Detail Project Report -- Wells Fargo Whole sell LOAN RECOVERY Prediction

## About the Project

Wells Fargo Commercial card management system has customer’s 7years of transactional and payment data. When an wholesale customer applied for loan, this process it to predict the risk associated using all the financial data for the corporation and other obligation information available

## What is the size of data

In wells Fargo Commercial Card System, we have last 7 year customer credit card transaction and Payment data.

## What was the data type

Data stored in Oracle relational database in as a data warehouse in a least normalized format for easy retrieval.

## What was the team size and Distribution

The team is divided into multiple functional unit, distributed across different geographical location for 24\*7 work in progress. All the team has individual scrum team with product owner and scrum muster and having 8-12 team member for each team.

Credit Card Statement team

Payment processing team

Front-end team

Reporting team

Production Support team

## Any Big data environment

We did not had any exposer to big data.

## Environments

CCDMS system has multiple environment DEV, SIT, UAT, PFIX, PROD, and BCP. In addition, we have multiple divisions like DEV1, DEV2, SIT1, SIT2 or UAT1, UAT2 for parallel work and testing.

## Creating and maintain logs

We take logging very seriously in CCDMS system. We have well structure reusable logging mechanizing to use in UNIX and Python scripting and for Database coding. Also well-established error alert and log archival process in place.

## Techniques used in data preprocessing and visualization

The credit card statement data is well structure and good part is we have access to direct database other than PDF credit card statement. As a result, not much data processing is required. Data visualization work is achieved using python matptolib and seaborn to be used for displaying product owner.

## Marinating the failure cases

For failing cases trigger is in place to generate notification and responsible support team to take care is possible, other cases development team to be contacted for details that are more technical.

## Automation done for Data processing

The data processing is fully automated, Daily transaction and Payment data is processed in a batch process and loaded into CCDMS database to create customer credit card statement.

## Scheduler

Autosys scheduler is in use .

## Job Monitoring

Dedicated production support team to monitor prod jobs. Development team takes care of the UAT jobs.

## My roles and responsibilities

I was part of loan data processing team responsible for data analysis, coding, data processing, documentations and scheduling jobs.

## My day to day task

Day to day activities include

* Daily scrum connecting with product owner
* Data analysis
* Creating SSD, FSD and flow diagrams and Approach document to show case.
* Reviewing documents with business owner
* Coding to try different machine learning models
* Discussing output with AI leads to select most appropriate model
* Migrating code to different environment
* Scheduling jobs
* Connecting to front end team
* Analysis on retraining and setup retraining process
* Helping prod support team for failure

## Core area of contribution

My core area of contribution was data processing and analysis and applying machine-learning algo to and discussing with AI architects about output. Then finalize the documentation and coding

## In which technology I am most comfortable

I am most comfortable with machine learning process along with all the supporting technologies

## How will I rate myself on the technology

I will rate myself a 7.5 in machine learning related technical area.

## How many project worked in total

I have worked in 2 projects and 1 POC in the field of machine learning

## Deployment Process

The deployment is on premise, did not had any cloud environment. Me as a developer was responsible to prepare game plan for deployment and hand it over to platform support tam to execute steps in game plan and post deployment validation and support.

## Challenges faced during project

Challenged faced during processing obligation information as no concrete source of quantified data. Extreme EDA preformed to decode the data and making it quantity data.

## Expectation

My expectation for Future project is to update myself with different problem statements, technologies, and solution approach

## Future Objectives

To be updated about most of the data science technologies and update AI and myself with latest fields

## Why I am leaving Current project

Current project is done .Looking for new opportunities.

## How did I do data validation

Data validation is done in a prod parallel testing process. I have develop data validation scripts to validate data for different scenarios,

## How did I do data enrichment

Data enrichment was an important aspect as most of the data was coming from legacy system. We processed the data for each tran code and payment type , decoded them more readable format and detokenized data to be better prediction.

## How would I rate myself in Machine learning

I will rate myself a 7.5 in machine learning

## How much time model take to train

The first time train had taken almost 120 hours to on 7 years of history data

## What is the frequency for retraining and updating model

Re training happens every month fifth business day.

## Which mode project get deployment

The project was on premise in company’s internal network.

## Areas of Machine leaning explored

Linear Regression.

Logistic Regression.

Decision Tree.

SVM.

Naive Bayes.

kNN.

K-Means.