PROJECT REPORT - IBM BUILD-A-THON
Telecom Customer Churn Prediction using Watson Auto AI

Abhishek Tripathi

Email: abhishektripathi911@gmail.com

LinkedIn: https://www.linkedin.com/in/abhishek--tripathi/

Github: https://github.com/tripathi-abhishek

Project Link: https://github.com/SmartPracticeschool/SPS-7809-Telecom-Customer-Churn-Prediction-using-Watson-Auto-AI

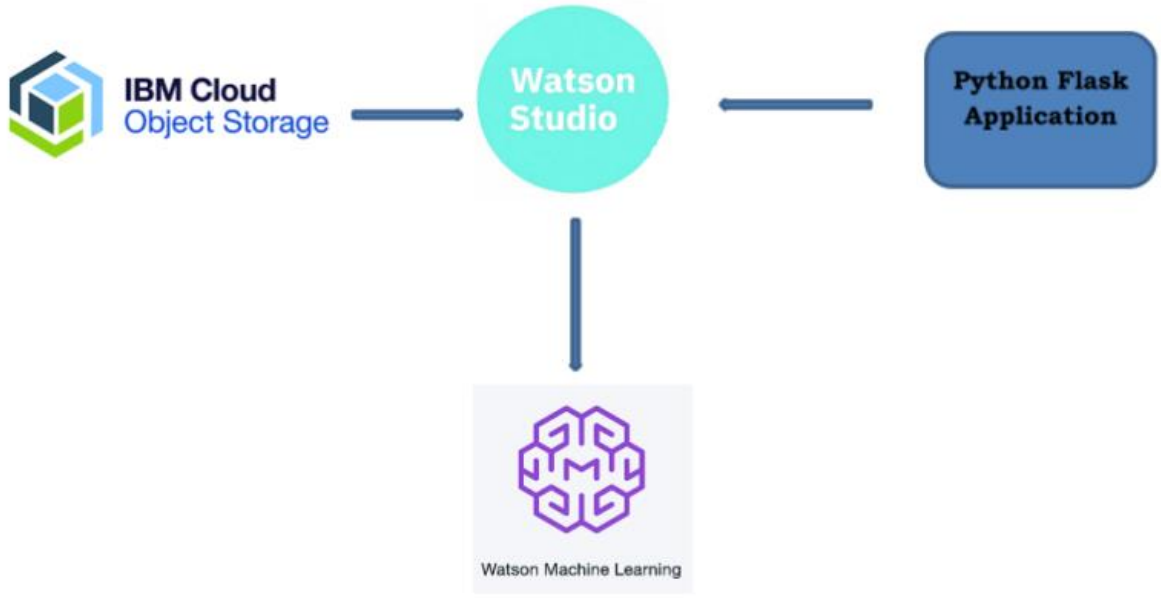
OVERVIEW

Churn prediction is one of the most popular Big Data use cases in business. It consists of detecting customers who are likely to cancel a subscription to a service. This can be telecom companies, SaaS companies, and any other company that sells a service for a monthly fee.
In the telecom industry, customers can choose from multiple service providers and actively switch from one operator to another. In this highly competitive market, the telecommunications industry experiences an average of 15-25% annual churn rate. Given the fact that it costs 5-10 times more to acquire a new customer than to retain an existing one, customer retention has now become even more important than customer acquisition.
For many incumbent operators, retaining high profitable customers is the number one business goal.

PURPOSE

Customer churn prediction can help you see which customers are about to leave your service so you can develop proper strategy to re-engage them before it is too late. This is a vital tool in a business' arsenal when it comes to customer retention.
Having the ability to accurately predict future churn rates is essential because it helps your business gain a better understanding of future expected revenue. Predicting churn rates can also help your business identify and improve upon areas where customer service is lacking.
To reduce customer churn, telecom companies need to predict which customers are at high risk of churn.
In this project, the customer-level data of a leading telecom firm, build predictive models to identify customers who will stay in the company (or) who will leave the company based on a set of parameters.

THEORITICAL ANALYSIS



The block diagram depicts the workflow of the entire system.

Watson Studio acts the central point of computation, and is used for running python notebooks and creating, monitoring, and managing deployments.

The runtime environment is powered by Watson Machine Learning Service. The UI is designed using HTML and the backend process is automated using Flask framework, which also facilitates deployment of the ML models using the scoring endpoint

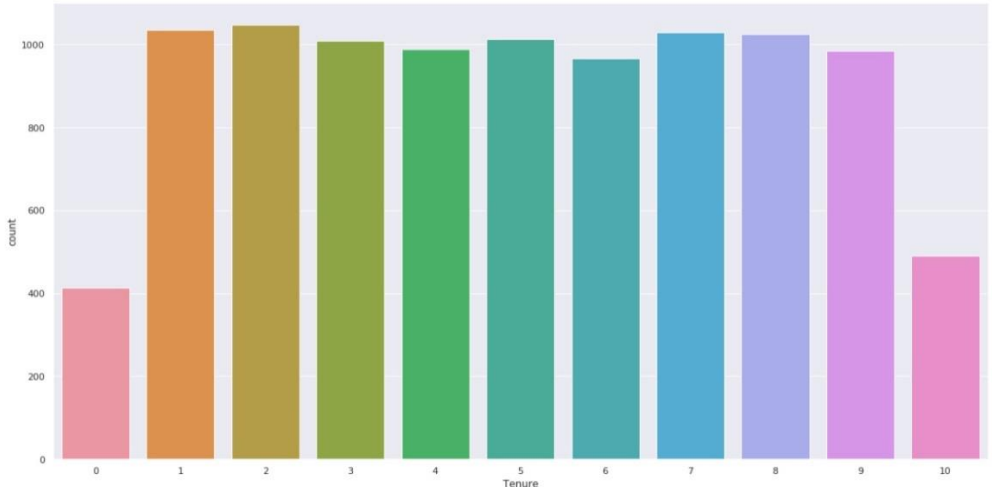
The software specification for the proposed system is as follows:

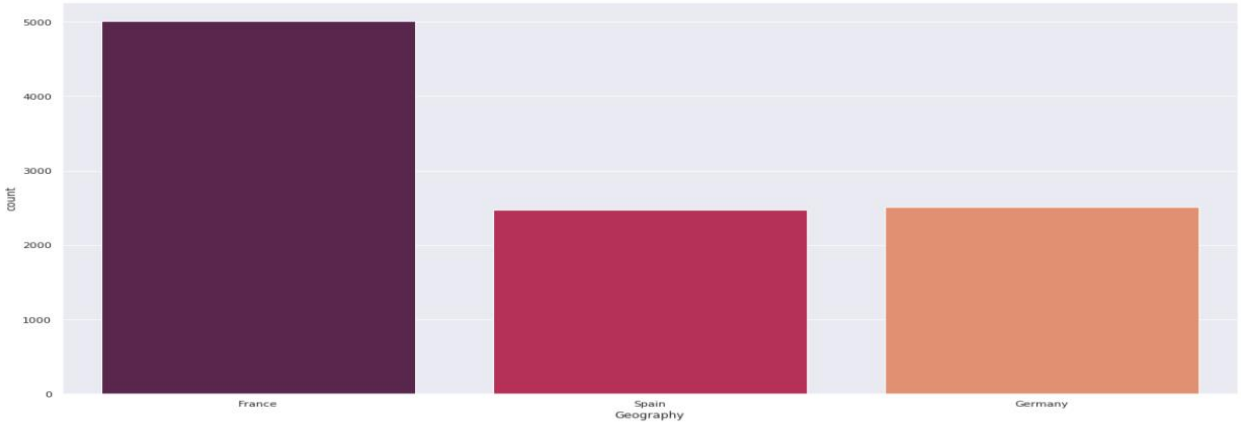
IBM Watson Studio:
Watson ML Package - 'Lite'
Instance Type - 'v2'
Environment Definition - Default Python 3.6XS
Virtual Hardware Configuration - 2 vCPU 8GB RAM
COS Instance Region - 'London'

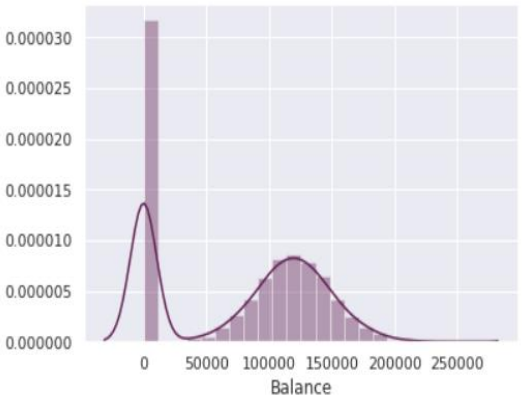
Python Flask Application:
HTML5
Flask

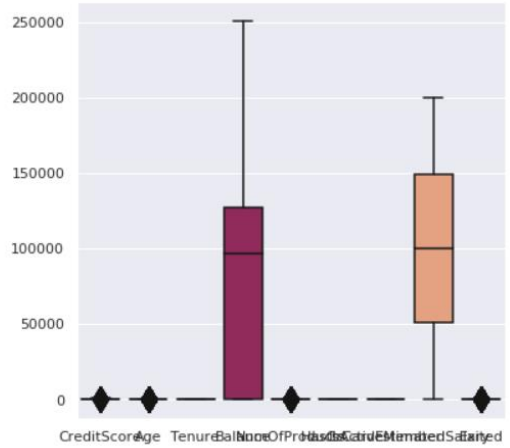
Python Libraries required:
scikit-learn
pandas
numpy
seaborn
json
sklearn.preprocessing
sklearn.model_selection
sklearn.feature_selection

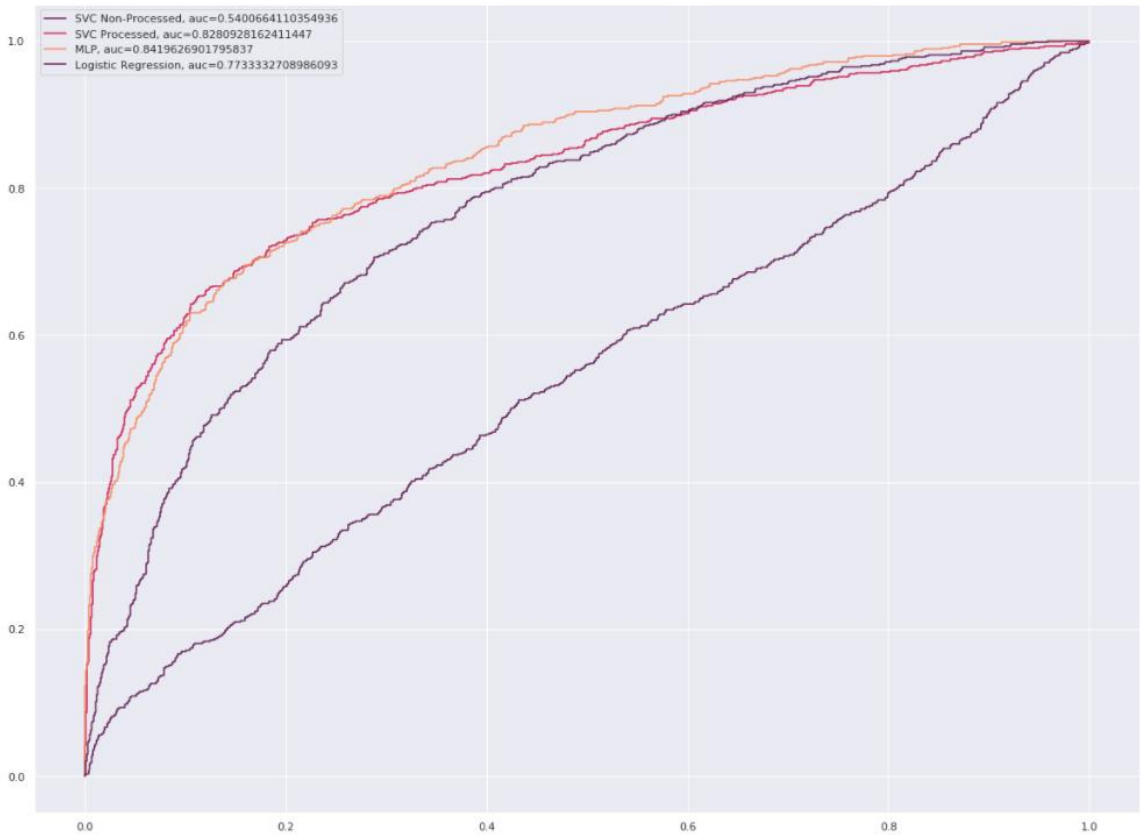
PLOTS

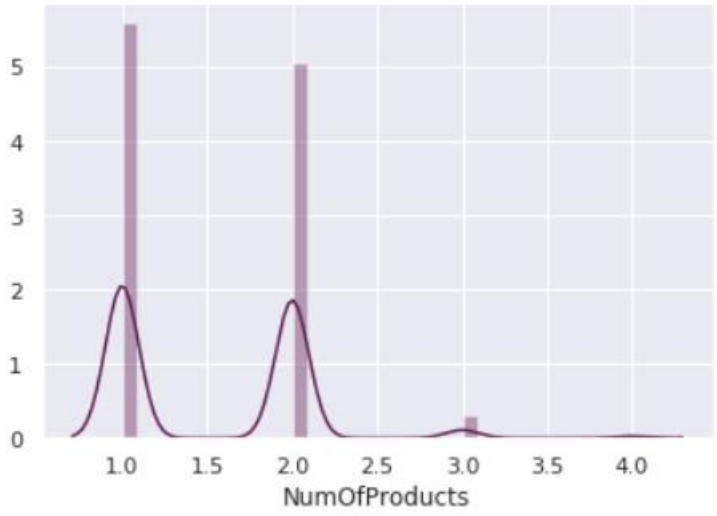

count
Tenure

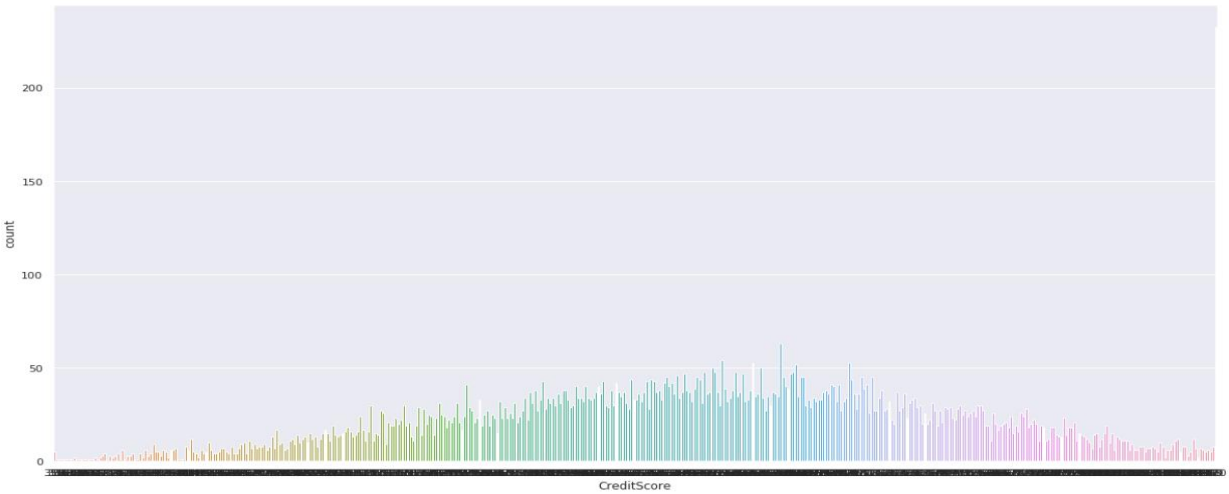

count
France
Spain
Germany
Geography

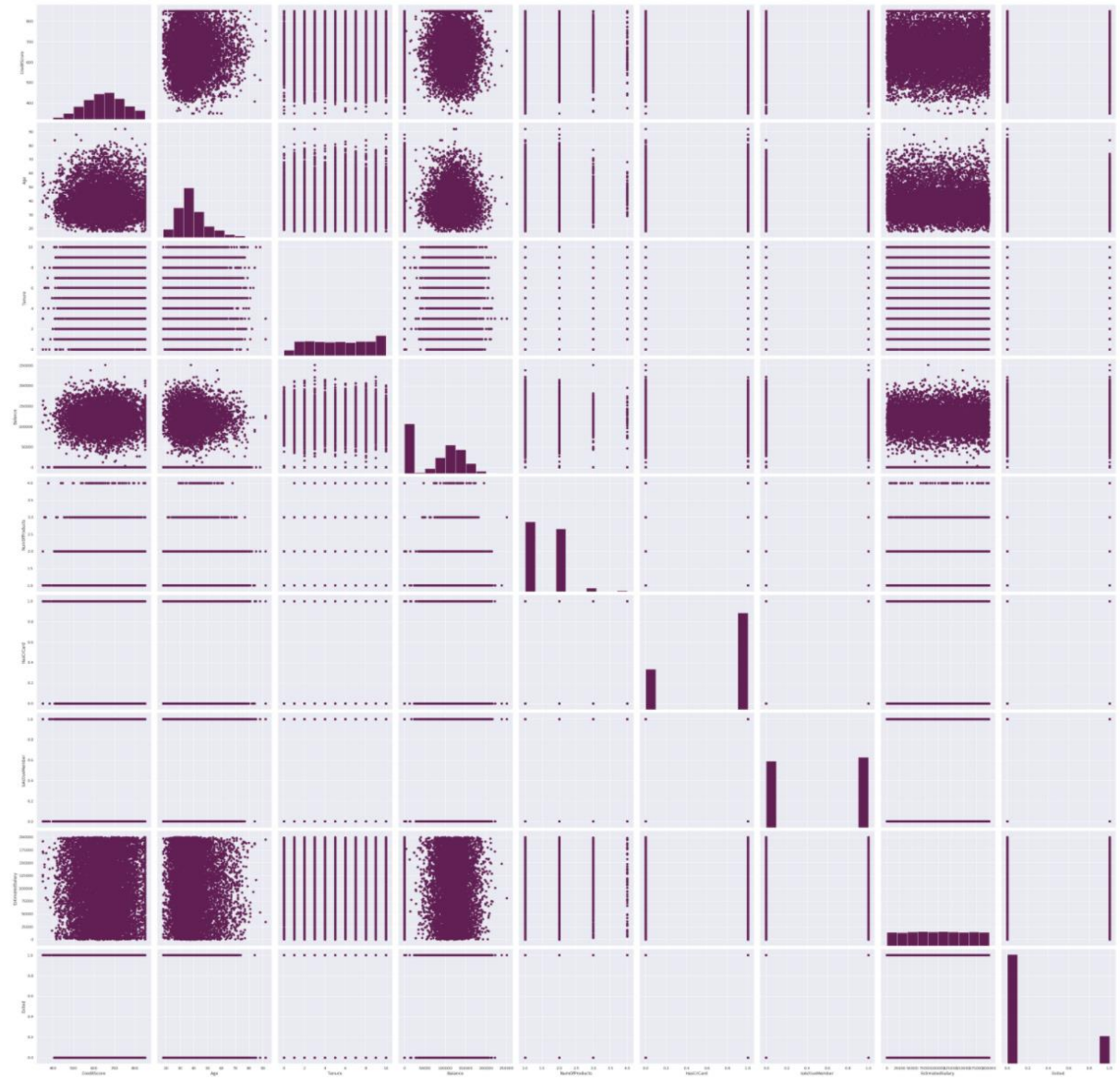

Balance



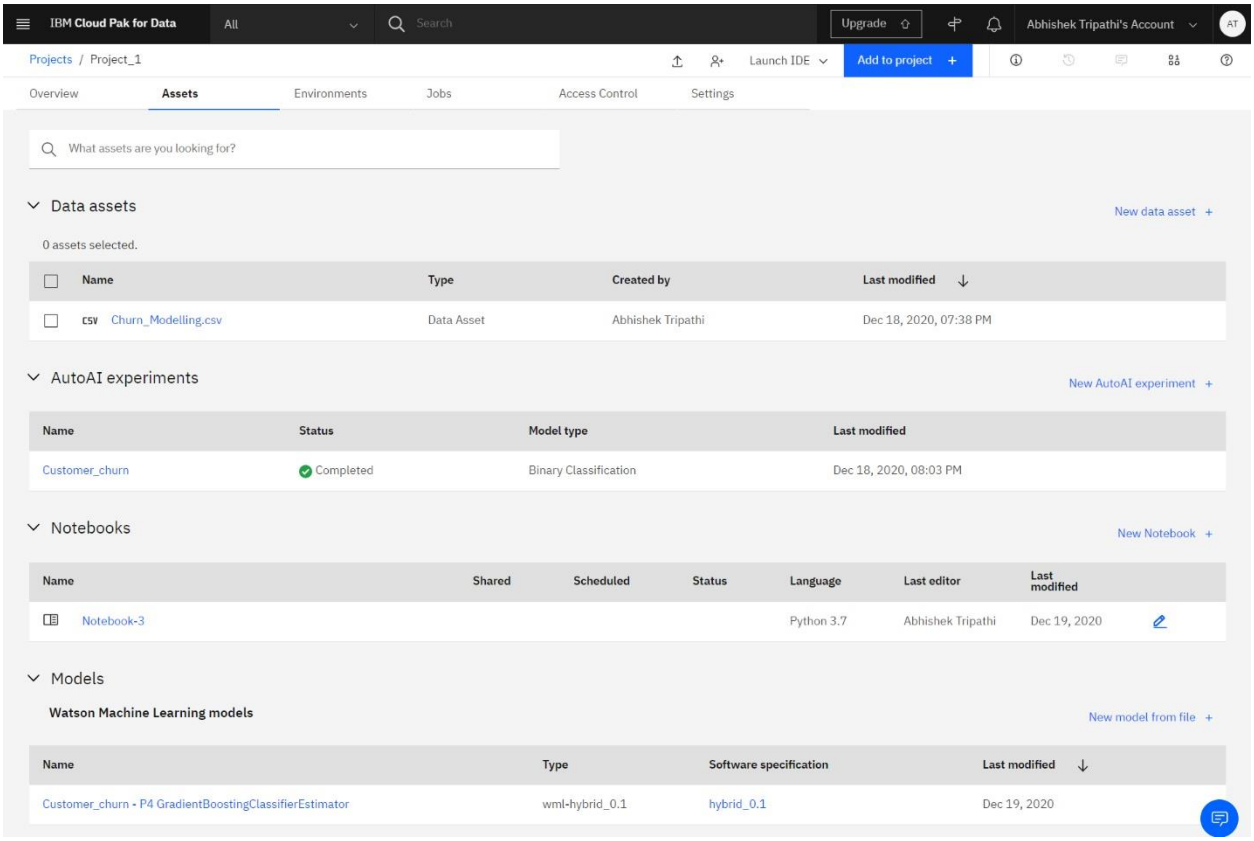

SVC Non-Processed, auc=0.5400664110354936
SVC Processed, auc=0.8280928162411447
MLP, auc=0.8419626901795837
Logistic Regression, auc=0.7733332708986093
1.0
0.8
0.6
0.4
0.2
0.0
0.0
0.2
0.4
0.6
0.8
1.0


5
4
3
2
1
0
1.0
1.5
2.0
2.5
3.0
3.5
4.0
NumOfProducts

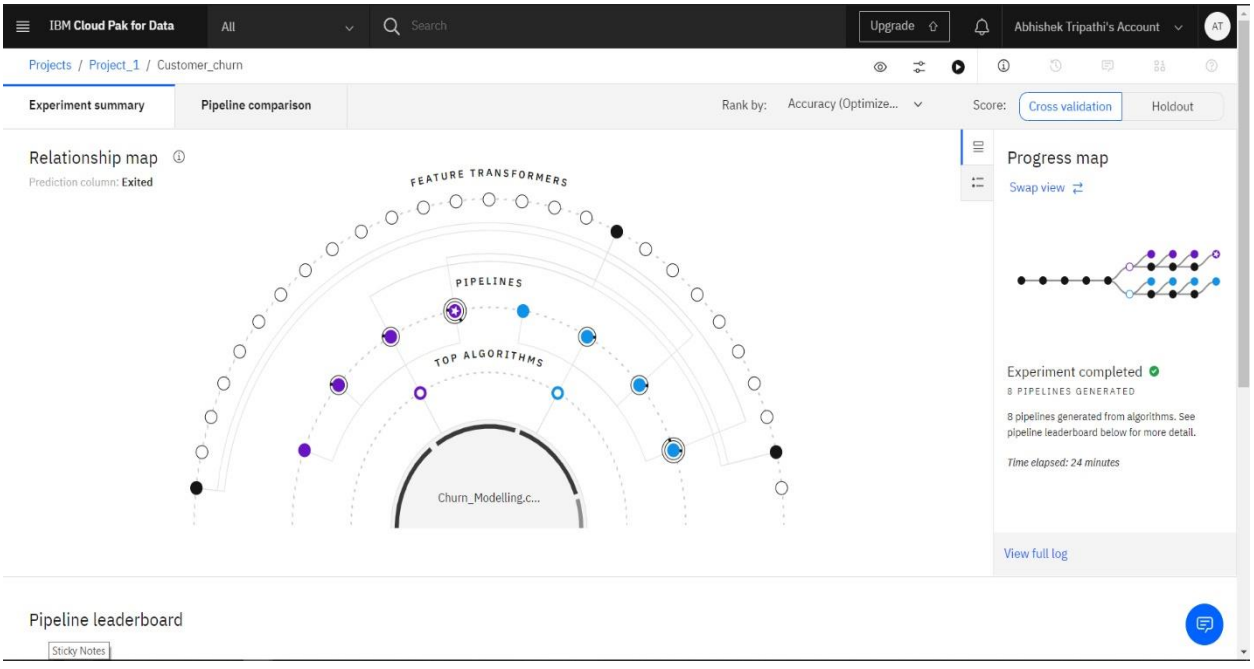

count
200
150
100
50
0
CreditScore



Creation of project and addition of assets –



Experiment summary –

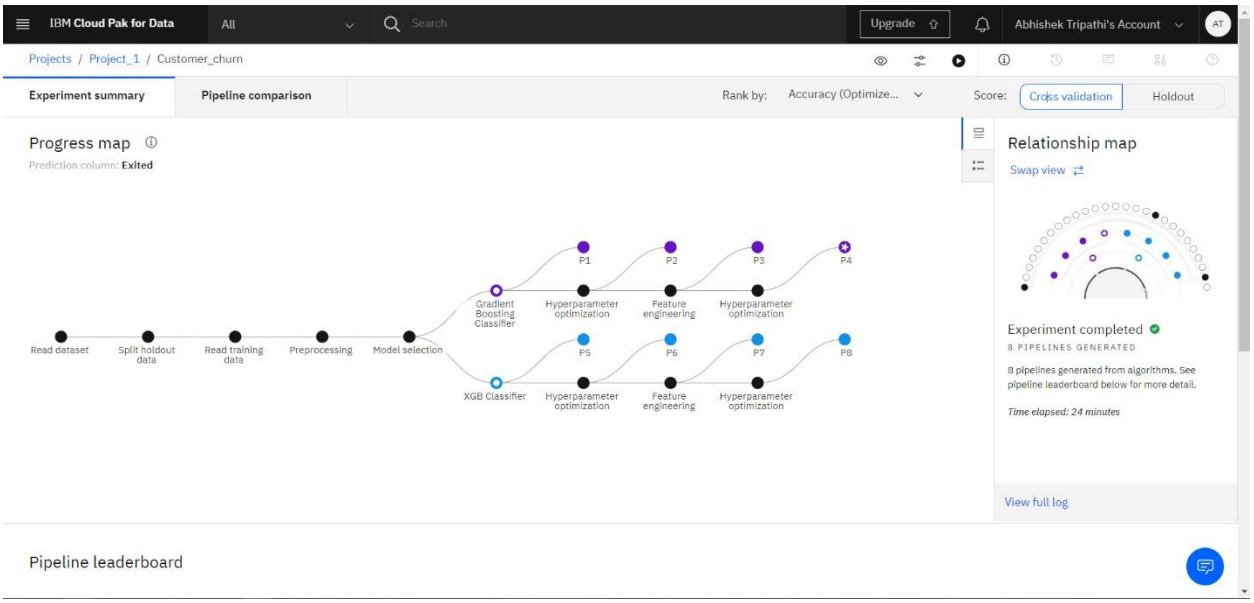


Pipeline leaderboard –

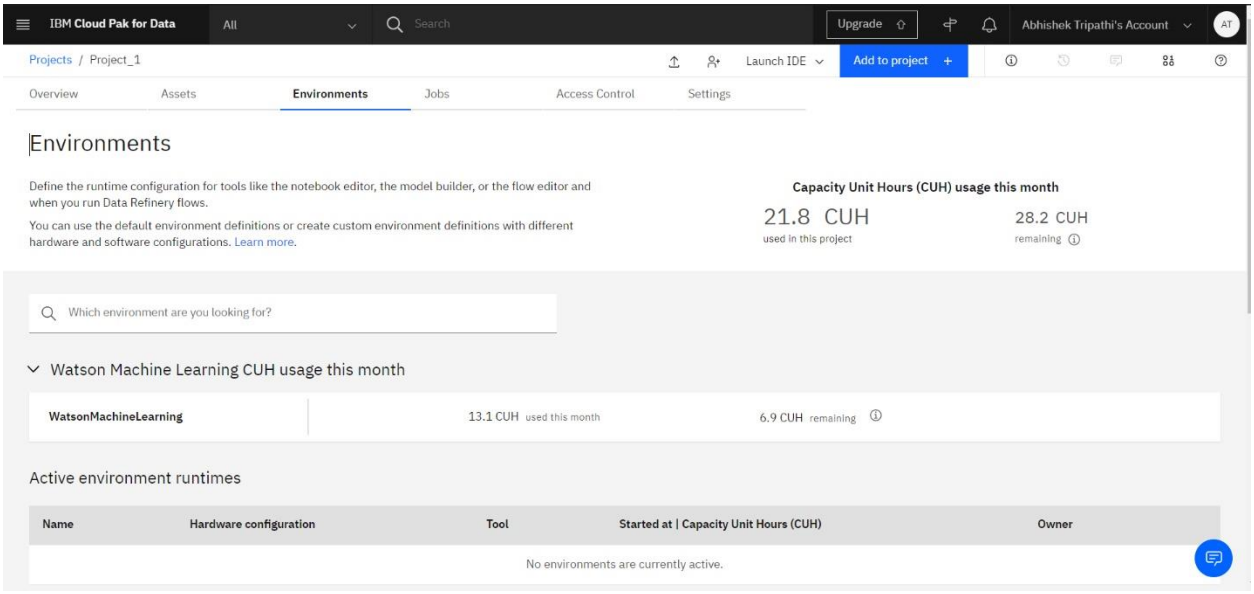
Pipeline leaderboard

Rank	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1	Pipeline 4	Gradient Boosting Classifier	0.866	HPO-1 FE HPO-2	00:05:32
2	Pipeline 3	Gradient Boosting Classifier	0.865	HPO-1 FE	00:09:00
3	Pipeline 1	Gradient Boosting Classifier	0.863	*None*	00:00:13
4	Pipeline 2	Gradient Boosting Classifier	0.863	HPO-1	00:00:52
5	Pipeline 7	XGB Classifier	0.855	HPO-1 FE	00:01:08
6	Pipeline 8	XGB Classifier	0.855	HPO-1 FE HPO-2	00:03:49
7	Pipeline 5	XGB Classifier	0.854	*None*	00:00:02
8	Pipeline 6	XGB Classifier	0.854	HPO-1	00:01:21

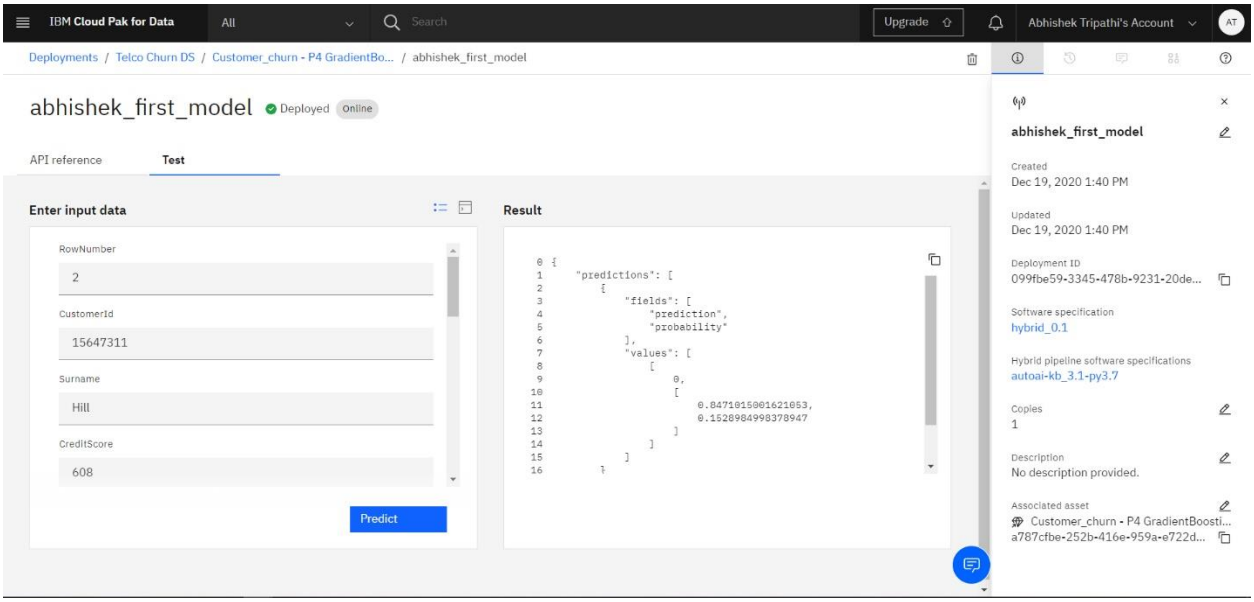
Progress Map-



Environments –



Testing Result –



FUTURE SCOPES

The project can be enhanced from different view-points namely:

- Optimized Machine Learning algorithms
- More feature engineering techniques
- Analyzing vital parameters for targeted customers
- Flask UI with improved functionalities
- Multiple deployments for different business scenarios

BIBILIOGRAPHY

1. *Essam Shaaban, Yehia Helmy, Ayman Khedr, Mona Nasr* | **International Journal of Engineering Research and Applications (IJERA)** | A Proposed Churn Prediction Model

2. *Sandra Mitrović, Bart Baesens, Wilfried Lemahieu, Jochen De Weerdt* | On the Operational Efficiency of Different Feature Types for Telco Churn Prediction

3. *Veronikha Effendy, Adiwijaya, Z.K.A. Baizal.* | **2014 2nd International Conference on Information and Communication Technology (ICoICT)** | Handling Imbalanced Data in Customer Churn Prediction Using Combined Sampling and Weighted Random Forest.

4. *Yiqing Huang, Fangzhou Zhu, Mingxuan Yuan, Ke Deng, Yanhua Li, Bing Ni, Wenyuan Dai, Qiang Yang, Jia Zeng* | **Advancing Computing as a Science & Profession** | Telco Churn Prediction with Big Data.