Machine Learning Explainability in Mortgage Credit Approving: An QII Approach – Proposal

Yichen Gong, Yihan Xu, Frank Zhao

I. Introduction

Improvements in algorithmic decision-making techniques and better access to high quality borrower data have altered the way lenders make credit decisions [1]. Nowadays, machine learning programs are ubiquitous in financial institutions, ranging from banks, insurance companies, mortgage providers, etc. This modern technique is critical to financial institutions in the sense that it has stronger power in credit risk prediction than traditional methods and thus can bring higher profitability. However, there is also concern that the algorithmic decision-making process is often opaque [2].

An improvement that will be beneficial for the multiple stakeholders is to uncover the "black-box" decision-making process and add some *explainability* to the model. According to a recent study by the Bank of England, at least six types of stakeholders will benefit from this result, namely: developers for the ML application; 1st line model checkers; managements for the application; 2nd line model checkers; conduct regulators and prudential regulators [3].

In this paper, we aim to simulate the credit approval decision-making process using real mortgage data. More specifically, we are studying a classification process of whether a mortgage application would be approved or not, based on multiple lender-specific inputs. Firstly, we want to train several models for classification. We classify these models into three types according to their explainability.

(i) Explainable: These are linear models which directly gives weights to the inputs, so the coefficients will be exactly what we need for model explanation.

(ii) Probabilistic: These are the models that are not directly interpretable, i.e., not linear, but give probabilistic estimates of the classification result that we want. This type of model is the one that we are most interested in in this paper, and we would like to examine how the probability estimates will change by some intervention on the inputs.

(iii) Non-Probabilistic: These are the models that directly output classification results instead of probabilities. Explaining these models require a better understanding of the context because a quantity of interest other than the classification probability must be constructed prior to the study [2].

We plan to train one or two models of each type and compare their AUROC measures. PCA and other methods may be used for feature selection and pre-processing. Secondly and more importantly, we would like to explain one of the Probabilistic models which gives us the best AUROC, using the method of Quantitative Input Influence as proposed in Datta, et al. QII captures the degree of influence of inputs on outputs of systems. It can explain decisions about individuals and can quantify the joint influence of a set of inputs (e.g., age and income) on outcomes (e.g., loan decisions) since single inputs may not always have high influence [2].

For simplicity, we will only conduct Unary QII, which isolates the effect of each feature and intervene them independently. During the process, although correlation is not considered, we should be able to get the program response to a change of certain features. In the examination of whether the program is discriminative against a certain group, this approach will be especially helpful. We will compare the result with the baseline explainable model, i.e., logistic regression, to examine how the models differ and finally propose some insights on whether our best probabilistic model is suitable for use.

Explainable	Logistic Regression
Probabilistic	XGBT Trees, Bayesian Classifiers, Neural Networks (with Softmax output layer)
Non-Probabilistic	SVM, Decision Tree

Table 1. Types of classification models (Ratnayaka, 2020) [4].

II. Data

Each year, the Consumer Financial Protection Bureau (CFPB) require "both depository financial institutions and non-depository financial institutions" to submit information of covered loans in accordance with Home Mortgage Disclosure Act (HMDA) [5]. The information in the Home Mortgage Disclosure Act Dataset (HMDA Data in the later part of the proposal) includes Loan Type, Loan Purpose, Action Taken, Ethnicity, Loan-to-Value Ratio, etc. Among these features, "Action Taken" is at our predicting interest, as it "indicate the action [that the financial institution] taken on the covered loan or application". Other inputs will be used for prediction and our preliminary selection of features are listed in Appendix A.

The HMDA Data is available from 2007-2020, however, we are only using the data in a three-year window between 2018, 2019 and 2020 mainly due to three reasons: Firstly, the latest change of data schema happened after 2017. Secondly, the dataset itself is large enough with around 150,000,000 observations per year, so the lack of observations is not a valid concern. Thirdly, recent data have more predictive power for the mortgage happening in the future. In line with the time-series nature of the HMDA Data, we would like to use the mortgage applications in 2018 and 2019 as training and validation set and to use the 2020 applications as the testing dataset. However, we do recognize that COVID-19 imposed a lot of challenges to the mortgage markets as "many lenders are facing large volumes of customer relief requests that require adequate operational capacity and flexibility" [6]. As a result, whether the 2020 data is clean enough for testing our model is still to be examined. If not, we will use only data from 2018 and 2019.

Another issue we need to address before starting to train the model is that we face the limitation of computing power when dealing with the whole dataset (almost 500,000,000 observations). At the current stage, we plan to randomly down-sample the data to an extent that suits our computing capability.

In this proposal we will use the 2019 HDMA Data to demonstrate our pre-processing step. After down-sampling, we still have 175,591 observations in the 2019 dataset. Not all types of action taken are at our interest. Among 8 types, only two fully represents the lender's decision without reflecting the actions made by the applicant (e.g., withdrawn) or by other parties (e.g., purchased loan). These two types are "loan originated" and "application denied". Then, we try to process the missing values. Fortunately, most missing values are of continuous variables, so filling them with the mean is a possible approach. Another feasible way is to drop these observations, given that the number of missing values is not too large. The final thing worth noting is that, at this stage, we drop the observations whose income feature is “Exempt,” because this contains information about the financial institution instead of the loan itself. Summary statistics and more details can be found in Appendix B.

Appendix A Selected Features and Numbers of Missing Values

Feature	#Missing
activity_year	0
derived_loan_product_type	1
derived_dwelling_category	1
derived_ethnicity	0
derived_race	0
derived_sex	0
action_taken	0
preapproval	0
loan_type	0
loan_purpose	0
loan_amount	0
combined_loan_to_value_ratio	59,162
income	22,115
N	175,591

Table 2. Selected Features and Number of Missing Values, Downsampled, Before Dropping

Feature	#Missing
activity_year	0
derived_loan_product_type	0
derived_dwelling_category	0
derived_ethnicity	0
derived_race	0
derived_sex	0
action_taken	0
preapproval	0
loan_type	0
loan_purpose	0
loan_amount	0
combined_loan_to_value_ratio	7,877
income	6,856
N	114,993

Table 3. Selected Features and Number of Missing Values, Downsampled, After Dropping

Appendix B Summary Statistics

derived_race	#obs
White	79,798
Race Not Available	17,556
Black or African American	7,897
Asian	6,389
Joint	2,185
American Indian or Alaska Native	604
Native Hawaiian or Other Pacific Islander	349
2 or more minority races	188
Free Form Text Only	27
N	114,,993

Table 4. Summary Statistics: by derived_race

derived_sex	#obs
Joint	43,565
Male	37,253
Female	24,386
Sex Not Available	9,789
N	114,993

Table 5. Summary Statistics: by derived_sex

loan_purpose	#obs
Home purchase	49,994
Refinancing	26,902
Cash-out refinancing	19,355
Home improvement	9,595
Other purpose	9,062
Not applicable	85
N	114,993

Table 6. Summary Statistics: by loan_purpose

derived_loan_product_type	#obs
Conventional:First Lien	71,043
Conventional:Subordinate Lien	17,977
FHA:First Lien	14,809
VA:First Lien	9,951
FSA/RHS:First Lien	1,145
FHA:Subordinate Lien	61
VA:Subordinate Lien	7
N	114,993

Table 7. Summary Statistics: by derived_loan_product_type

References

[1] Andreas Fuster, Paul Goldsmith-Pinkham, Tarun Ramadorai, and Ansgar Walther. "Predictably Unequal? The Effects of Machine Learning on Credit Markets". *The Journal of Finance, Rapid Publication*, 28 October 2021.

[2] Anupam Datta, Shayak Sen, and Yair Zick. "Algorithmic Transparency via Quantitative Input Influence: Theory and Experiments with Learning Systems". *Proceedings of IEEE Symposium on Security & Privacy 2016*, pages 598-617, 2016.

[3] Philippe Bracke, Anupam Datta, Carsten Jung, and Shayak Sen. "Machine Learning Explainability in Finance: An Application to Default Risk Analysis". *Bank of England, Staff Working Paper No.816*, 2019.

[4] Gathika Ratnayaka. "Probability and Machine Learning? — Part 1- Probabilistic vs Non-Probabilistic Machine Learning Models". *Medium,* 2020. Available at medium.com/nerd-for-tech/probability-and-machine-learning-570815bad29d.

[5] "A Guide to HMDA Reporting". *Federal Financial Institutions Examination Council*, 2021. Available at ffiec.gov/hmda/pdf/2021Guide.pdf.

[6] "Covid-19: The Impacts on Global Residential Mortgage Markets". *Deloitte*, 2020.