits due to imbalanced data.
2. XGBoost & LightGBM with SMOTE: Similar high accuracy as the RF model. Also raises concerns about overfitting.
3. RF & XGBoost with ADASYN: Marginally lower than SMOTE, meaning that they are sensitive to the oversampling method used.
4. LightGBM with ADASYN: Comparable to SMOTE models, robust to oversampling method.
5. RF & LightGBM with Borderline-SMOTE: Lower than SMOTE and ADASYN. Could potentially indicate better at handling borderline cases.
6. XGBoost with Borderline-SMOTE: Lowest accuracy. Could potentially be due to conservative classification of the minority class.

Precision-Recall (AUC-PR)

1. Class 0 (Bad Debt): The AUC-PR varies a lot, but XGBoost models mostly have higher values. They strike a better balance between precision and recall for this class.
2. Class 1 (Good Debt): Consistently perfect AUC-PR score across all models. This however masks performance on the minority class which is the ‘bad debt’ one.

Mean Squared Error (MSE)

- RF Models: Exhibited lower MSE compared to the boosting models.
- Boosting Models (XGBoost & LightGBM): Slightly higher MSE, which might be because the models are sensitive to the noise created by oversampling techniques.

R-squared (R^2)

R^2 values were not significantly different across models. All models explain the variance in the dataset to a similar extent.

Error Rates (False Positive Rate (FPR) & False Positive Rate (FNR))

- RF Models: Strike a balance between FPR and FNR. This indicates a balanced error type trade-off.
- XGBoost Models: Lower FPR but higher FNR. Indicates a conservative approach to predicting the 'bad debt' class.
- LightGBM Models: Showed a similar trend to RF but with slightly higher FNR. This could be critical depending on the cost of false negatives.

Overall Performance

- Random Forest Models are more balanced but do not offer the best precision for the 'bad debt' class.
- XGBoost Models showed potential for better 'bad debt' prediction but at the risk of missing out on the actual 'bad debt' cases (higher FNR).
- LightGBM Models are in the middle ground. They offer a compromise between the other two model types.

5. Recommendations and Business Value

- Reduced Defaults: Models like XGBoost with Borderline-SMOTE target a lower false negative rate, catching more potential defaults and reducing financial risk.
- Increased Approval Accuracy: XGBoost with ADASYN lowers false positives. This ensures creditworthy customers are serviced, potentially boosting market share.
- Balanced Risk Management: RF with SMOTE or Borderline-SMOTE offers equilibrium. This helps mitigate the risk while supporting customer acquisition.

6. Further Considerations

Further model tuning should consider domain expertise to set appropriate thresholds that balance recall and precision based on the business impact of misclassification. Additionally, Incorporating more data points could potentially enhance model performance, especially for the minority class. Moving forward, exploring advanced ensemble techniques or deep learning methods combined with more computational capacity could offer improvements over traditional machine learning models while making it easier to set appropriate thresholds. Having access to more computational resources would be greatly beneficial for tuning the hyperparameters of complex models, as it can significantly reduce the time required for such processes with limited resources.

7. Summary and Conclusions

We conclude our project analysis by highlighting the significance of choosing the appropriate prediction model based on specific business objectives and the costs associated with misclassifications. XGBoost with Borderline-SMOTE is recommended for minimizing the risk of overlooking bad debts, while XGBoost with ADASYN is preferable for reducing false positives. For a balanced approach to error management, RF models with either SMOTE or Borderline-SMOTE offer an effective solution. Implementing these tailored strategies can significantly enhance credit risk assessment processes, ensuring both financial stability and customer satisfaction.