Credit Default Detector

Joel Choe

Amogh Gupte

Chintan Mehta

Amit Rajwani

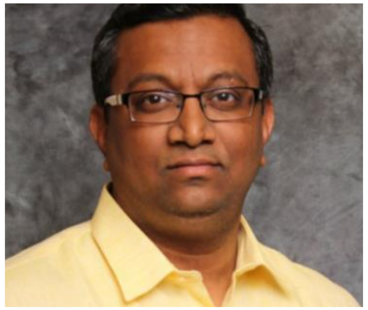
The Team

Joel Choe

Associate on the OTC Derivatives team within Basel Measurement & Analytics

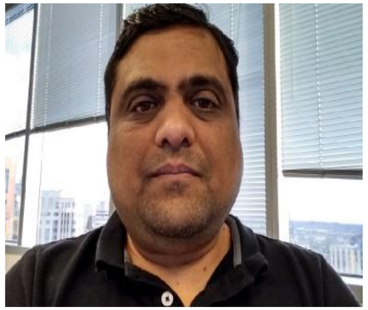
First got interested in data/analytics because of fantasy basketball

Favorite Quote: "You come at the king, you best not miss." - Omar Little from "The Wire"



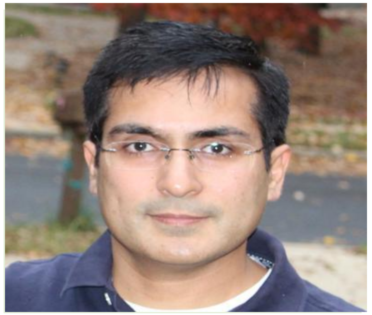
Amogh Gupte

Amogh has been with JP Morgan for the last 12 years as a Java developer in the corporate tech space. He loves hiking spending time with kids.



Chintan Mehta

Chintan has been with JP Morgan Chase more than 8 years and currently working as Strategic Business Analyst with CCB Remediation and Corrections team. I like to travel and explore new places, listening classical Indian music and Bollywood songs and watching movies.



Amit Rajwani

Amit works in CCB Credit Card Technology for the C3 Authorizations Platform as an Agility Lead. He has been with the firm for 8 years and likes to travel and be outdoors with nature.

Executive Summary

- In 2019 JP Morgan reported $5.6 billion net charge-offs.
- Our aim is to build a scalable, automated solution to predict defaults.
- Our best model (with open-sourced data) was able to catch about 65% of the defaults.

AI Canvas : Using Machine Learning for Credit Card Loan Approval Decisions

- Investigate use of machine learning to see what insights and potential benefits can be created when applied to the loan approval process
- Looking for more nuance than traditional decision-making process based on credit history/score that maximizes revenue
- Made use of dataset via Kaggle containing information about credit card loan transactions (166 attributes and over 300k records)

The AI Canvas

This data set is uploaded in order to get the insights of Credit card Defaultees based on the respective attributes. The consumer lending line of business @ JPMC is in the business of lending money to customers for loans, credit cards, mortgage, etc. and a model like this which can predict potential defaults would be immensely helpful in making lending decisions. The aim of the project is to provide this service, based on a ML model, which will be repeatable, scallable and retrainable.

Prediction

Predict whether an applicant will default

Judgment

For this use-case catching the default case is more important, as if a client defaults the firm will loose the whole credit amount.

Action

A customer's application can be approved or declined based on the inputs provided.

Outcome

Recall on the testing data is the main criteria here

Training

To over come the skewed input data we performed over sampling.
We reduced the inputs to the model to 7 from 166 odd inputs.

Input

DAYS_EMPLOYED
DAYS_BIRTH
AMT_ANNUITY
AMT_CREDIT
AMT_INCOME_TOTAL
AMT_GOODS_PRICE
AMT_REQ_CREDIT_BUREAU_YEAR

Feedback

Create an automated trigger or periodic review to evaluate the deployed model for tuning/refresh based on performance of the model on real data after implementation

A customer's application can be approved or declined based on the inputs provided. AI will not replace the staff, but will help in better decision making and tremendously improve the decision timing. Also it will bring consistency in the decision making process and provide additional scalability and easy maintainability. The staff feedback can help improve the AI.

Key observations of the dataset & Pre-processing

- This is an imbalanced dataset. There are far more loans that were repaid on time than loans that were not repaid
- Mean employment tenure is skewed due to outliers and needs to be filtered to get a better model
- The Risk of Failure to repay was higher in the younger age group as compared to older applicants. As the age increased the risk decreased
- We handled missing data, data errors and outliers
- A large number of attributes were dropped since they either provided duplicate information or were insignificant in model performance

Input

Column Name	Column Description	Weight
AMT_ANNUITY	Loan annuity	0.14
AMT_CREDIT	Credit amount of the loan	1.21
AMT_GOODS_PRICE	For consumer loans it is the price of the goods for which the loan is given	-1.41
AMT_INCOME_TOTAL	Income of the client	-0.09
AMT_REQ_CREDIT_BUREAU_YEAR	Number of enquiries to Credit Bureau about the client one day year (excluding last 3 months before application)	0.09
DAYS_BIRTH	Client's age in years at the time of application	0.15
DAYS_EMPLOYED	How many months before the application the person started current employment	0.24

Model Selection

With our project:

- A false positive will mean that the firm will lose the business of a potential 'good' client.
- A false negative will result in the firm losing money due to loaning to a defaulter client.

From the above we believe that a false negative will have a higher impact to model performance. We have compared the below models and come up with Logistic Regression model as it will give us the best recall for defaulters in test data. KNN also has the same recall, but it has a higher false positive rate than Logistic Regression.

Comparing our chosen model to the **baseline** (**marking all applicants as not default**):

Baseline Accuracy Score : 0.9159
Our Best Accuracy Score : 0.91527
Baseline Recall Score : 0.0
Model Recall Score : 0.658

Name	Accuracy	Recall	precision
Decision Tree	60.6%	57.9%	52.5%
Random Forest	91.4%	0.7%	91.2%
Logistic	**56.3%**	**65.8%**	**53.5%**
KNN	84.8%	10.8%	85.1%
XGBoost	89.5%	7.1%	87.4%

As it can be seen our model gives considerably better Recall over the baseline for defaulter case.

Hyper Parameter tuning

- We ran the various classifiers through different configurations
- For Logistic Regression (Chosen solution), we tried:
  - Constant: [1.0, 0.5, 0.2, 1.5]
  - Penalty: [L2, none]
  - Solver: [newton-cg, lbfgs, liblinear, sag, saga]
- With best results from:
  - Constant: 1.0
  - Penalty: none
  - Solver: sag

Demo & Repo

- https://slackers-ml.herokuapp.com/
- https://amoghugupte.github.io/Slackers-Capstone/
- Data set is from Kaggle : https://www.kaggle.com/mishra5001/credit-card

Demo

Credit Card Application

Making Prediction

Please provide customer information:

Loan annuity

36000.00 − +

Credit amount of the loan

1293502.00 − +

For consumer loans it is the price of the goods for which the loan is given

1129500.00 − +

Income of the client

270000.00 − +

Number of enquiries to Credit Bureau about the client one day year (excluding last 3 months before application)

0.00 − +

Client's age in years at the time of application

20.00 − +

How many months before the application the person started current employment

36.00 − +

This customer is predicted to be: **Default**