

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.

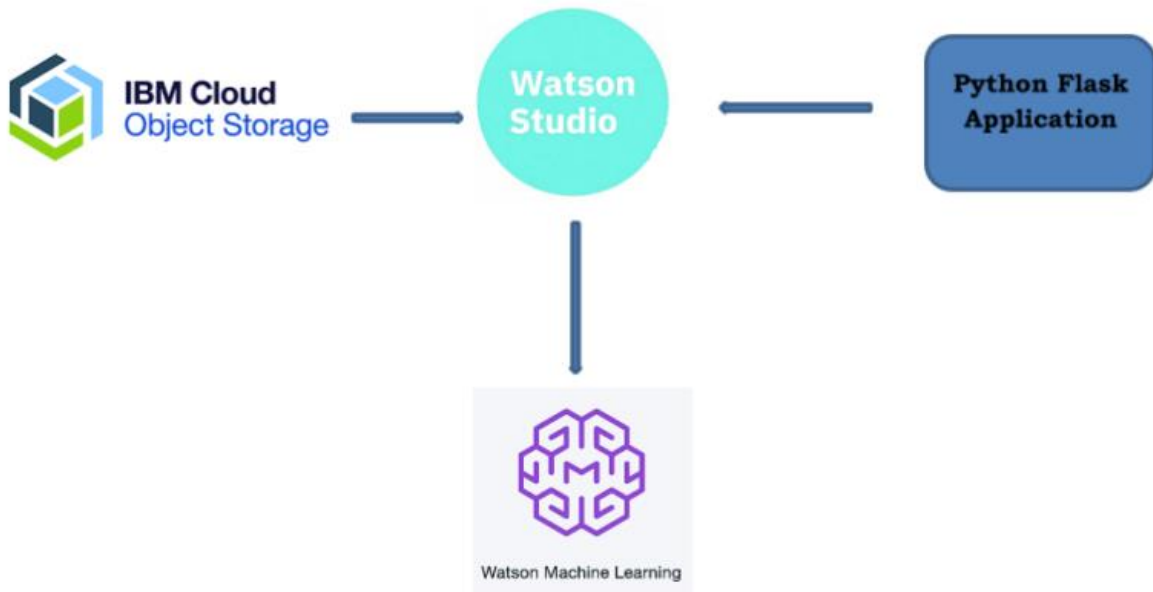
PURPOSE

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

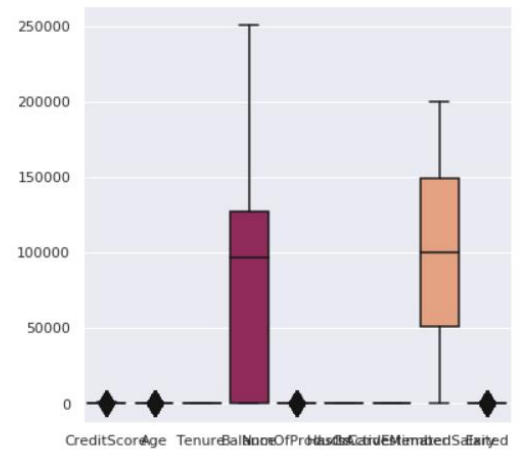
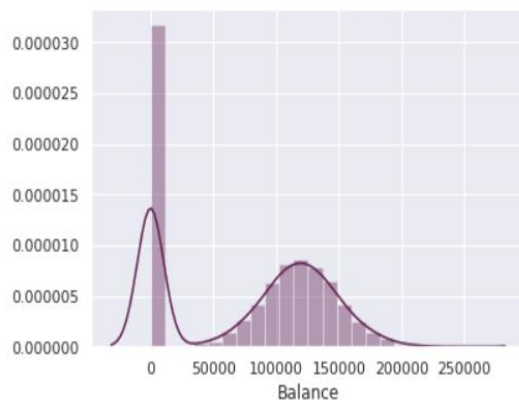
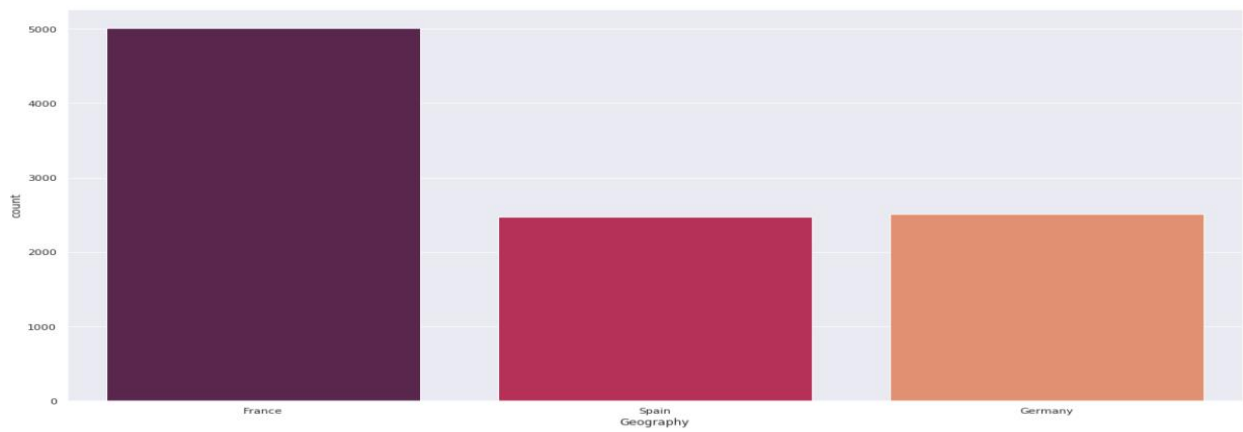
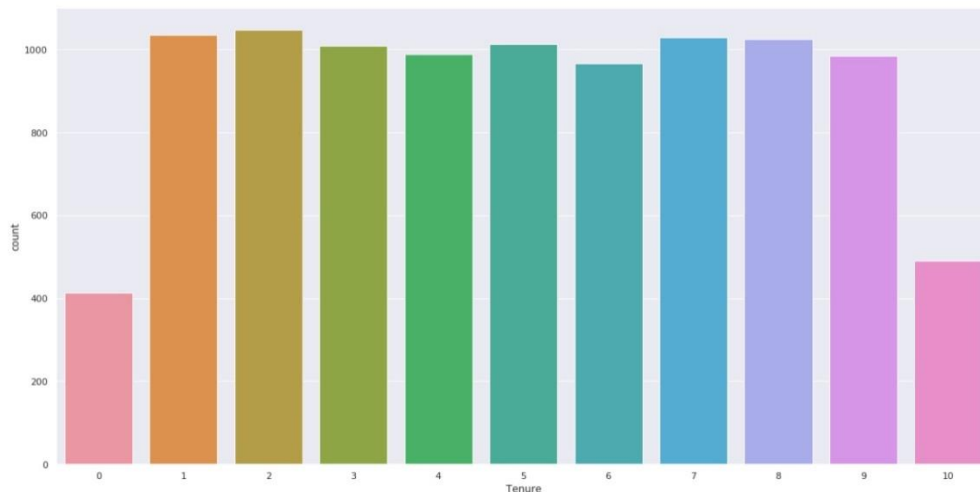
json

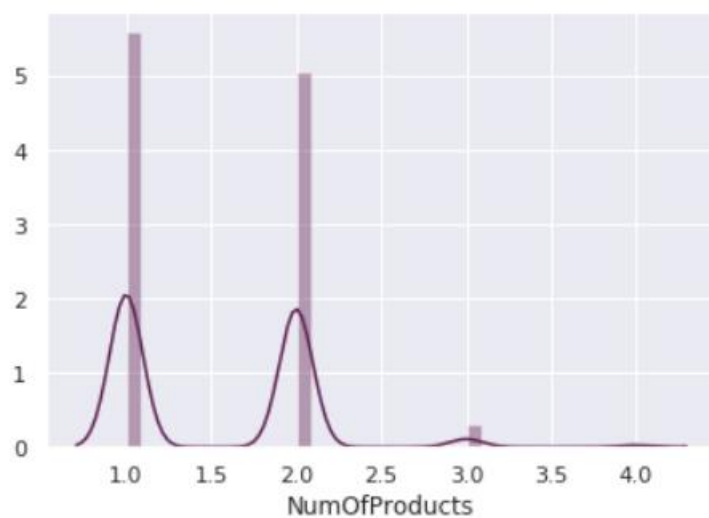
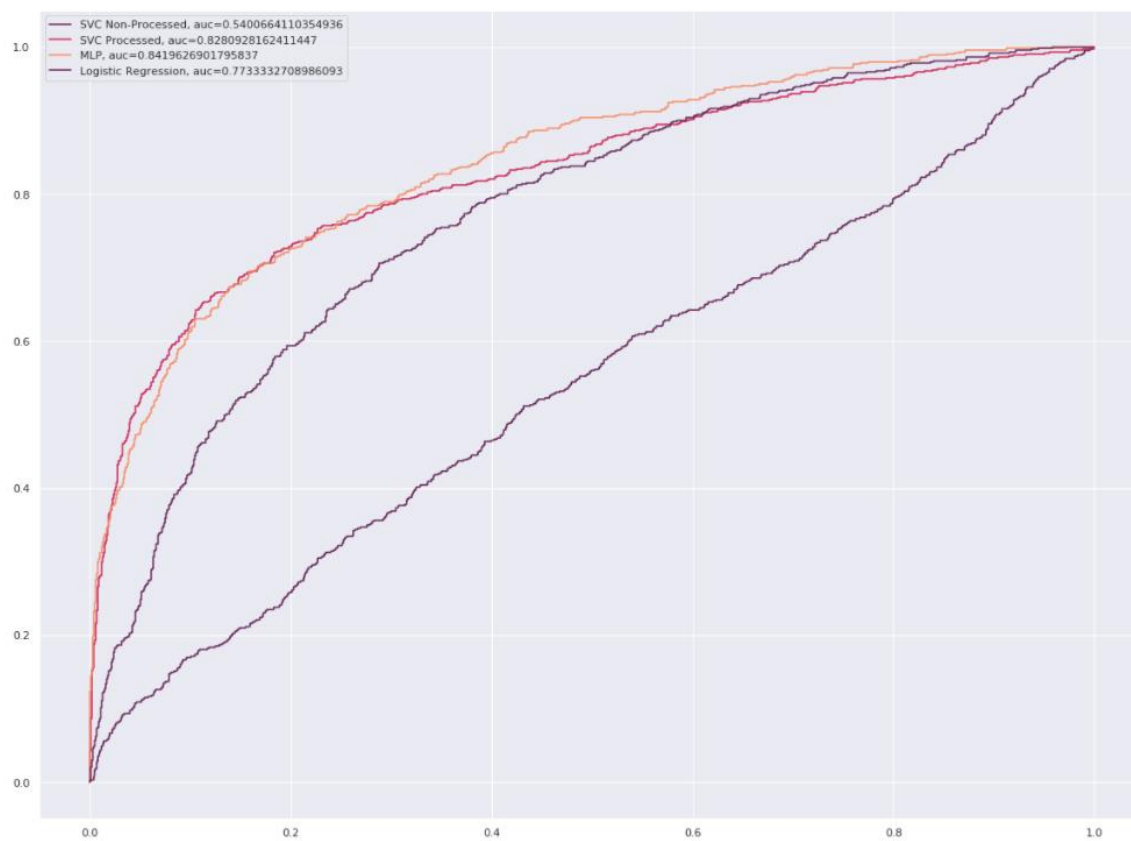
sklearn.preprocessing

sklearn.model_selection

sklearn.feature_selection

PLOTS





Creation of project and addition of assets –

IBM Cloud Pak for Data

All

Search

Upgrade

Abhishek Tripathi's Account

AT

Projects / Project_1

OverviewAssetsEnvironmentsJobsAccess ControlSettings

What assets are you looking for?

▼ Data assets

New data asset +

0 assets selected.

<input type="checkbox"/>	Name	Type	Created by	Last modified	↓
<input type="checkbox"/>	CSV Churn_Modelling.csv	Data Asset	Abhishek Tripathi	Dec 18, 2020, 07:38 PM	

▼ AutoAI experiments

New AutoAI experiment +

Name	Status	Model type	Last modified
Customer_churn	Completed	Binary Classification	Dec 18, 2020, 08:03 PM

▼ Notebooks

New Notebook +

Name	Shared	Scheduled	Status	Language	Last editor	Last modified	
Notebook-3				Python 3.7	Abhishek Tripathi	Dec 19, 2020	

▼ Models

Watson Machine Learning models

New model from file +

Name	Type	Software specification	Last modified	↓
Customer_churn - P4 GradientBoostingClassifierEstimator	wml-hybrid_0.1	hybrid_0.1	Dec 19, 2020	

Experiment summary –

IBM Cloud Pak for Data

All

Search

Upgrade

Abhishek Tripathi's Account

AT

Projects / Project_1 / Customer_churn

Experiment summaryPipeline comparisonRank by: Accuracy (Optimize...Score: Cross validationHoldout

Relationship map

Prediction column: Exited

Progress map

Swap view

Experiment completed

8 PIPELINES GENERATED

8 pipelines generated from algorithms. See pipeline leaderboard below for more detail.

Time elapsed: 24 minutes

View full log

Pipeline leaderboard

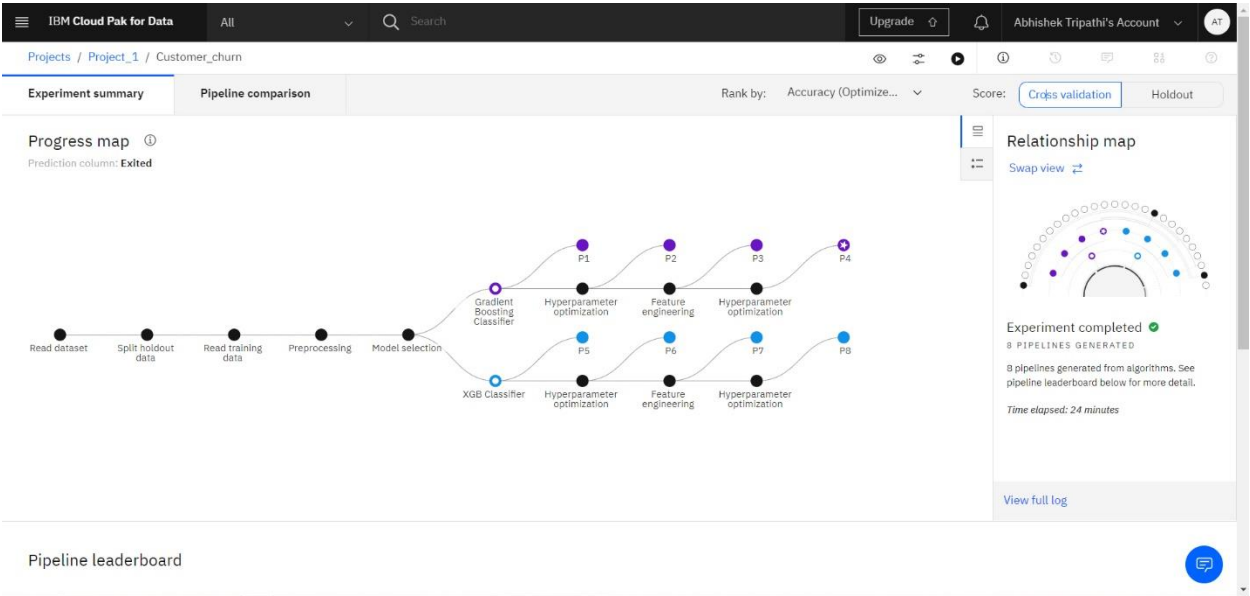
Sticky Notes

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 —

IBM Cloud Pak for Data

All

Search

Upgrade

Abhishek Tripathi's Account

AT

Projects / Project_1

Launch IDE

Add to project

Info

Refresh

Help

Settings

Overview

Assets

Environments

Jobs

Access Control

Settings

Environments

Define the runtime configuration for tools like the notebook editor, the model builder, or the flow editor and when you run Data Refinery flows.

You can use the default environment definitions or create custom environment definitions with different hardware and software configurations. [Learn more.](#)

Capacity Unit Hours (CUH) usage this month

21.8 CUH
used in this project

28.2 CUH
remaining

Which environment are you looking for?

▼ Watson Machine Learning CUH usage this month

WatsonMachineLearning

13.1 CUH used this month

6.9 CUH remaining

Active environment runtimes

Name	Hardware configuration	Tool	Started at Capacity Unit Hours (CUH)	Owner
No environments are currently active.				

Testing Result —

IBM Cloud Pak for Data

All

Search

Upgrade

Abhishek Tripathi's Account

AT

Deployments / Telco Churn DS / Customer_churn - P4 GradientBo...

abhishek_first_model

Info

Refresh

Help

Settings

API reference

Test

Enter input data

Result

RowNumber

2

CustomerId

15647311

Surname

Hill

CreditScore

608

Predict

```
0 {
1   "predictions": [
2     {
3       "fields": [
4         "prediction",
5         "probability"
6       ],
7       "values": [
8         [
9           0,
10          [
11            0.8471915901621053,
12            0.1528984998378947
13          ]
14        ]
15      ]
16    }
17  ]
18 }
```

abhishek_first_model

Created

Dec 19, 2020 1:40 PM

Updated

Dec 19, 2020 1:40 PM

Deployment ID

099f9e59-3345-476b-9231-20de...

Software specification

hybrid_0.1

Hybrid pipeline software specifications

autoai-kb_3.1-py3.7

Copies

1

Description

No description provided.

Associated asset

Customer_churn - P4 GradientBoosti...

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

BIBLIOGRAPHY

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 Weerd* | 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, Wenyan Dai, Qiang Yang, Jia Zeng* | **Advancing Computing as a Science & Profession** | Telco Churn Prediction with Big Data.