
Telecom Churn Prediction using AWS Sagemaker

Presented By: Sonu Adhikari



Table of contents



01

Introduction

02

Problem Statement

03

Objectives

04

Related Works

05

AWS Architecture

06

Future Work





Introduction

- Customer Churn means customer leaving a company or discontinuing their association
- Telecommunications industry is characterized by intense competition and high customer churn rates
- Timely identification of customers at risk of churn remains a significant obstacle



Problem Statement



Telecom companies face challenges in retaining customers due to high churn rates.



Existing Machine Learning aren't using fully managed Cloud environment like Sagemaker for training, deploying, and managing machine learning models



Organisations aren't still utilizing SageMaker's managed infrastructure that eliminates the need for upfront investments in hardware or ongoing maintenance costs.





Objectives

- Develop a machine learning model using AWS SageMaker's classification algorithm to predict customer churn, in this case only used the XGBoost algorithm
 - Deploy the model as a real-time prediction endpoint
 - Ultimately enable telecom operators to implement the endpoints in their systems for proactive retention strategies based on churn predictions.
-



Related Works

1. ***A Customer Churn Prediction Model using XGBoost for the Telecommunication Industry in Nepal:*** *XGBoost has been used for a Telecommunications service in Nepal*
 2. ***Churn Prediction in the Telecommunication sector using Support Vector Machines:*** *The paper has used a Support Vector Machines algorithm with four kernel functions to implement the predictive models. The performance of the models is evaluated and compared using gain measure.*
 3. ***Churn Prediction in Telecommunication using Logistic Regression and Logit Boost:*** *In this proposed model, two machine-learning techniques were used for predicting customer churn Logistic regression and Logit Boost. Experiment was carried out in the WEKA Machine-learning tool*
-



Dataset

- The dataset contains information of all 7,043 customers from a Telecommunications company in California in Q2 2022.
- Each record represents one customer, and contains details about their demographics, location, tenure, subscription services and more.

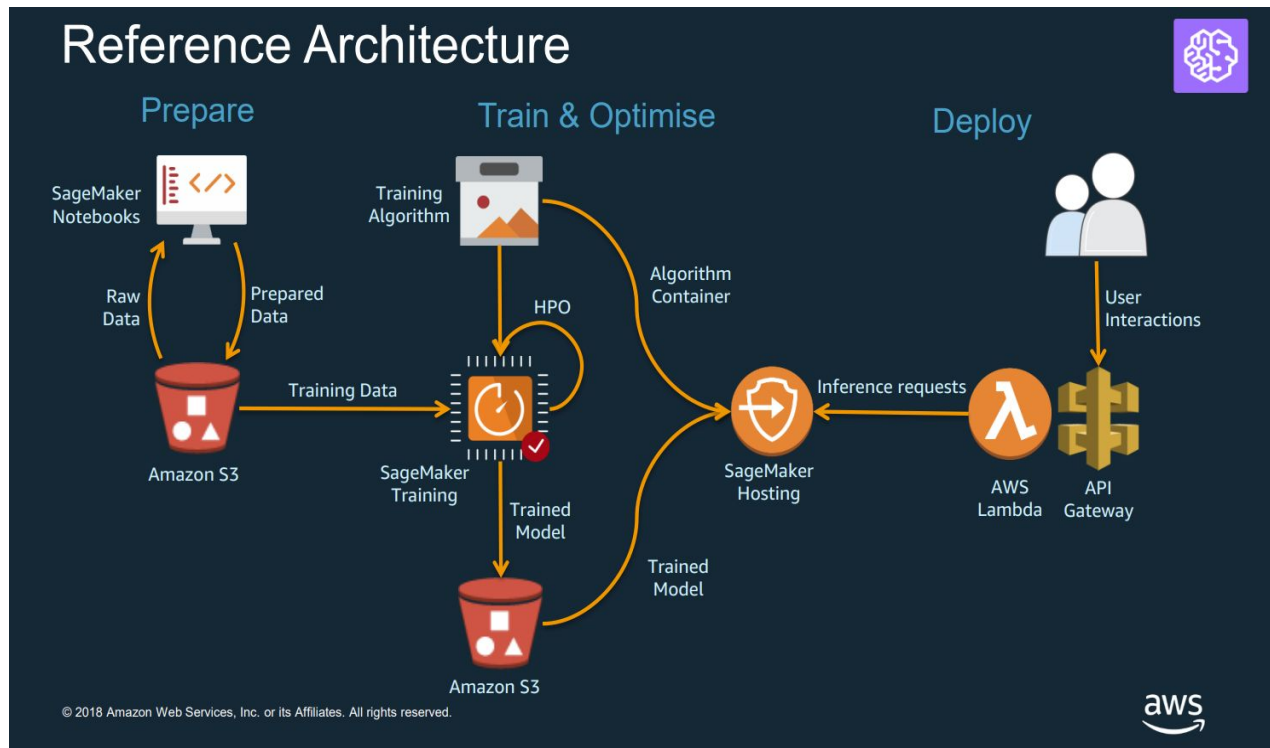
	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	DeviceProtection	TechSupport	StreamingTV	StreamingMovies	Contract	PaperlessBilling	PaymentMethod	MonthlyCharges	TotalCharges	Churn
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	No	No	No	No	Month-to-month	Yes	Electronic check	29.85	29.85	No
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	Yes	No	No	No	One year	No	Mailed check	56.95	1889.5	No
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	No	No	No	No	Month-to-month	Yes	Mailed check	53.85	108.15	Yes
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	Yes	Yes	No	No	One year	No	Bank transfer (automatic)	42.30	1840.75	No
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	No	No	No	No	Month-to-month	Yes	Electronic check	70.70	151.65	Yes

5 rows x 21 columns



AWS Architecture

- S3 for initial raw data collection and storage of preprocessed data
- Sagemaker for training the dataset
- Storing the trained model in S3 again
- Sagemaker hosting takes the trained model after getting a request from Lambda function
- The users can interact with the predictive model using API Gateway





Well-Architected Framework



- ❖ **Security:** Using IAM Role instead of Root Account with least privilege principle
- ❖ **Cost Optimisation:**
 - Setting up cloudwatch alarm for cost tracking since it was a problem
 - Continuously monitor SageMaker resource usage and right-size instances based on workload demands
- ❖ **Sustainability:**
 - Using the minimum amount of hardware to meet your needs and instance types with the least impact
 - Stop the creation and maintenance of unused assets
- ❖ **Performance Efficiency:**
 - Select appropriate SageMaker instance types and sizes based on requirements
- ❖ **Operational Excellence:** Prepare for operations by automating processes by using Cloudformation



Challenges and Future Enhancements

- While using Sagemaker, cost optimization was a big challenge(even led to account suspension)
- Fetching the desired output while testing the endpoint was an issue

- End-to-end MLOps Implementation, including pipelining
- Fine Tuning the Sagemaker Endpoint
- Experimenting with more algorithms
- Integration of the endpoint with a real-time system



Thank You !

Do you have any questions?
