

Financial Distress Prediction for Credit Scoring

蔡旻容、3位成員

Business Goal

- **Brief description of the company/context**

Banks play a crucial role in market economies. They decide who can get finance and on what terms and can make or break investment decisions. For markets and society to function, individuals and companies need access to credit. Credit scoring algorithms, which make a guess at the probability of default, are the method banks use to determine whether or not a loan should be granted.

- **Who is the stakeholder or client? What is the action triggered by the prediction?**

Bank, borrower, business, examiner. Banks make the best financial decisions to determine whether or not a loan should be granted.

- **A description of the business goal. What are the business benefits of implementing this idea? What opportunity is it creating? What shortcoming does it address?**

Improve on the state of the art in credit scoring by predicting the probability that somebody will experience financial distress in the next two years. The goal is to build a model that borrowers can use to help make the best financial decisions.

Improve the accuracy of credit scores, reduce the risk of human judgment errors, increase interest benefits, reduce bad debt rates, and lend funds to suitable individuals and companies.

For those with higher credit scores, banks can offer better loan options (higher loan amounts, lower interest rates, or longer grace periods)

For those with lower credit scores, banks will provide more conservative loan options (lower loan amount, higher interest rate or shorter grace period)

Machine Learning Goal

- **A description of the analytics objective.**

Use multiple feature analysis to predict whether a borrower will default on repayment.

- **Is this a supervised or unsupervised task? Predictive or descriptive? Retrospective or forward-looking?**

This is a supervised, Predictive task.

- **What is the main outcome variable(s) of interest?**

Age & MonthlyIncome

Data

- **Brief description of the available data.**

kaggle dataset : <https://www.kaggle.com/c/GiveMeSomeCredit/overview>

The data columns and descriptions are as follows:

Variable Name	Description	Type
SeriousDlqin2yrs	Person experienced 90 days past due delinquency or worse	Y/N
RevolvingUtilizationOfUnsecuredLines	Total balance on credit cards and personal lines of credit except real estate and no installment debt like car loans divided by the sum of credit limits	percentage
age	Age of borrower in years	integer
NumberOfTime30-59DaysPastDueNotWorse	Number of times borrower has been 30-59 days past due but no worse in the last 2 years.	integer
DebtRatio	Monthly debt payments, alimony, living costs divided by monthly gross income	percentage
MonthlyIncome	Monthly income	real
NumberOfOpenCreditLinesAndLoans	Number of Open loans (installment like car loan or mortgage) and Lines of credit (e.g. credit cards)	integer
NumberOfTimes90DaysLate	Number of times borrower has been 90 days or more past due.	integer
NumberRealEstateLoansOrLines	Number of mortgage and real estate loans including home equity lines of credit	integer
NumberOfTime60-89DaysPastDueNotWorse	Number of times borrower has been 60-89 days past due but no worse in the last 2 years.	integer
NumberOfDependents	Number of dependents in family excluding themselves (spouse, children etc.)	integer

- **Some guidance on the data subset that will be used and the re-processing or preparation that might be needed based on your past experience.**

1. Data preprocessing
 - Data cleaning (missing data...etc.)
 - Data transformation (feature selection...etc.)
2. Training Data model
 - Decision Tree
 - Scikit-Learn : LogisticRegression, RandomForest...etc.

3. Data visualization

- **Sample of ten rows (records) with ten columns (variables) that will be used, including the outcome column.**

	A	B	C	D	E	F	G	H	I	J	K	L	M
1		SeriousDl	Revolving age		NumberOf	DebtRatio	MonthlyIr	NumberO	NumberO	NumberR	NumberO	NumberOf	Dependents
2	1	1	0.766127	45	2	0.80298213	9120	13	0	6	0	2	
3	2	0	0.957151	40	0	0.1218762	2600	4	0	0	0	1	
4	3	0	0.65818	38	1	0.08511338	3042	2	1	0	0	0	
5	4	0	0.23381	30	0	0.03604968	3300	5	0	0	0	0	
6	5	0	0.907239	49	1	0.0249257	63588	7	0	1	0	0	
7	6	0	0.213179	74	0	0.37560697	3500	3	0	1	0	1	
8	7	0	0.305682	57	0	5710 NA		8	0	3	0	0	
9	8	0	0.754464	39	0	0.20994002	3500	8	0	0	0	0	
10	9	0	0.116951	27	0	46 NA		2	0	0	0	NA	

Methods

- **What are some machine learning methods to consider?**
Decision Tree, Logistic Regression, Random Forest, KNN
- **Which performance measures are appropriate? How do they map to the business goal?**
Use classification methods to find models with higher prediction accuracy to achieve the goal of more accurately predicting whether borrowers will default on their repayments.

Implementation & Production

- **Operational requirements or constraints (who exactly will use the system and how? will the solution run in realtime? Will it require collecting new data? one-time analysis or ongoing?)**
Financial distress prediction is a key issue in measuring corporate solvency. It can help to identify risks early, make plans according to the actual situation, and adjust business strategy.
Banks or borrowers will use this system to understand the user's credit rating and whether they can lend to this person. The solution to run in realtime and require collecting new data. It's an ongoing analysis.