

Telecom Customer Churn Prediction Using Watson Auto AI

Introduction:

■ Overview:

The project is about telecom customer churn prediction. The main idea of this project is to use Watson Auto AI to predict the customer churn. This project falls under the category of Machine Learning as we are training and testing a model to predict the output depending on certain inputs given.

■ Purpose:

The purpose of this project is to create a model that predicts and identifies those customers who are most likely to leave a company.

Literature Survey:

■ Existing Problem:

Customer churn is a major problem and one of the most important concerns for large industries. Telecommunication industry always suffers from very high churn rates when one industry offers a better plan than the previous, there is a high possibility of customer churning from the present due to a better plan in such a scenario it is very difficult to avoid losses due to the direct effect on the revenues of the companies, especially in the telecom field.

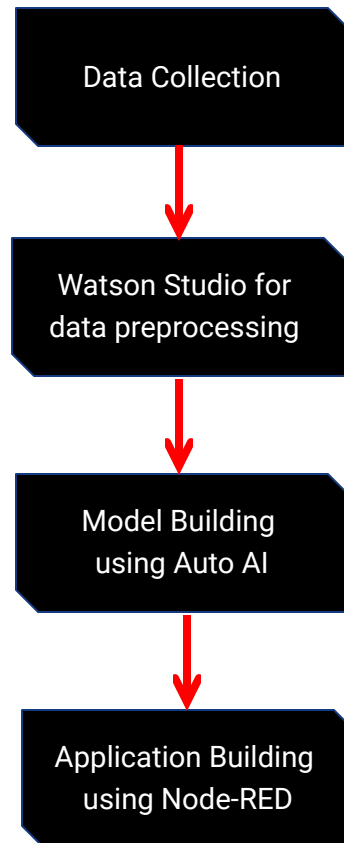
■ Proposed Solution:

We are building a Machine Learning model to predict the customer churn using IBM Watson Auto AI Machine Learning Service. The model is deployed on IBM cloud to get scoring end point which can be used as API in mobile app or web app building. We are developing a web application which is built using node red service. We make use of the scoring end point to give user input values to the deployed model. The model prediction is then showcased on User Interface.

Theoretical Analysis:

■ **Block Diagram:**

The block diagram and steps for building a model that determines the churn prediction are as follows -



■ **Hardware/Software Designing:**

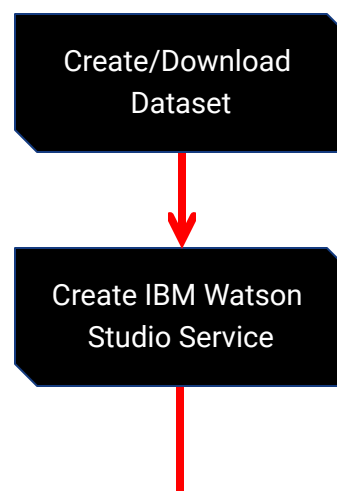
- IBM cloud
- IBM Watson Studio
- Machine Learning Service
- Auto AI Experiment
- UI using Node-Red

Experimental Investigations:

We decipher the following points based on the analysis of the provided dataset -

- The dataset falls under the category of supervised learning as inputs are given and output is to be predicted.
- The output of the dataset is not continuous but is discrete in nature hence, it falls under classification algorithm.
- The output is considered as dependent variable and all the input columns are considered as independent variables.
- The output of the dataset is of binary classification algorithm and hence accuracy is determined.
- Node-RED service is used to create a user interface.

Flowchart:



```
graph TD; A[Create Machine Learning Service] --> B[Create Watson Studio Project]; B --> C[Import Dataset]; C --> D[Create Auto AI Experiment]; D --> E[Select Pipeline]; E --> F[Deploy and Test Model]; F --> G[Build UI using Node-RED];
```

Create Machine
Learning Service

Create Watson
Studio Project

Import Dataset

Create Auto AI
Experiment

Select Pipeline

Deploy and Test
Model

Build UI using
Node-RED

Result:

A machine learning model using IBM Auto AI and Node-RED user interface is successfully built for accurately calculating churn predictions.

Advantages:

- It is useful for making predictions related to customer churn in advance.
- Loses can be kept to a minimal level by creating a machine learning model.
- It helps reduce the direct affect on revenue by giving prior notice.
- The model has high accuracy.
- It can be used for large datasets.

Disadvantages:

- Model might be slow due to network glitch.
- May receive errors if the Node-RED model coding is not done properly.
- Run-time might be long for large datasets.

Applications:

- Predict customers' chance of leaving a phone subscription plan.
- The customer churn rates in telecom industry are highly increasing, the model helps predict if a person is going to change his plan or not before the person actually decides in order to prevent loses.
- It is used in telecome industries.
- It helps the telecom industry to create better plans in future.

Conclusion:

A machine learning model to predict telecom customer churn using Auto AI has been created to predict if a person is likely to leave a service or not.

Future Scope:

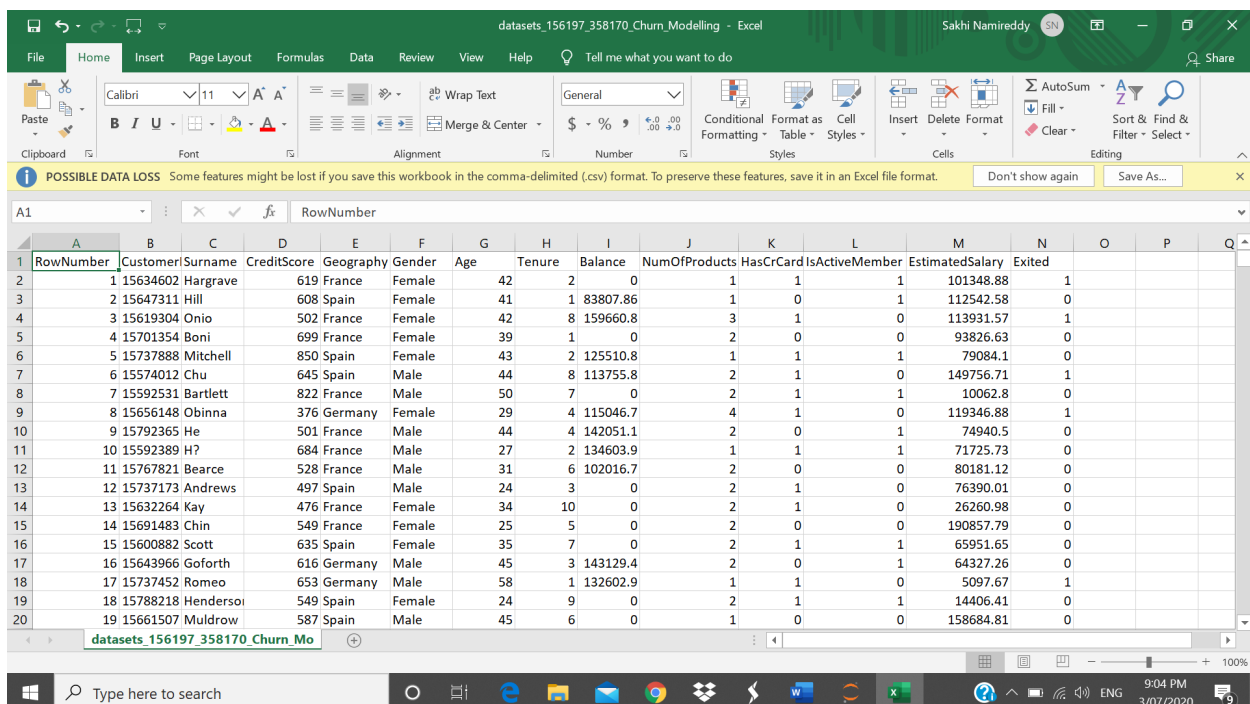
Telecom customer churn has been increasing rapidly thereby, causing losses to certain telecom industries. If a company offers better subscription plan then customer churning takes place. It is important to know if a person is likely to leave a service or not in order to minimize the losses caused to a company. This model will be helpful in predicting the output.

Bibliography:

- <https://smartbridge.teachable.com/>
- <https://smartinternz.com/>
- <https://cloud.ibm.com/>
- <https://node-red-bsqzs.eu-gb.mybluemix.net/>
- <https://www.kaggle.com/>

Appendix:

■ Data Collection:



RowNumber	CustomerID	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
1	15634602	Hargrave	619	France	Female	42	2	0	1	1	1	101348.88	1
2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	0
3	15619304	Onio	502	France	Female	42	8	159660.8	3	1	0	113931.57	1
4	15701354	Boni	699	France	Female	39	1	0	2	0	0	93826.63	0
5	15737888	Mitchell	850	Spain	Female	43	2	125510.8	1	1	1	79084.1	0
6	15574012	Chu	645	Spain	Male	44	8	113755.8	2	1	0	149756.71	1
7	15592531	Bartlett	822	France	Male	50	7	0	2	1	1	10062.8	0
8	15656148	Obinna	376	Germany	Female	29	4	115046.7	4	1	0	119346.88	1
9	15792365	He	501	France	Male	44	4	142051.1	2	0	1	74940.5	0
10	15592389	H7	684	France	Male	27	2	134603.9	1	1	1	71725.73	0
11	15767821	Bearce	528	France	Male	31	6	102016.7	2	0	0	80181.12	0
12	15737173	Andrews	497	Spain	Male	24	3	0	2	1	0	76390.01	0
13	15632264	Kay	476	France	Female	34	10	0	2	1	0	26260.98	0
14	15691483	Chin	549	France	Female	25	5	0	2	0	0	190857.79	0
15	15600882	Scott	635	Spain	Female	35	7	0	2	1	1	65951.65	0
16	15643966	Goforth	616	Germany	Male	45	3	143129.4	2	0	1	64327.26	0
17	15737452	Romeo	653	Germany	Male	58	1	132602.9	1	1	0	5097.67	1
18	15788218	Henderso	549	Spain	Female	24	9	0	2	1	1	14406.41	0
19	15661507	Muldro	587	Spain	Male	45	6	0	1	0	0	158684.81	0

■ Importing Dataset and Calculating Correlation:

IBM Cloud Node-RED : n Home Page - Telecom c x Student Dash IISPS_INT_317 Node-RED Dc New tab

localhost:8888/notebooks/Telecom%20customer%20churn%20prediction%20using%20Watson%20Auto%20AI.ipynb

jupyter Telecom customer churn prediction using Watson Auto AI Last Checkpoint: Last Wednesday at 2:54 PM (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [1]: import numpy as np
import pandas as pd
import matplotlib.pyplot as plt

In [2]: dataset=pd.read_csv('datasets_156197_358170_Churn_Modelling edited.csv')
dataset
```

Out[2]:

	RowNumber	CustomerId	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
0	1	15634602	619	France	Female	42	2	0.00	1	1	1	101348.88	
1	2	15647311	608	Spain	Female	41	1	83807.86	1	0	1	112542.58	
2	3	15619304	502	France	Female	42	8	159660.80	3	1	0	113931.57	
3	4	15701354	699	France	Female	39	1	0.00	2	0	0	93826.63	
4	5	15737888	850	Spain	Female	43	2	125510.82	1	1	1	79084.10	
...
9995	9996	15606229	771	France	Male	39	5	0.00	2	1	0	96270.64	
9996	9997	15569892	516	France	Male	35	10	57369.61	1	1	1	101699.77	
9997	9998	15584532	709	France	Female	36	7	0.00	1	0	1	42085.58	
9998	9999	15682355	772	Germany	Male	42	3	75075.31	2	1	0	92888.52	
9999	10000	15628319	792	France	Female	28	4	130142.79	1	1	0	38190.78	

10000 rows x 13 columns

Type here to search

9:05 PM 3/07/2020

IBM Cloud Node-RED : n Home Page - Telecom c x Student Dash IISPS_INT_317 Node-RED Dc New tab

localhost:8888/notebooks/Telecom%20customer%20churn%20prediction%20using%20Watson%20Auto%20AI.ipynb

jupyter Telecom customer churn prediction using Watson Auto AI Last Checkpoint: Last Wednesday at 2:54 PM (unsaved changes) Logout

File Edit View Insert Cell Kernel Widgets Help Trusted Python 3

```
In [3]: dataset.corr()
```

Out[3]:

	RowNumber	CustomerId	CreditScore	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary	Exited
RowNumber	1.000000	0.004202	0.005840	0.000783	-0.006495	-0.009067	0.007246	0.000599	0.012044	-0.005988	-0.016571
CustomerId	0.004202	1.000000	0.005308	0.009497	-0.014883	-0.012419	0.016972	-0.014025	0.001665	0.015271	-0.006248
CreditScore	0.005840	0.005308	1.000000	-0.003965	0.000842	0.006268	0.012238	-0.005458	0.025651	-0.001384	-0.027094
Age	0.000783	0.009497	-0.003965	1.000000	-0.009997	0.028308	-0.030680	-0.011721	0.085472	-0.007201	0.285323
Tenure	-0.006495	-0.014883	0.000842	-0.009997	1.000000	-0.012254	0.013444	0.022583	-0.028362	0.007784	-0.014007
Balance	-0.009067	-0.012419	0.006268	0.028308	-0.012254	1.000000	-0.304180	-0.014858	-0.010084	0.012797	0.118533
NumOfProducts	0.007246	0.016972	0.012238	-0.030680	0.013444	-0.304180	1.000000	0.003183	0.009612	0.014204	-0.047820
HasCrCard	0.000599	-0.014025	-0.005458	-0.011721	0.022583	-0.014858	0.003183	1.000000	-0.011866	-0.009933	-0.007138
IsActiveMember	0.012044	0.001665	0.025651	0.085472	-0.028362	-0.010084	0.009612	-0.011866	1.000000	-0.011421	-0.156128
EstimatedSalary	-0.005988	0.015271	-0.001384	-0.007201	0.007784	0.012797	0.014204	-0.009933	-0.011421	1.000000	0.012097
Exited	-0.016571	-0.006248	-0.027094	0.285323	-0.014007	0.118533	-0.047820	-0.007138	-0.156128	0.012097	1.000000

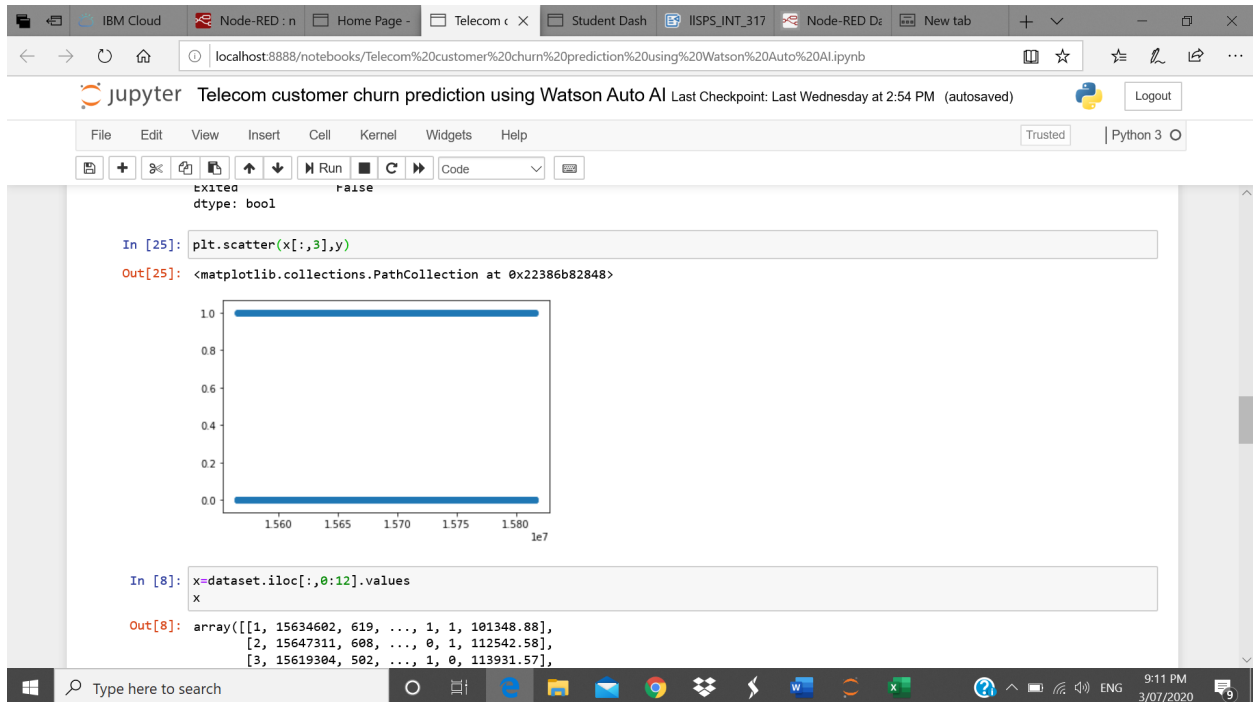
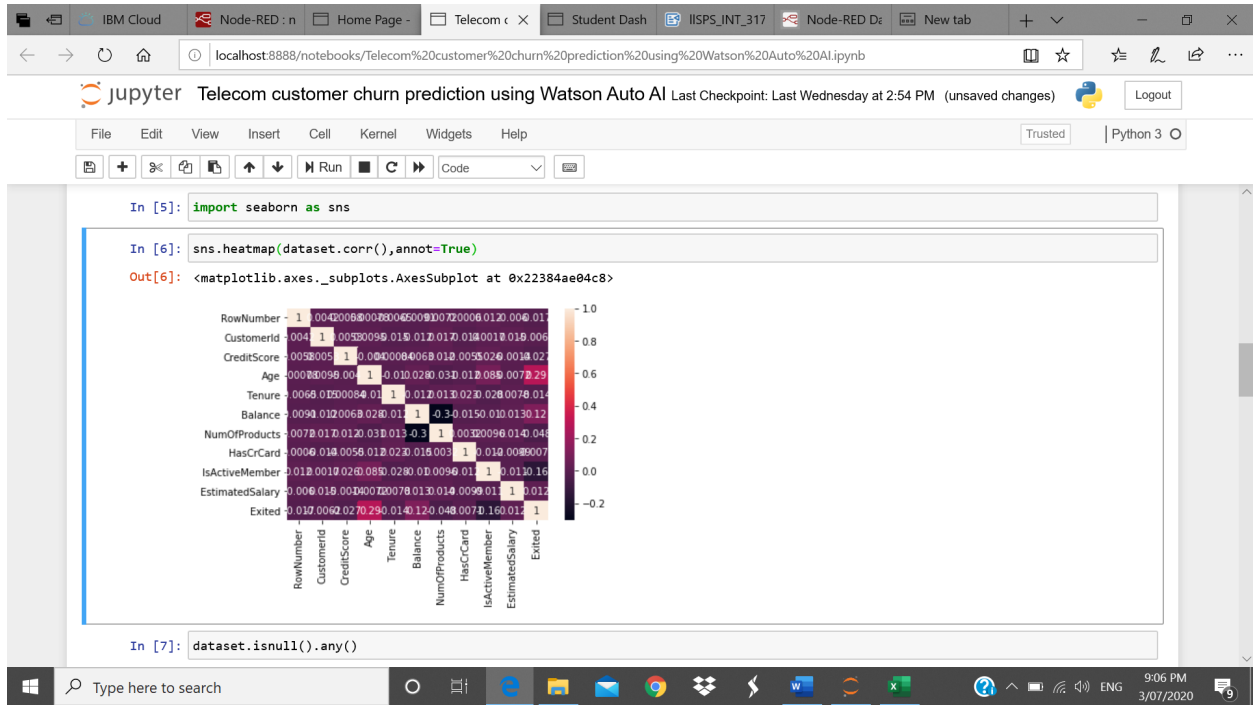
```
In [4]: dataset.info()

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 13 columns):
#   Column                Non-Null Count  Dtype
0   RowNumber              10000 non-null  int64
1   CustomerId             10000 non-null  int64
2   CreditScore             10000 non-null  int64
3   Geography               10000 non-null  object
4   Gender                  10000 non-null  object
5   Age                    10000 non-null  int64
6   Tenure                  10000 non-null  int64
7   Balance                 10000 non-null  float64
8   NumOfProducts           10000 non-null  int64
9   HasCrCard               10000 non-null  int64
10  IsActiveMember          10000 non-null  int64
11  EstimatedSalary          10000 non-null  float64
12  Exited                   10000 non-null  int64
```

Type here to search

9:05 PM 3/07/2020

■ Data visualization:



■ IBM Cloud Dashboard:

The screenshot shows the IBM Cloud Dashboard in a web browser. The browser's address bar displays <https://cloud.ibm.com/>. The dashboard header includes a search bar and navigation links for Catalog, Docs, Support, and Manage. The main content area is titled "Dashboard" and features a "Resource summary" section with a count of 11 resources. Below this, a list of resource categories is shown with their respective counts and status indicators. A "Planned maintenance" section is also visible, along with a "Recent support cases" section. The bottom of the dashboard includes a "For you" section with a "News" link. The Windows taskbar at the bottom shows the time as 9:13 PM on 3/07/2020.

Resource summary

Resource	Count	Status
Cloud Foundry apps	1	✓
Cloud Foundry services	1	
Services	4	✓
Storage	1	✓
Apps	1	
Developer tools	3	

Planned maintenance

Clear skies!
You can view your scheduled maintenance events here.

Recent support cases

For you

Get started with using AI and Cloud Object

News

IBM CEO Arvind Krishna Keynotes IBM Think Gov

■ Creating a Project using IBM Watson Studio:

The screenshot shows the IBM Watson Studio interface in a web browser. The browser's address bar displays https://dataplatform.cloud.ibm.com/home?context=wdp&apps=data_science_experience&nocache=true. The dashboard header includes a search bar and navigation links for Upgrade, Sakhi Namireddy's Account, and a user profile icon. The main content area is titled "Start by creating a project" and includes a "Create a project" section with instructions. Below this, a "Recently updated projects" section is shown with a table of projects. The Windows taskbar at the bottom shows the time as 9:17 PM on 3/07/2020.

Start by creating a project

A project is how you organize your resources to work with data and collaborate with team members.

Create a project

Create a project, and then add the tools and assets you need.

Recently updated projects

Name	Role	Collaborators	Date created	Last updated
Telecom customer churn prediction using Watson Auto AI	Admin	SN	Jul 02, 2020	Jul 02, 2020
Sample	Admin	SN	Jun 25, 2020	Jun 25, 2020

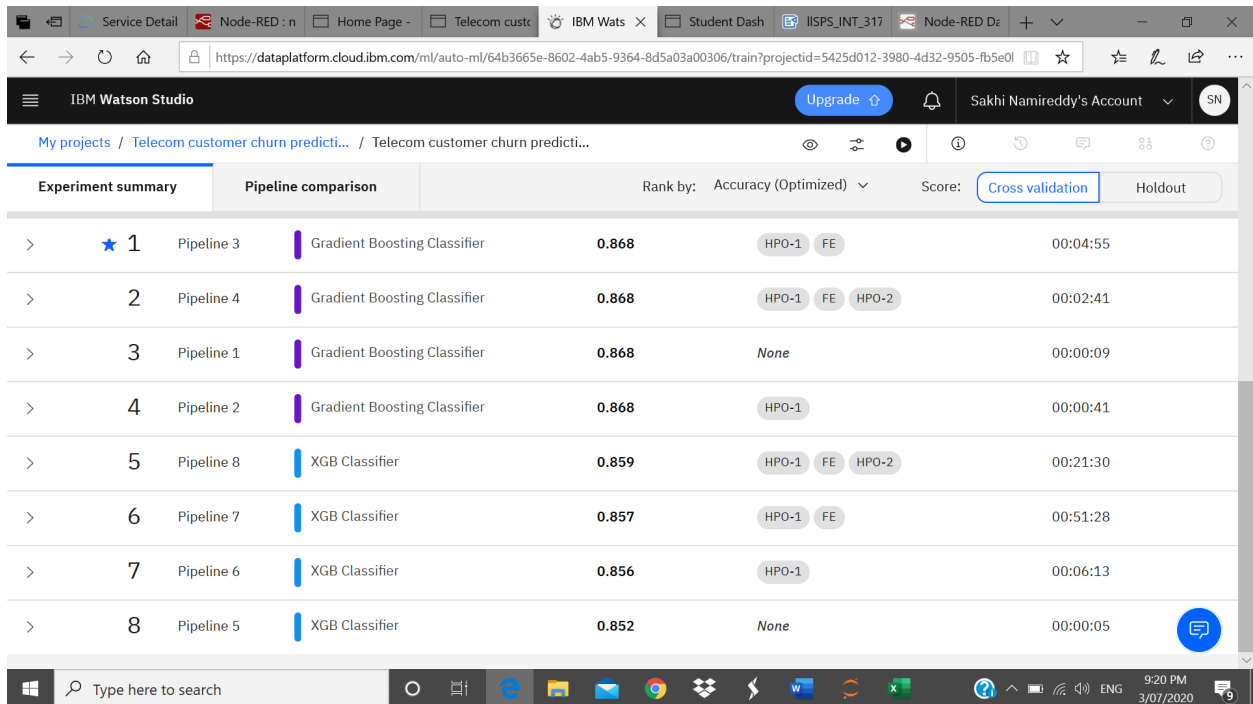
■ Importing Dataset and Creating Auto AI experiment:

The screenshot shows the IBM Watson Studio web interface. The top navigation bar includes the IBM Watson Studio logo, an 'Upgrade' button, and the user's account information (Sakhi Namireddy's Account). The main content area is divided into three sections: 'Data assets', 'AutoAI experiments', and 'Deep learning experiments'. The 'Data assets' section shows a table with one asset: 'datasets_156197_358170_Churn_Modelling.csv'. The 'AutoAI experiments' section shows a table with one experiment: 'Telecom customer churn prediction', which is 'Completed'. The 'Deep learning experiments' section is currently empty. On the right side, there is a 'Data' panel with tabs for 'Load', 'Files', and 'Catalog'. The 'Load' tab is active, showing a message: 'Drop files here or browse for files to upload.' The bottom of the screen shows a Windows taskbar with various application icons and the system clock indicating 9:18 PM on 3/07/2020.

■ Running the Auto AI Experiment:

The screenshot shows the IBM Watson Studio interface displaying the results of an AutoAI experiment. The top navigation bar is the same as the previous screenshot. The main content area is divided into two sections: 'Experiment summary' and 'Pipeline comparison'. The 'Experiment summary' section shows a 'Relationship map' with a 'Prediction column: Exited'. The 'Pipeline comparison' section shows a 'Rank by: Accuracy (Optimized)' and a 'Score: Cross validation'. The 'Relationship map' is a circular diagram showing the flow of data from 'datasets_156197_3...' through 'FEATURE TRANSFORMERS', 'PIPELINES', and 'TOP ALGORITHMS'. The 'Pipeline comparison' section shows a 'Progress map' with a 'Swap view' button. The 'Experiment completed' status is shown with a green checkmark. The text indicates that 8 pipelines were generated from algorithms, and the pipeline leaderboard is available for more detail. The 'Time elapsed: 91 minutes' is also displayed. The bottom of the screen shows a Windows taskbar with various application icons and the system clock indicating 9:19 PM on 3/07/2020.

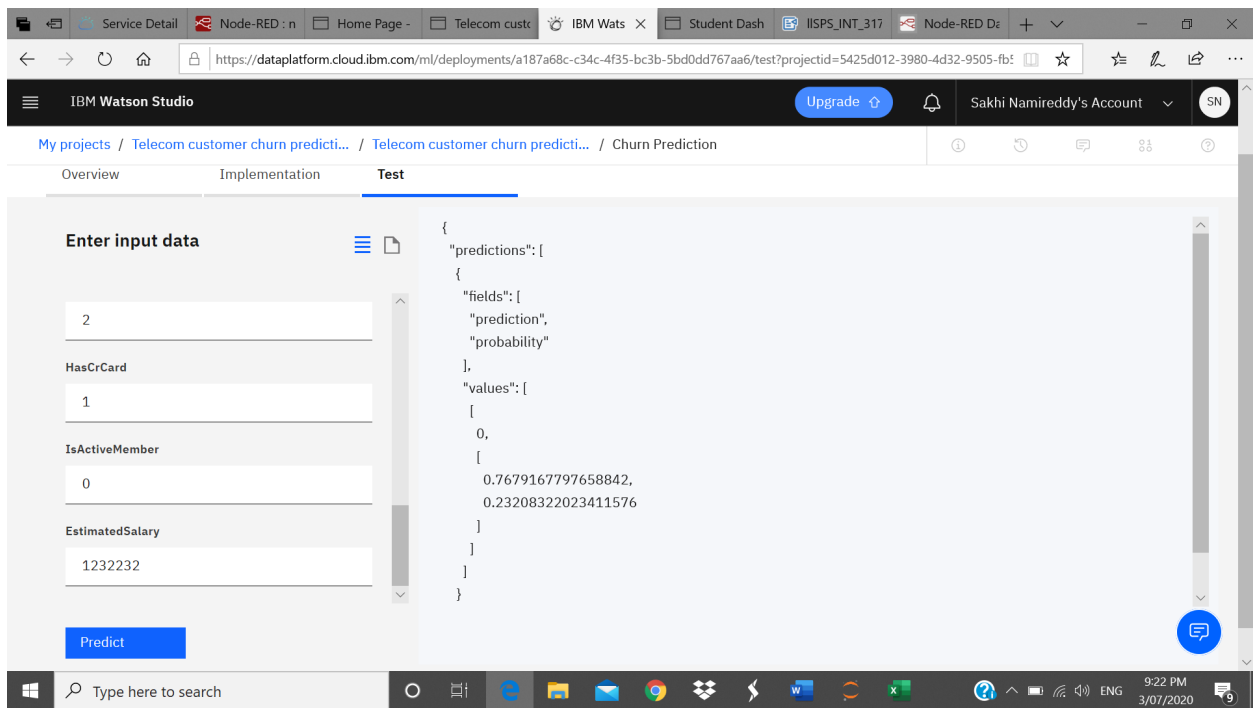
■ Selecting a Pipeline:



The screenshot shows the IBM Watson Studio interface with the 'Pipeline comparison' tab selected. The table lists 8 pipelines, ranked by Accuracy (Optimized). The top three pipelines (1, 2, and 3) all have an accuracy of 0.868 and are based on a Gradient Boosting Classifier. Pipeline 1 is the selected one. The bottom three pipelines (5, 6, and 7) are based on an XGB Classifier with accuracies of 0.859, 0.857, and 0.856 respectively. Pipeline 8 is also a Gradient Boosting Classifier with an accuracy of 0.852.

Rank	Pipeline	Model	Accuracy	Features	Time
1	Pipeline 3	Gradient Boosting Classifier	0.868	HPO-1, FE	00:04:55
2	Pipeline 4	Gradient Boosting Classifier	0.868	HPO-1, FE, HPO-2	00:02:41
3	Pipeline 1	Gradient Boosting Classifier	0.868	None	00:00:09
4	Pipeline 2	Gradient Boosting Classifier	0.868	HPO-1	00:00:41
5	Pipeline 8	XGB Classifier	0.859	HPO-1, FE, HPO-2	00:21:30
6	Pipeline 7	XGB Classifier	0.857	HPO-1, FE	00:51:28
7	Pipeline 6	XGB Classifier	0.856	HPO-1	00:06:13
8	Pipeline 5	XGB Classifier	0.852	None	00:00:05

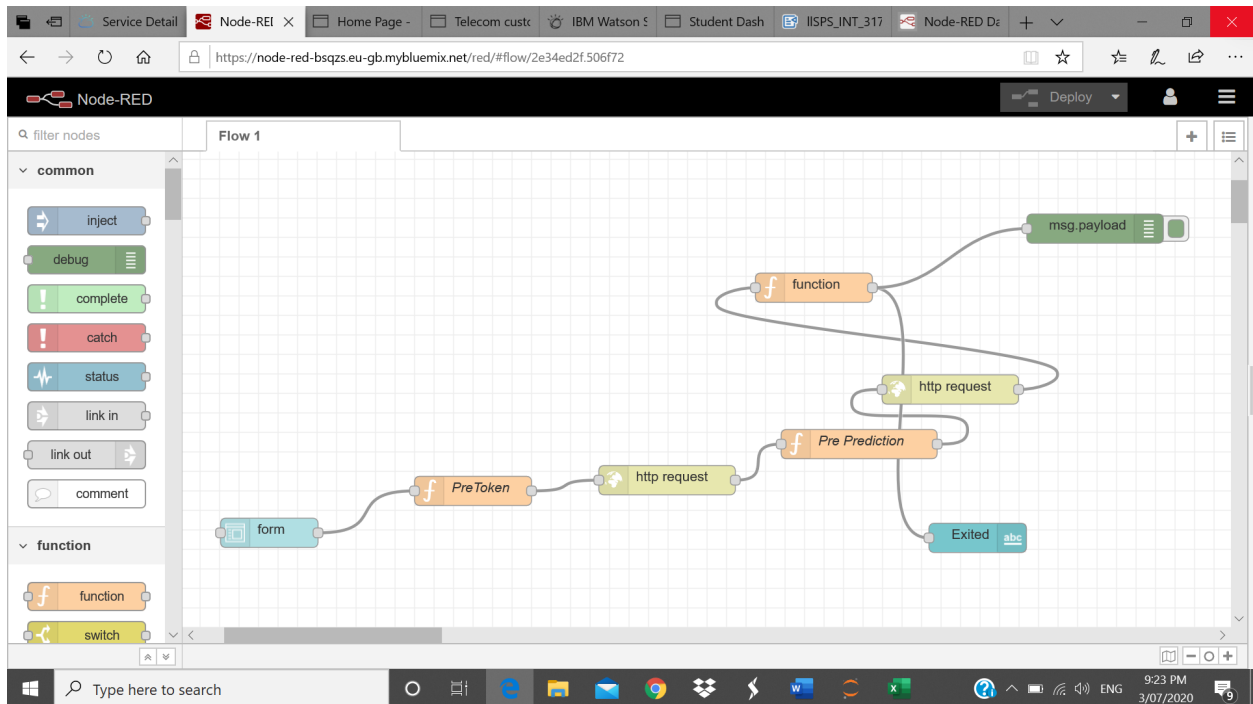
■ Testing the Model:



The screenshot shows the IBM Watson Studio 'Test' interface. On the left, under 'Enter input data', there are four input fields: '2', 'HasCrCard: 1', 'IsActiveMember: 0', and 'EstimatedSalary: 1232232'. A 'Predict' button is at the bottom. On the right, the JSON output is displayed:

```
{
  "predictions": [
    {
      "fields": [
        "prediction",
        "probability"
      ],
      "values": [
        0,
        [
          0.7679167797658842,
          0.23208322023411576
        ]
      ]
    }
  ]
}
```

■ Creating a Node-RED UI:



■ Predicting Output Using Node-RED:

The screenshot shows the Node-RED web interface displaying the predicted output for a churn prediction application. The browser's address bar shows the URL: <https://node-red-bsqzs.eu-gb.mybluemix.net/ui/#/0?socketid=1fFc3iSuN5E0jCTAAAD>. The main content area is titled 'Home' and displays a 'Telecom Customer Churn Prediction' form. The form contains the following input fields and their values:

Field	Value
RowNumber *	4
CustomerId *	1232454
Surname *	Lee
CreditScore *	456
Geography *	France
Gender *	Female
Age *	45
Tenure *	1
Balance *	1243434
NumOfProducts *	1

The Windows taskbar at the bottom shows the time as 10:06 PM on 3/07/2020.

Home Page - SelTelecom customApplication DataNode-RED : nodStudent DashboIISPS_INT_3172_1Node-RED D

<https://node-red-bsqzs.eu-gb.mybluemix.net/ui/#/I/0?socketid=IfFc3iSuN5E0jCTAAAD>

Home

Female

Age *45

Tenure *1

Balance *1243434

NumOfProducts *1

HasCrCard *1

IsActiveMember *0

EstimatedSalary *12344545

SUBMITCANCEL

Exited1

Type here to search

10:06 PM3/07/2020ENG9