



INDIANA UNIVERSITY
BLOOMINGTON

ENGR-E516 Engineering Cloud Computing

Project Proposal

Jayesh Prasad Anandan

Afeefa Bano

Navya Sree Santhapeta

February 17th, 2023

Retail Sales Prediction and Application

using AWS SageMaker, AWS Lambda and AWS Quicksight

Introduction and Motivation

Retail businesses are data-driven!

Techniques for making future predictions based upon the present and past data have always been an area with direct application to various real-life problems. Sample historical sales data and other relevant information such as departments, holidays, stores, markdown are gathered from the retail sales management system to forecast future sales for easy scalability and maintenance. Accurate sales forecasting can help businesses optimize their operations, reduce costs, and improve their profitability.

The main objective of this project is to develop a machine learning (ML) predictive model that can generate accurate retail sales forecasts using AWS SageMaker.

Existing Tech

Amazon SageMaker is a comprehensive platform for managing machine learning workflows that includes tools and services for all stages of the process, from data preparation to model deployment. For our project, we plan to utilize XGBoost in AWS SageMaker, which is a supervised learning technique that uses the gradient-boosted decision tree algorithm. In this algorithm, the data is divided into subsets, and decision trees are constructed on each subset. The output from each tree is combined to create the final prediction, producing an ensemble of decision trees that can improve the accuracy of the model. This approach is particularly useful for regression tasks like retail sales prediction, where we are forecasting continuous numerical values.

By leveraging AWS SageMaker, we can take advantage of the platform's managed training and hosting capabilities to create a scalable and precise predictive model for retail sales forecasting. With access to a variety of tools and services, we can streamline the machine learning workflow and produce accurate sales predictions that help businesses optimize their operations, reduce costs, and improve their profitability.

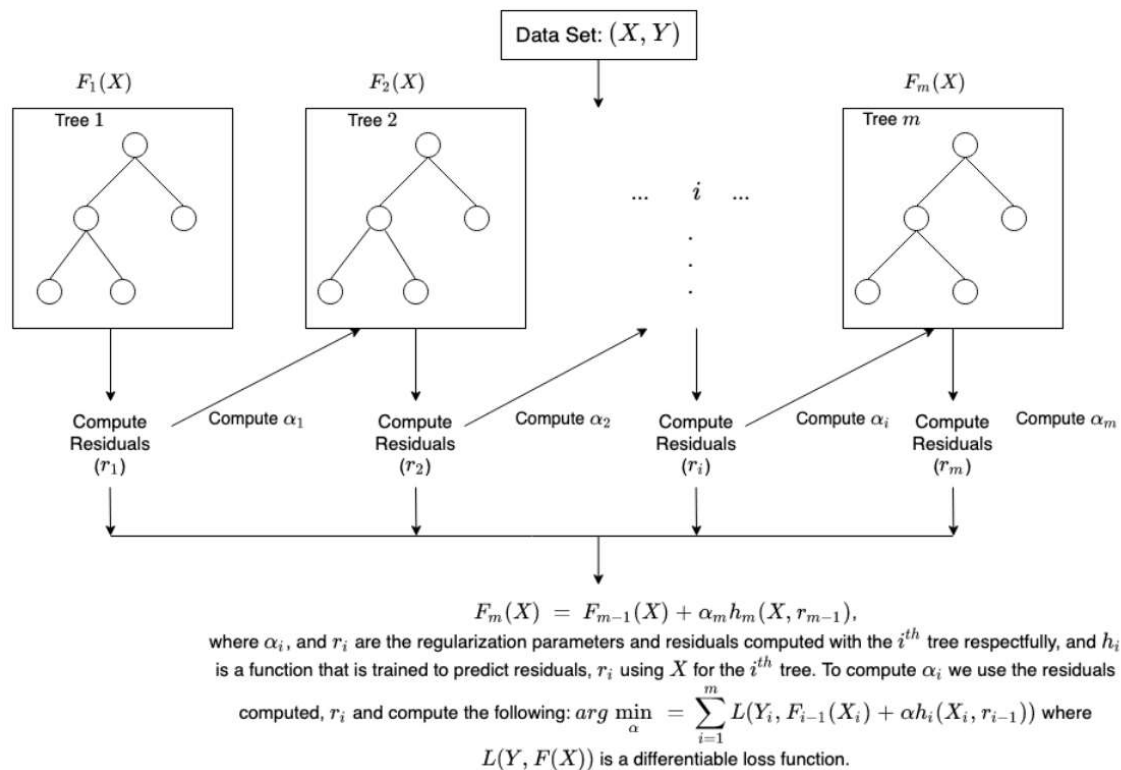


Image Source: <https://docs.aws.amazon.com/sagemaker/latest/dg/xgboost-HowItWorks.html>

Related Work

There have been several related works in the field of retail sales prediction using machine learning techniques. Some of the notable ones are as follows:

1. "Predicting Future Sales and Profit Margin of Retail Stores using Machine Learning" by Vijayabhaskar et al. (2018): In this study, the authors proposed a predictive model for retail sales and profit margin using machine learning techniques. The model was trained on a dataset of historical sales data and used a combination of regression and clustering algorithms to generate accurate forecasts.
2. "Retail Sales Forecasting using Machine Learning Algorithms" by Srinivas et al. (2020): In this study, the authors proposed a retail sales forecasting model using machine learning algorithms, including XGBoost. The model was trained on a dataset of historical sales data and used several input features, including product attributes, promotions, and seasonal trends, to generate accurate sales forecasts.
3. "Time Series - ARIMA, DNN, XGBoost Comparison" from Kaggle Blog Post. This blog is about comparison of three models; ARIMA, DNN and XGBoost on Time Series Dataset. The format of the data set on which the comparisons were drawn matches the format of the data set on which we can make our own predictions.

All these related works highlight the importance of accurate sales forecasting in retail and the potential benefits that can be gained from using machine learning techniques. The proposed project, Retail Sales Prediction Using AWS SageMaker XGBoost (Regression), aims to build upon these existing works by using AWS SageMaker's managed training and hosting platform to create a scalable and accurate predictive model for retail sales forecasting.

Proposed Method

We will begin the project by configuring our AWS environment and providing access to all team members. Initially, we will upload our dataset of retail sales prediction to our AWS S3 bucket for further use. Later, the preprocessing and EDA of the data will be performed. Further, the prediction model will be implemented in AWS Sagemaker and trained with the retail sales data. Once the model is trained, we will be testing the model with untrained data to observe the model's accuracy. After that, we will implement high-performance tuning for finding the best hyperparameters for the model.

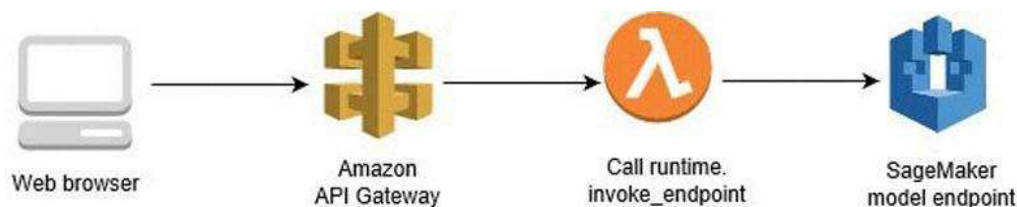


Image Source: <https://aws.amazon.com/blogs/machine-learning/call-an-amazon-sagemaker-model-endpoint-using-amazon-api-gateway-and-aws-lambda/>

For the second half of the project, we will use AWS Lambda to get an endpoint to the AWS Sagemaker and connect it with AWS API Gateway to create an API for our model. We can use this API to create a website where a user can utilize our model to find future sales predictions. Also, we will be creating a dashboard from our data to visualize the trends, and insights into the sales. Finally, we will consolidate our observations and report it.

Team Allocation

The project contribution will be equally divided between our team members. According to our skillset, the development and implementation of the Prediction Model till building an API will be taken care of by Jayesh and Afeefa. The website development and deployment will be taken care of by Navya Sree. Jayesh will solely build dashboards in AWS Quicksight. However, everyone will be helping one another with every task. In the end, the report will be consolidated and developed by everyone.

Timeline

Week 1: System environment configuration and Dataset upload in AWS S3.

Week 2: Preprocessing and EDA on the Data using AWS SageMaker.

Week 3: Implement Prediction Models on the data in AWS SageMaker.

Week 4: High-performance tuning and Integration of AWS Lambda.

Week 5: Creation of API Gateway for our model. Development of Website to utilize the Model.

Week 6: Creating Dashboard to visualize the data and results using AWS Quicksight.

Week 7-8: Summarize observations and submit the project report.

References

1. [Retail Sales Dataset from Kaggle](#)
2. [AWS SageMaker Inbuilt Model Tuning](#)
3. [AWS API Gateway Integration with AWS Lambda and AWS SageMaker](#)
4. [AWS Sagemaker XGBoost](#)
5. [XGBoost: A Scalable Tree Boosting System](#)