Model training and Prediction on Home Credit Default Risk

Can you predict how capable each applicant is of repaying a loan?

GOALS

Process data at scale

· Efficiently train and update models based on large-scale data
· Prediction based on large amounts of data

High AUC

Ensure that：
· Clients capable of repayment are not rejected
· Principal can be repaid on time.

Real-time Prediction

· Real-time preliminary loan review after filling in information online or at outlets
· Avoid loss of potential customers due to long manual review cycle

TOOLS WE USE

Amazon S3

DATA STORAGE

Storage of training and testing data

Amazon SageMaker

MODEL TRAINING

· Feature engineering
· Model training & optimization

python

DATA STREAMING

Process data streams

Amazon QuickSight

VISUALIZATION

· Deliver insights
· interactive visualization

OVERVIEW OF THE SOLUTION

MODEL TRAINING

Amazon S3
Historical raw data → Amazon SageMaker
Data Processing → Amazon SageMaker
Model Training → Trained Model

STREAM PREDICTION

Live data → Stream the data → Amazon SageMaker
Prediction： AUC ~ 80 → Amazon QuickSight
Visualization: deliver insights

BUSINESS VALUE

Revenue Increase

With high accuracy and recall score of the model, bank can lend to as many qualified lenders as possible, boosting revenue.

Respond in time

With real-time prediction the results can be presented in minutes, improve customer experience.

Business Insights

Get business insights intuitively from the interactive visual environment.

TEAM 5

Anjali Chikkudukaila chikk009@umn.edu
Ariana Irani irani018@umn.edu
Pranav Chakradhar chakr184@umn.edu
Rohan Rajeev rajee006@umn.edu
William Missling missl008@umn.edu
Zhewen Zhang zhan8190@umn.edu