Cloud-Based Vehicle Loan Default Prediction

Team Members:
1. Jil Patel
2. Malavika Sreedhar Tankala
3. Shubham Singla
4. Sreevikram Chandrasekhar

Motivation

- Loans are one of the most important aspects of the finance world. One of the major sources of profit for financial institutions comes from the same.
- When a customer applies for a loan, the concerned organization has to review multiple parameters.
- This process is very time-consuming which increases the workload within the organization.
- Also, any human error in this process can have a drastic negative effect.

Problem Statement

- The objective is to predict whether customers are capable of repaying the loan.
- We propose that this review process must be automated using a Machine Learning Model hosted on the cloud.
- Our project has achieved the same using the various services provided by the Amazon Web Services.

Background - Dataset

- The data has been split as follows:

  Training =70% & Testing =30%

- For the purpose of this project the dimensions of the dataset have been reduced.
- We obtained the dataset from Hackathon.
- The dataset is of Crescent bank.

Background -Machine Learning (ML)

What is Machine Learning?

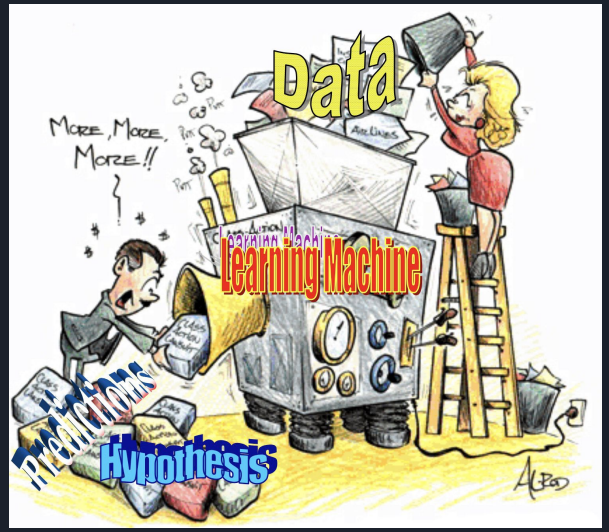
"Field of study that gives computers the ability to learn without being explicitly programmed." - Arthur Samuel (1959)

- Machine learning is taking large amounts of data, run it through a model to detect patterns, summarise, learn trends.
- So that predictions can be made for future data.

Data -> Hypothesis/Model -> Prediction
(Input X) (Output Y)

Machine Learning is broadly classified into two categories:

- **Supervised Learning** (Learning with labels)
  - Used in most real world applications
- **Unsupervised Learning** (Learning w/o labels)



Picture courtesy: CS10-315, Prof. Aarti Singh, CMU

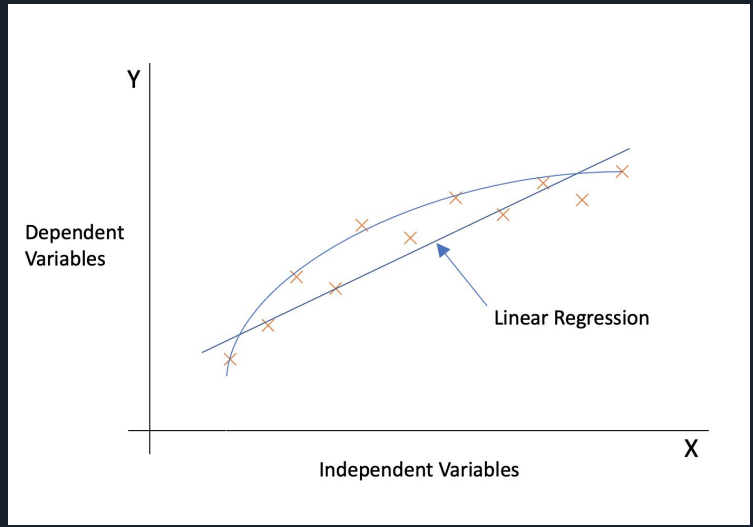
Background -Machine Learning (ML)

Supervised Learning:

- Regression:
  Predict a number from infinitely possible outcomes
- Classification (predict categories)

Linear Regression is a commonly used form of predictive analysis. It provides a linear relationship between one dependent variable and one or more independent variables.

XGBoost (Extreme Gradient Boosting) is a popular and powerful machine learning algorithm that uses decision trees to make predictions. It is known for its speed, scalability, and high performance in various machine learning tasks.



Background - AWS SageMaker

- AWS SageMaker is a fully-managed service that enables developers and data scientists to easily build, train, and deploy machine learning models at scale.
- With SageMaker, users can access a range of tools and features, including pre-built algorithms and frameworks, automatic model tuning, and deployment options for real-time and batch inference.
- SageMaker is designed to be highly flexible, allowing users to bring their own algorithms and frameworks, use different programming languages and environments, and integrate with other AWS services.

Background - AWS Simple Storage Service(S3)

- AWS S3 provides object storage, which is built for storing and recovering any amount of information or data from anywhere over the internet.
- An object consists of the data, key and metadata. A bucket is used to store objects.
- AWS S3 provides many benefits including but not limited to durability, low cost and scalability.

Background - AWS Amplify

- AWS Amplify is a cloud-based development platform that helps developers build web and mobile applications quickly and easily.
- User doesn’t need to manage the underlying infrastructure.
- With AWS Amplify, developers can easily integrate features such as authentication, APIs, databases, and storage into their applications.

Background - Amazon Cognito

- Amazon Cognito provides authentication, authorization and user management for web and mobile applications.
- The 2 main components are User pools and Identity pools.
- User pools are user directories that provide sign-up and sign-in options.
- Identity pools allow granting users access to other Amazon services.

Background - AWS Lambda

- AWS Lambda is an event driven computing service that responds to triggers.
- It runs code on high-availability compute infrastructure and takes care of complete administration of compute resources.
- Pay per request pricing model and Autoscaling(scale compute and memory resources based on-demand) are its main advantages.

Background - AWS API Gateway

- APIs act as the "front door" for applications to access data, business logic, or functionality from your backend services.
- AWS API Gateway supports creating, deploying and managing RESTful or WebSocket APIs to expose the backend HTTP endpoints, AWS Lambda functions or other AWS services.

Background - AWS SNS & AWS SES

- AWS Simple Notification Services (SNS) is a cloud-based services that delivers push notifications to subscribers.

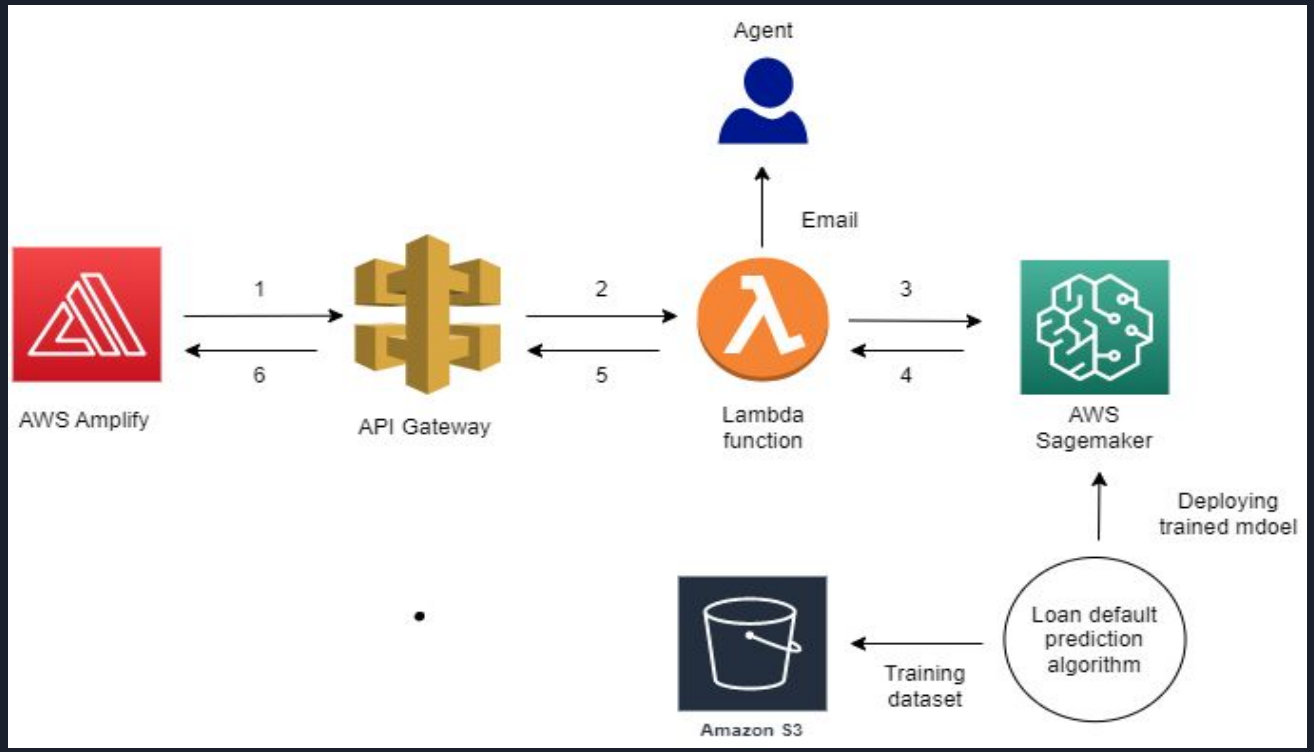
- AWS Simple Email Services(SES) is a cloud-based email service that allows businesses to send emails to users in a cost-effective and scalable way.

Background- Related Work

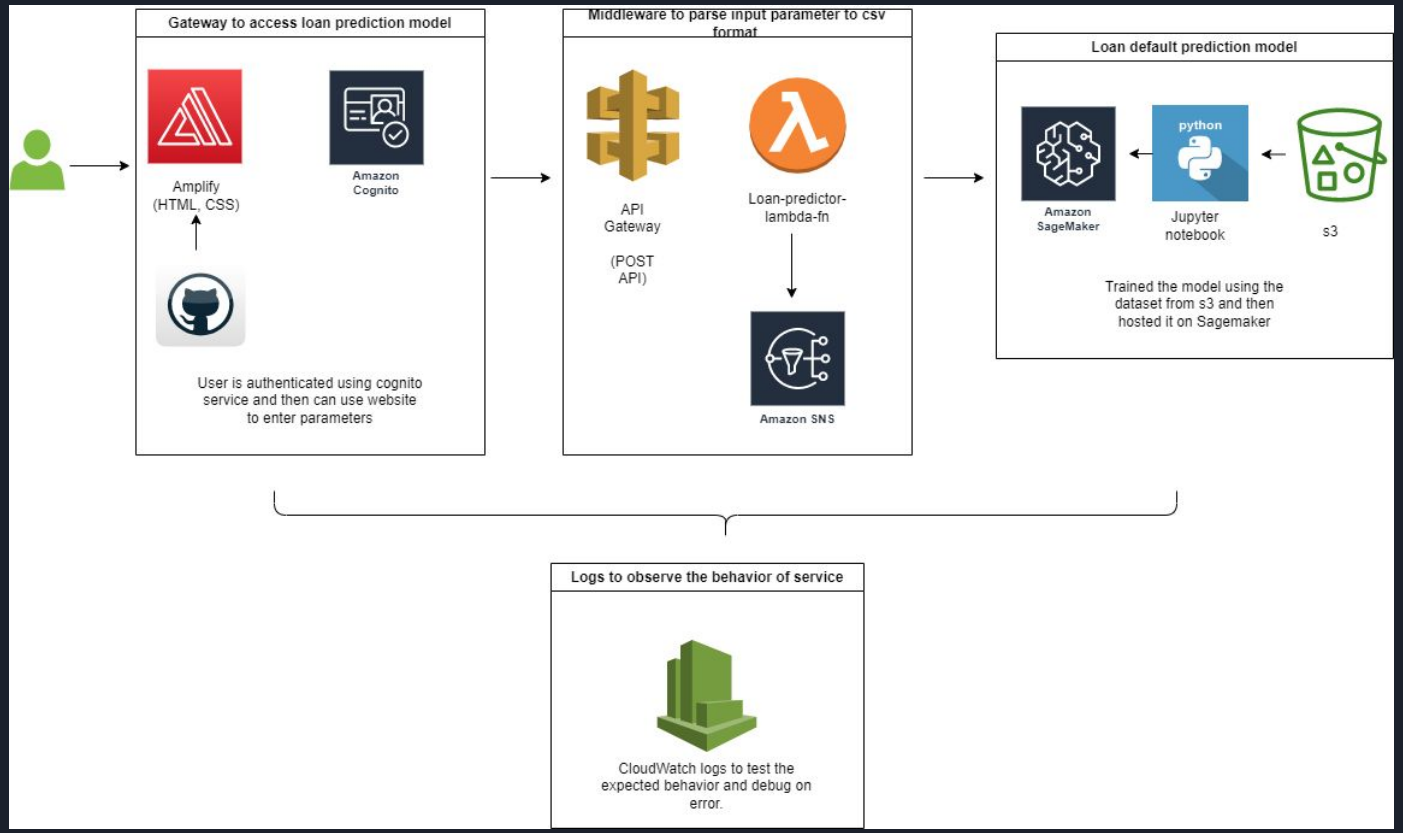
- Related work was found in research papers where they implemented various Machine Learning models and libraries including but not limited to Logistic Regression, Decision Trees, Random Forest and Support Vector Machine.

- Our implementation uses the Regression XGBoost model and has a significant advantage as the application is hosted on the cloud. This allows the reviewing financial institutions to leverage cloud specific features like scalability, high-availability, security, reliability and cost-effectiveness.

System Architecture



Software Architecture



Outcome

- We were able to understand how cloud services helps us to train ML models on huge datasets using various distributed system concepts to divide the load.
- To host the models on cloud server.
- Understand the functioning of FaaS (Function as a service) by practically implementing them in our use case.
- How to create the website and host it by abstracting various distributed system principles such as scalability, security, maintainability etc.

Possible Future Work

- Currently, the application works for the business case of Vehicle Loan Default Prediction, however we can scale the application for different use cases as well.
- Currently we are just displaying the percentage of loan default value but in future we can tell client what are the negative values affected.
- Create dashboard to record history of loan that were approved or rejected.

Demo

References

- https://www.sciencedirect.com/science/article/pii/S1877050919320277
- https://iopscience.iop.org/article/10.1088/1757-899X/1022/1/012042/pdf
- https://www.itm-conferences.org/articles/itmconf/pdf/2022/04/itmconf_icacc2022_03019.pdf
- https://ijcrt.org/papers/IJCRT2106313.pdf
- https://docs.amazonaws.cn/en_us/cognito/latest/developerguide/what-is-amazon-cognito.html
- https://docs.aws.amazon.com/lambda/latest/dg/welcome.html
- https://docs.aws.amazon.com/sagemaker/latest/dg/whatis.html
- https://docs.aws.amazon.com/AmazonS3/latest/userguide/Welcome.html
- https://docs.aws.amazon.com/amplify/latest/userguide/welcome.html
- https://docs.aws.amazon.com/apigateway/latest/developerguide/welcome.html
- https://docs.aws.amazon.com/sns/latest/dg/welcome.html

Thank You