TELECOM CUSTOMER CHURN PREDICTION USING WATSON AUTO AI

CHAPTER 1

INTRODUCTION

1.1 Overview:

Customer churn is a big problem for service providers because losing customers results in losing revenue and could indicate service deficiencies. 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 the customer churning from the present due to a better plan. There are many reasons why customers decide to leave services. With machine learning, we can identify the important factors of churning, create a retention plan, and predict which customers are likely to churn.

Acquisition and retention of new clients are one of the most significant concerns of businesses. While recipient companies concentrate on acquiring new customers, mature ones try to focus on retention of the existing ones in order to provide themselves with the opportunity of cross – selling. According to Freeman (1999) one of the most significant ways of increasing customers' value is to keep them for longer period of time.

The telecommunications sector has become one of the main industries in developed countries. The technical progress and the increasing number of operators raised the level of competition [1]. Companies are working hard to survive in this competitive market depending on multiple strategies. Three main strategies have been proposed to generate more revenues [2]: (1) acquire new customers, (2) upsell the existing customers, and (3) increase the retention period of customers. However, comparing these strategies taking the value of return on investment (RoI) of each into account has shown that the third strategy is the most profitable strategy [2], proves that retaining an existing customer costs much lower than acquiring a new one [3], in addition to being considered much easier than the upselling strategy [4]. To apply the third strategy, companies have to decrease the potential of customer's churn, known as "the customer movement from one provider to another" [5].

Customers' churn is a considerable concern in service sectors with high competitive services. On the other hand, predicting the customers who are likely to leave the company will represent potentially large additional revenue source if it is done in the early phase [3].Many research confirmed that machine learning technology is highly efficient to predict this situation. This technique is applied through learning from previous data [6, 7].

1.2 Purpose:

The main contribution of our work is to develop a churn prediction model, which assists telecom operators to predict customers who are most likely subject to churn.

CHAPTER 2

LITERATURE SURVEY

2.1 Existing Problem:

Many approaches were applied to predict churn in telecom companies. Most of these approaches have used machine learning and data mining. The majority of related work focused on applying only one method of data mining to extract knowledge, and the others focused on comparing several strategies to predict churn.

Gavril et al. [9] presented an advanced methodology of data mining to predict churn for prepaid customers using dataset for call details of 3333 customers with 21 features, and a dependent churn parameter with two values: Yes/No. Some features include information about the number of incoming and outgoing messages and voicemail for each customer. The author applied principal component analysis algorithm "PCA" to reduce data dimensions. Three machine learning algorithms were used: Neural Networks, Support Vector Machine, and Bayes Networks to predict churn factor. The author used AUC to measure the performance of the algorithms. The AUC values were 99.10%, 99.55% and 99.70% for Bayes Networks, Neural networks and support vector machine, respectively. The dataset used in this study is small and no missing values existed.

He et al. [10] proposed a model for prediction based on the Neural Network algorithm in order to solve the problem of customer churn in a large Chinese telecom company which contains about 5.23 million customers. The prediction accuracy standard was the overall accuracy rate, and reached 91.1%.

Idris [11] proposed an approach based on genetic programming with AdaBoost to model the churn problem in telecommunications. The model was tested on two standard data sets. One by Orange Telecom and the other by cell2cell, with 89% accuracy for the cell2cell dataset and 63% for the other one.

Huang et al. [12] studied the problem of customer churn in the big data platform. The goal of the researchers was to prove that big data greatly enhance the process of predicting the churn depending on the volume, variety, and velocity of the data. Dealing with data from the Operation Support department and Business Support department at China's largest telecommunications company needed a big data platform to engineer the fractures. Random Forest algorithm was used and evaluated using AUC.

Makhtar et al. [13] proposed a model for churn prediction using rough set theory in telecom. As mentioned in this paper Rough Set classification algorithm outperformed the other algorithms like Linear Regression, Decision Tree, and Voted Perception Neural Network.

Various researches studied the problem of unbalanced data sets where the churned customer classes are smaller than the active customer classes, as it is a major issue in churn prediction problem. Amin et al. [14] compared six different sampling techniques for oversampling regarding telecom churn prediction problem. The results showed that the algorithms (MTDF and rules-generation based on genetic algorithms) outperformed the other compared oversampling algorithms.

Burez and Van den Poel [8] studied the problem of unbalance datasets in churn prediction models and compared performance of Random Sampling, Advanced Under-Sampling, Gradient Boosting Model, and Weighted Random Forests. They used (AUC, Lift) metrics to evaluate the model. The result showed that undersampling technique outperformed the other tested techniques.

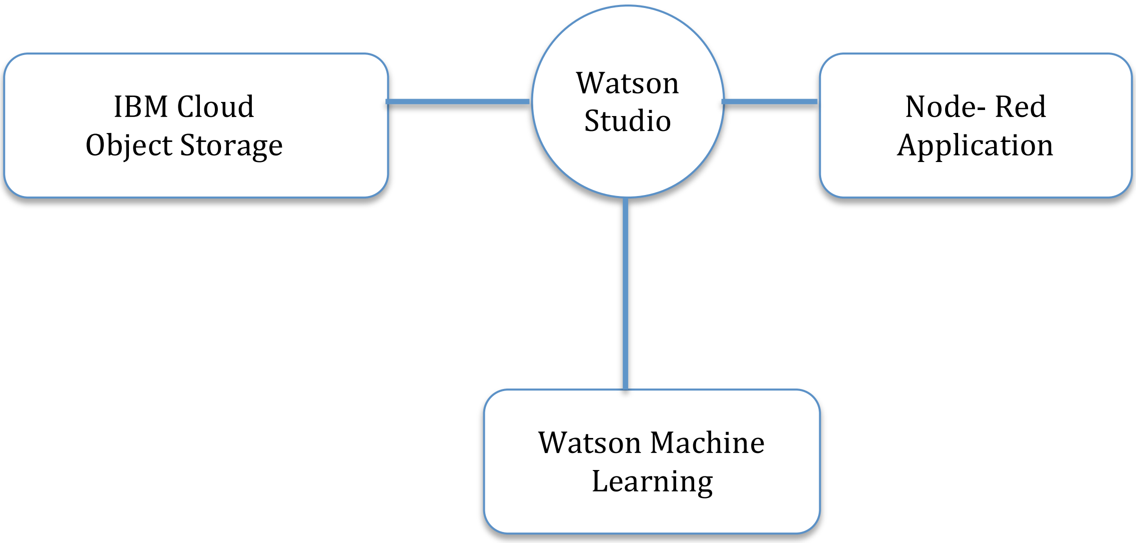
2.2 Proposed Problem:

In this paper, the churn prediction is done using IBM Watson Studio and Auto AI service. The best algorithm that worked good for the dataset was identified and result analysis was done. This project aims to predict if and when a customer could probably churn based on the company's data from the previous month, so as to offer those customers better services. This is a supervised learning problem, which classifies whether the customers will churn or not.

CHAPTER 3

THEORETICAL ANALYSIS

3.1 Block Diagram:



3.2 Hardware/Software Requirements:

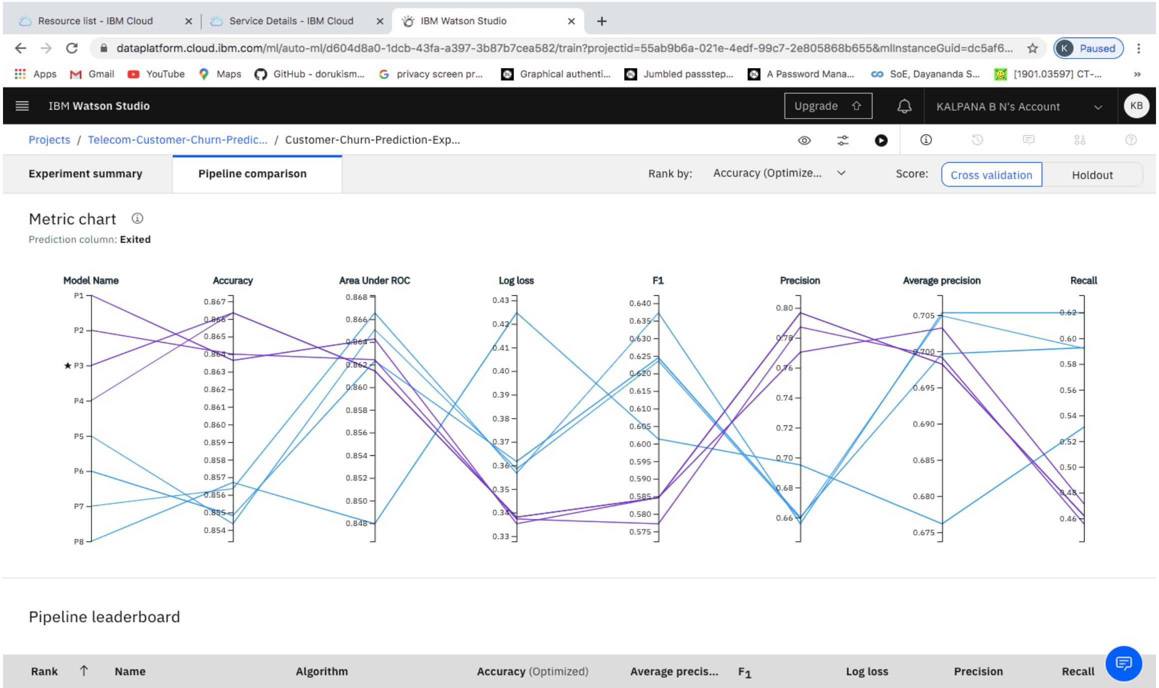
Software Required: IBM Watson Studio, Node-Red, Watson Machine Learning

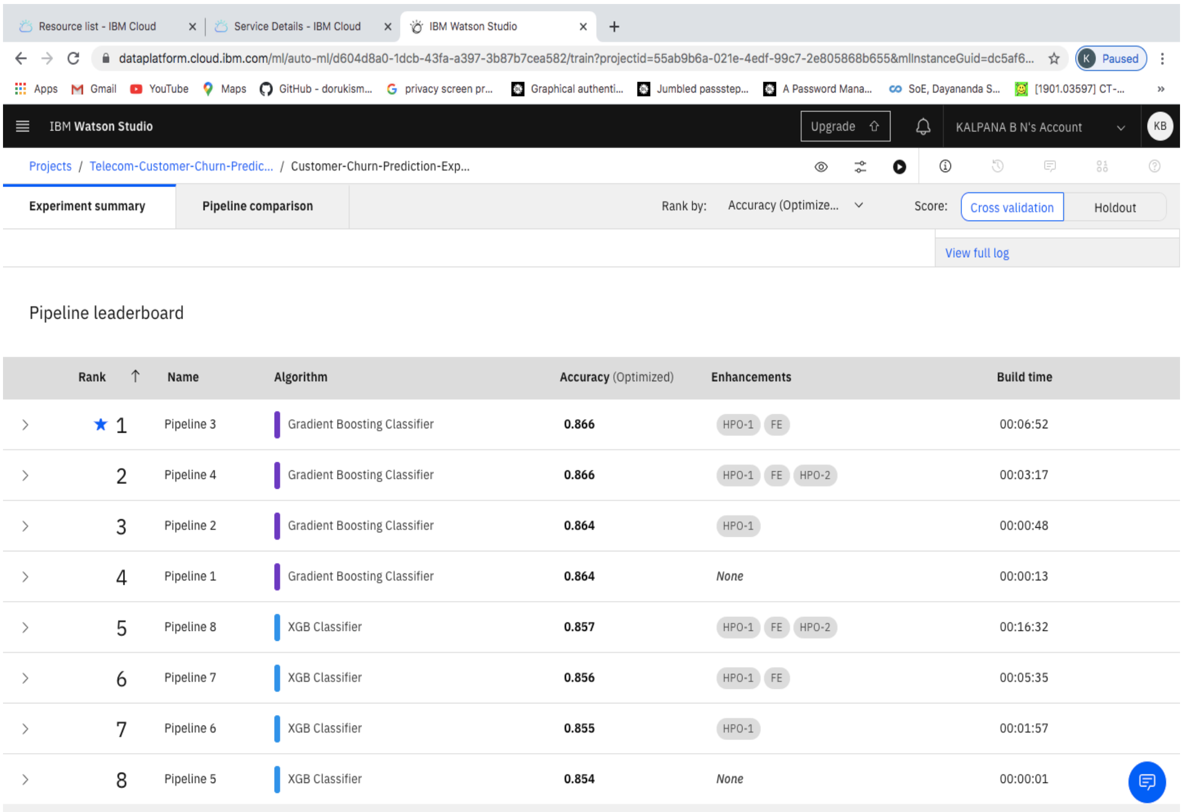
RAM: Minimum 2GB

OS: Windows/ Mac OS

CHAPTER 4

EXPERIMENTAL INVESTIGATION



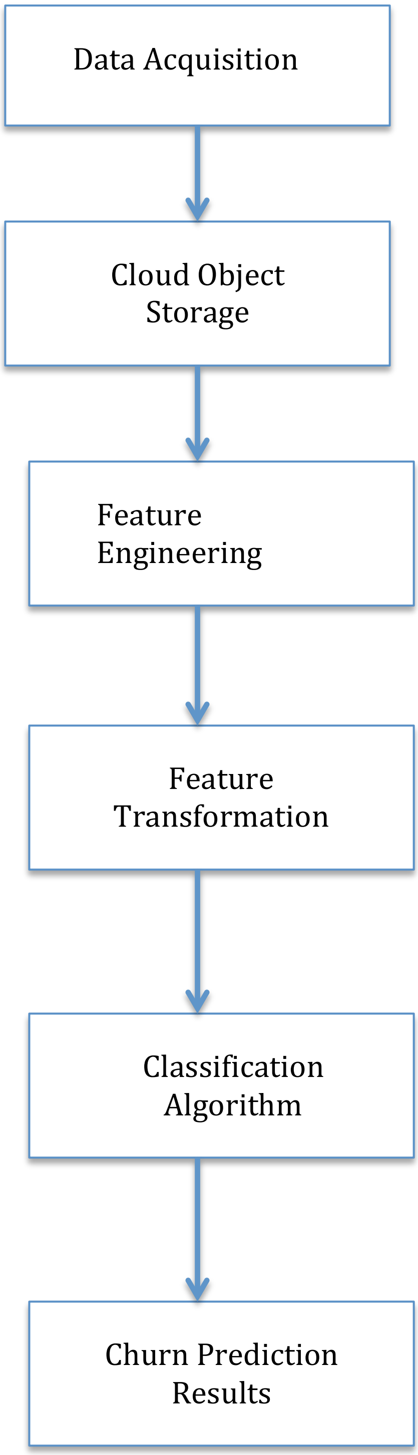


Rank	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1	Pipeline 3	Gradient Boosting Classifier	0.866	HPO-1 FE	00:06:52
2	Pipeline 4	Gradient Boosting Classifier	0.866	HPO-1 FE HPO-2	00:03:17
3	Pipeline 2	Gradient Boosting Classifier	0.864	HPO-1	00:00:48
4	Pipeline 1	Gradient Boosting Classifier	0.864	None	00:00:13
5	Pipeline 8	XGB Classifier	0.857	HPO-1 FE HPO-2	00:16:32
6	Pipeline 7	XGB Classifier	0.856	HPO-1 FE	00:05:35
7	Pipeline 6	XGB Classifier	0.855	HPO-1	00:01:57
8	Pipeline 5	XGB Classifier	0.854	None	00:00:01

The solution we proposed divided the data into two groups: the training group and the testing group. The training group consists of 80% of the dataset and aims to train the algorithms. The test group contains 20% of the dataset and is used to test the algorithms. The hyperparameters of the algorithms were optimized using cross-validation. Accuracy is one of the most common metrics for binary classifiers. The Gradient Boosting Classifier showed higher accuracy with 0.866.

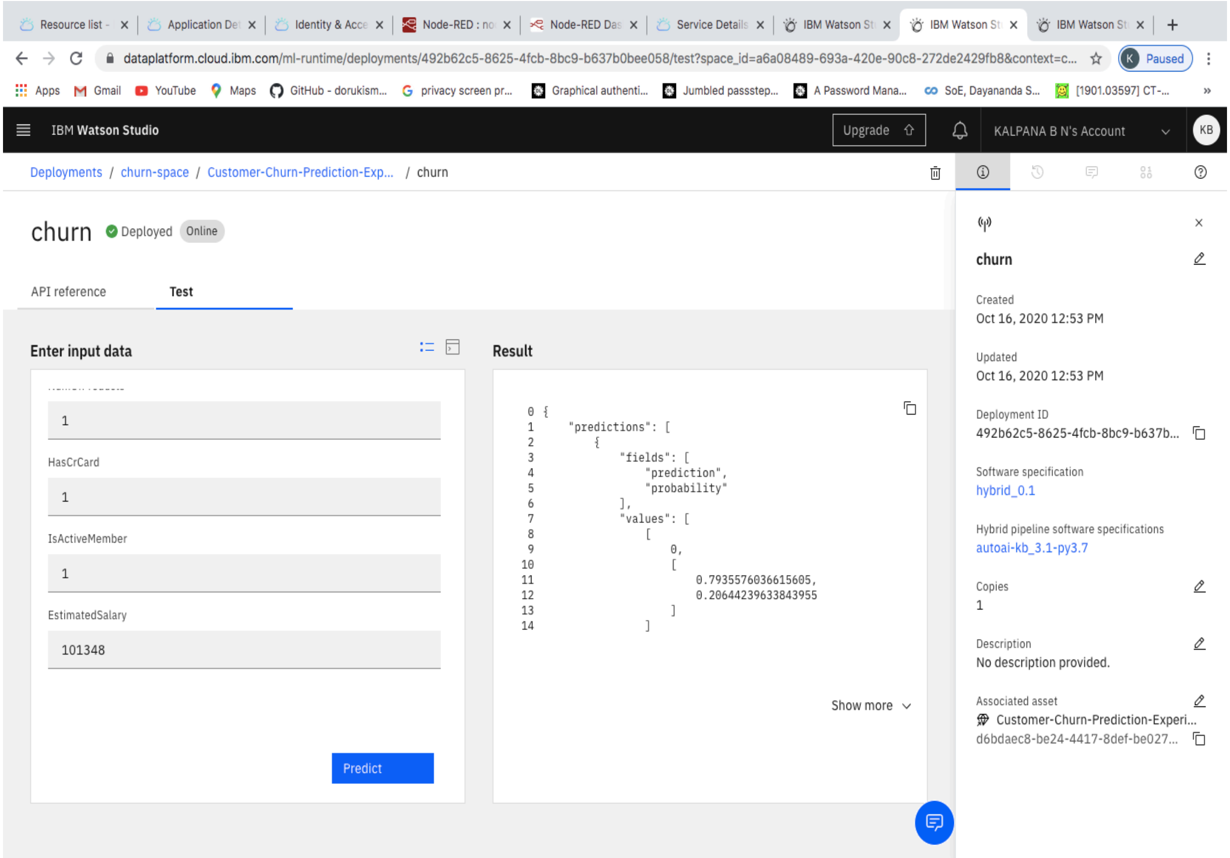
CHAPTER 5

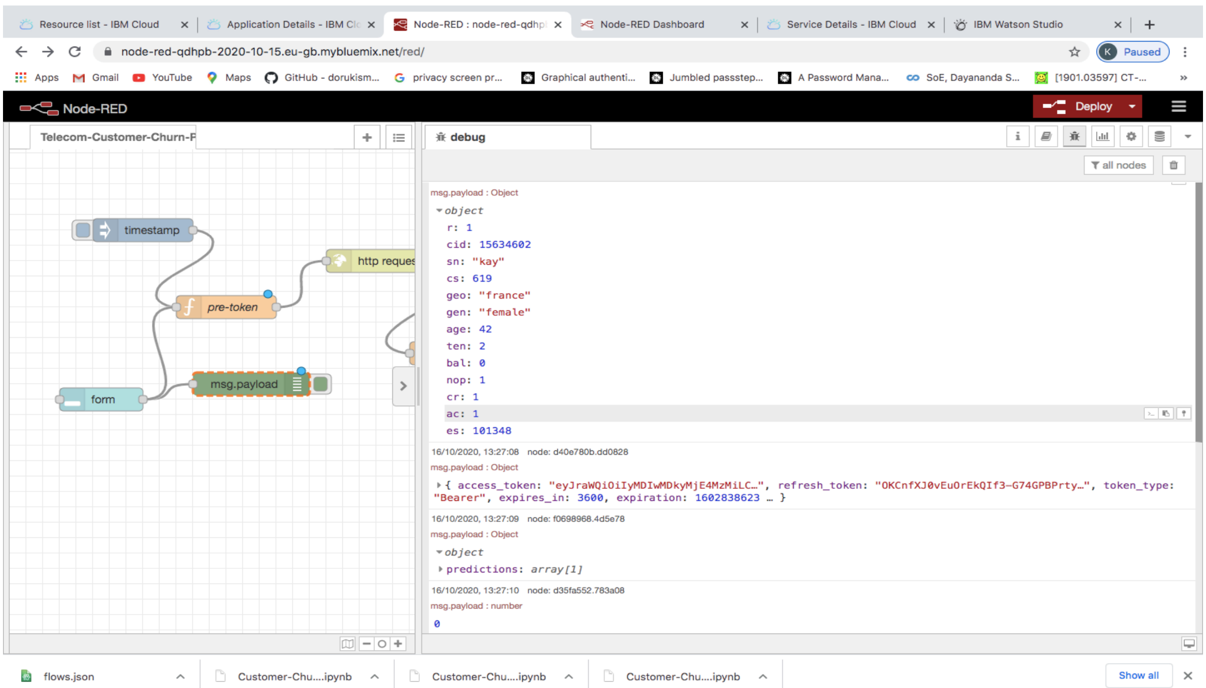
FLOWCHART

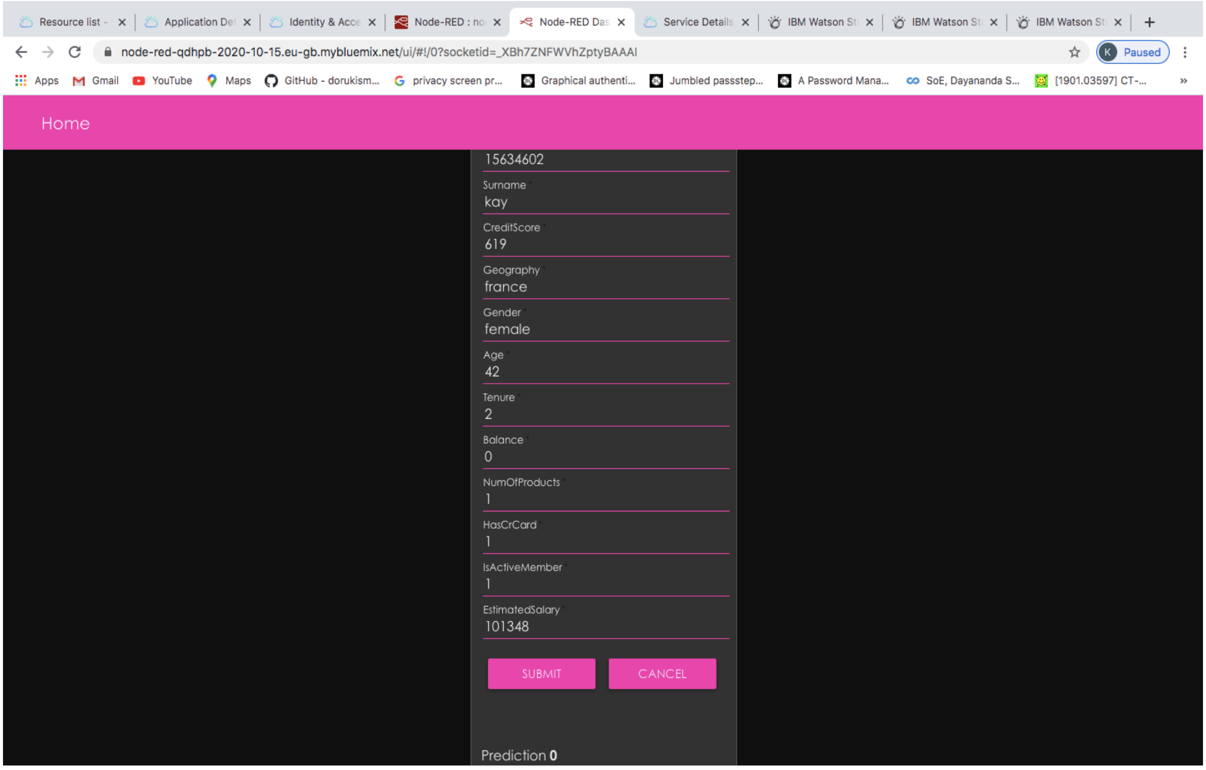


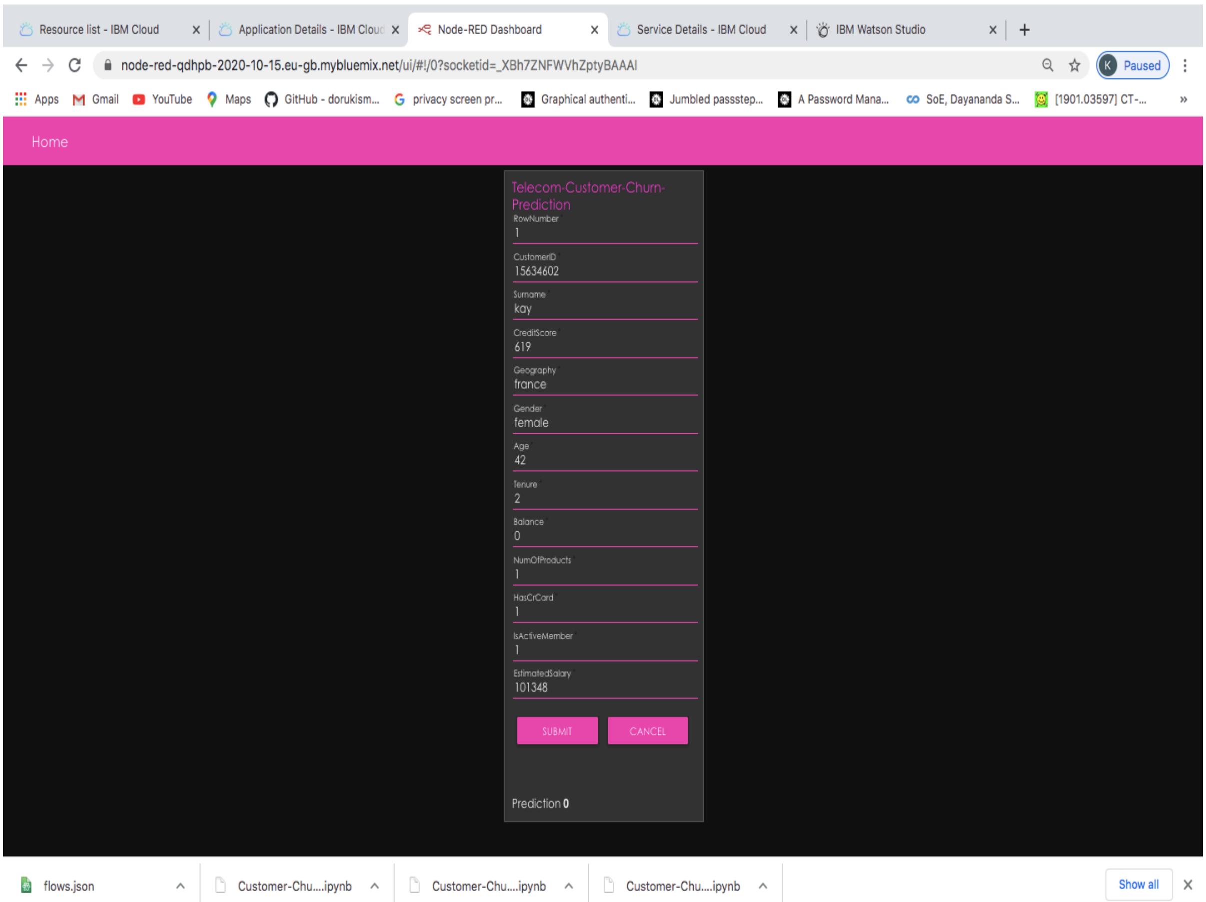
CHAPTER 6

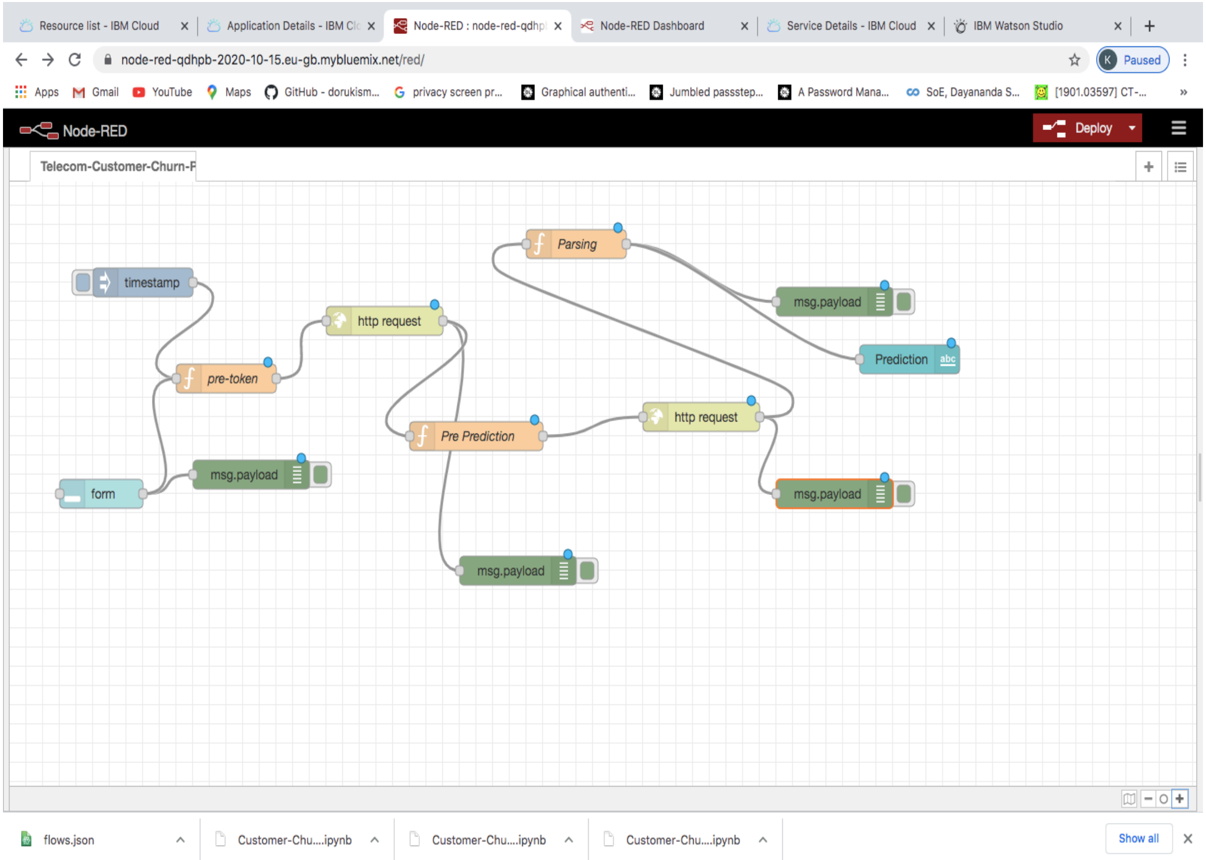
RESULT






node-red-qdhpb-2020-10-15.eu-gb.mybluemix.net/ui/#!/0?socketid=_XBh7ZNFWVhZptyBAAAI
Home
15634602
Surname
kay
CreditScore
619
Geography
france
Gender
female
Age
42
Tenure
2
Balance
0
NumOfProducts
1
HasCrCard
1
IsActiveMember
1
EstimatedSalary
101348
SUBMIT
CANCEL
Prediction 0


Resource list - IBM Cloud
Application Details - IBM Cloud
Node-RED Dashboard
Service Details - IBM Cloud
IBM Watson Studio
node-red-qdhpb-2020-10-15.eu-gb.mybluemix.net/ui/#!/0?socketid=_XBh7ZNFWVhZptyBAAAI
Home
Telecom-Customer-Churn-Prediction
RowNumber
1
CustomerID
15634602
Surname
kay
CreditScore
619
Geography
france
Gender
female
Age
42
Tenure
2
Balance
0
NumOfProducts
1
HasCrCard
1
IsActiveMember
1
EstimatedSalary
101348
SUBMIT
CANCEL
Prediction 0
flows.json
Customer-Chu....ipynb
Customer-Chu....ipynb
Customer-Chu....ipynb
Show all



node-red-qdhpb-2020-10-15.eu-gb.mybluemix.net/red/
Node-RED
Deploy
Telecom-Customer-Churn-F
timestamp
pre-token
http request
Parsing
msg.payload
Prediction
Pre Prediction
http request
form
msg.payload
msg.payload
msg.payload
flows.json
Customer-Chu....ipynb
Customer-Chu....ipynb
Customer-Chu....ipynb
Show all

CHAPTER 7

ADVANTAGES AND DISADVANTAGES

Advantages:

The built model worked fine with the chosen dataset.

Disadvantages:

The model when applied to other datasets prediction accuracy was low.

CHAPTER 8

APPLICATIONS

- This project can be used in industries where they need to predict the customers who are likely to churn, to improve their profit.
- Specifically in Telecom sector, this can be used.

CHAPTER 9

CONCLUSION

The importance of this type of research in the telecom market is to help companies make more profit. It has become known that predicting churn is one of the most important sources of income to telecom companies. Hence, this project aimed to build a system that predicts the churn of customers.

CHAPTER 10

FUTURE SCOPE

The dataset with more number of rows and columns can be chosen to train the model to get better accuracy. Lot of cleaning and preprocessing work has to be done and benchmark model can be used for better results.

CHAPTER 11

BIBLIOGRAPHY

References:

1. Gerpott TJ, Rams W, Schindler A. Customer retention, loyalty, and satisfaction in the German mobile cellular telecommunications market. Telecommun Policy. 2001;25:249–69.

2. Wei CP, Chiu IT. Turning telecommunications call details to churn prediction: a data mining approach. Expert Syst Appl. 2002;23(2):103–12.

3. Qureshii SA, Rehman AS, Qamar AM, Kamal A, Rehman A. Telecommunication subscribers' churn prediction model using machine learning. In: Eighth international conference on digital information management. 2013. p. 131–6.

4. Ascarza E, Iyengar R, Schleicher M. The perils of proactive churn prevention using plan recommendations: evidence from a field experiment. J Market Res. 2016;53(1):46–60.

5. Bott. Predicting customer churn in telecom industry using multilayer preceptron neural networks: modeling and analysis. Igarss. 2014;11(1):1–5.

6. Umayaparvathi V, Iyakutti K. A survey on customer churn prediction in telecom industry: datasets, methods and metric. Int Res J Eng Technol. 2016;3(4):1065–70.

7. Yu W, Jutla DN, Sivakumar SC. A churn-strategy alignment model for managers in mobile telecom. In: Communication networks and services research conference, vol. 3. 2005. p. 48–53.

8. Burez D, den Poel V. Handling class imbalance in customer churn prediction. Expert Syst Appl. 2009;36(3):4626–36.

9. Brandusoiu I, Toderean G, Ha B. Methods for churn prediction in the prepaid mobile telecommunications industry. In: International conference on communications. 2016. p. 97–100.

10. He Y, He Z, Zhang D. A study on prediction of customer churn in fixed communication network based on data mining. In: Sixth international conference on fuzzy systems and knowledge discovery, vol. 1. 2009. p. 92–4.

11. Idris A, Khan A, Lee YS. Genetic programming and adaboosting based churn prediction for telecom. In: IEEE international conference on systems, man, and cybernetics (SMC). 2012. p. 1328–32.

12. Huang F, Zhu M, Yuan K, Deng EO. Telco churn prediction with big data. In: ACM SIGMOD international conference on management of data. 2015. p .607–18.

13. Makhtar M, Nafis S, Mohamed M, Awang M, Rahman M, Deris M. Churn classification model for local telecommunication company based on rough set theory. J Fundam Appl Sci. 2017;9(6):854–68.

14. Amin A, Anwar S, Adnan A, Nawaz M, Howard N, Qadir J, Hawalah A, Hussain A. Comparing oversampling techniques to handle the class imbalance problem: a customer churn prediction case study. IEEE Access. 2016;4:7940–57.

APPENDIX

Source Code:

Flows.json

[{"id":"394fa569.c5ac2a","type":"tab","label":"Telecom-Customer-
Churn-
Prediction","disabled":false,"info":""},{"id":"968a1fe1.cedf1","type"
:"ui_base","theme":{"name":"theme-
dark","lightTheme":{"default":"#0094CE","baseColor":"#0094CE","baseFo
nt":"-apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-
Sans,Ubuntu,Cantarell,Helvetica Neue,sans-
serif","edited":true,"reset":false},"darkTheme":{"default":"#097479",
"baseColor":"#e60aab","baseFont":"Century
Gothic,CenturyGothic,AppleGothic,sans-
serif","edited":true,"reset":false},"customTheme":{"name":"Untitled
Theme 1","default":"#4B7930","baseColor":"#4B7930","baseFont":"-
apple-system,BlinkMacSystemFont,Segoe UI,Roboto,Oxygen-
Sans,Ubuntu,Cantarell,Helvetica Neue,sans-
serif"},"themeState":{"base-
color":{"default":"#097479","value":"#e60aab","edited":true},"page-
titlebar-backgroundColor":{"value":"#e60aab","edited":false},"page-
backgroundColor":{"value":"#111111","edited":false},"page-sidebar-
backgroundColor":{"value":"#ffffff","edited":false},"group-
textColor":{"value":"#f746c7","edited":false},"group-
borderColor":{"value":"#555555","edited":false},"group-
backgroundColor":{"value":"#333333","edited":false},"widget-
textColor":{"value":"#eeeeee","edited":false},"widget-
backgroundColor":{"value":"#e60aab","edited":false},"widget-
borderColor":{"value":"#333333","edited":false},"base-
font":{"value":"Century Gothic,CenturyGothic,AppleGothic,sans-
serif"}},"angularTheme":{"primary":"indigo","accents":"blue","warn":"
red","background":"grey"}},"site":{"name":"Node-RED
Dashboard","hideToolbar":"false","allowSwipe":"false","lockMenu":"fal
se","allowTempTheme":"true","dateFormat":"DD/MM/YYYY","sizes":{"sx":4
8,"sy":48,"gx":6,"gy":6,"cx":6,"cy":6,"px":0,"py":0}}},{"id":"6d0e156
7.df564c","type":"ui_tab","z":"","name":"Home","icon":"dashboard","di
sabled":false,"hidden":false},{"id":"618c85b2.0ac14c","type":"ui_grou
p","z":"","name":"Telecom-Customer-Churn-
Prediction","tab":"6d0e1567.df564c","order":1,"disp":true,"width":"6"
,"collapse":false},{"id":"ea5c8f27.ca778","type":"debug","z":"394fa56
9.c5ac2a","name":"","active":true,"tosidebar":true,"console":false,"t
ostatus":false,"complete":"payload","targetType":"msg","statusVal":""
,"statusType":"auto","x":430,"y":480,"wires":[]},{"id":"ecd2a7cc.c047
08","type":"function","z":"394fa569.c5ac2a","name":"pre-
token","func":"global.set(\"r\",msg.payload.r)\nglobal.set(\"cid\",ms
g.payload.cid)\nglobal.set(\"sn\",msg.payload.sn)\nglobal.set(\"cs\",
msg.payload.cs)\nglobal.set(\"geo\",msg.payload.geo)\nglobal.set(\"ge
n\",msg.payload.gen)\nglobal.set(\"age\",msg.payload.age)\nglobal.set
(\"ten\",msg.payload.ten)\nglobal.set(\"bal\",msg.payload.bal)\ngloba
l.set(\"nop\",msg.payload.nop)\nglobal.set(\"cr\",msg.payload.cr)\ngl
obal.set(\"ac\",msg.payload.ac)\nglobal.set(\"es\",msg.payload.es)\nv
ar
apikey=\"BEgIF2R6kF5PRAKG8_jBnYDPv9vljZ4ZplNOm09knzxq\";\nmsg.headers
={\"content-type\":\"application/x-www-form-
urlencoded\"}\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:gra
nt-type:apikey\",\"apikey\":apikey}\nreturn
msg;\n","outputs":1,"noerr":0,"initialize":"","finalize":"","x":480,"
y":280,"wires":[["3d043c77.d72b84"]]},{"id":"3d043c77.d72b84","type":
"http
request","z":"394fa569.c5ac2a","name":"","method":"POST","ret":"obj",
"paytoqs":"ignore","url":"https://iam.cloud.ibm.com/identity/token","
tls":"","persist":false,"proxy":"","authType":"","x":650,"y":360,"wir

es":[["d40e780b.dd0828","c2385056.a288e"]]},{"id":"d40e780b.dd0828","
type":"debug","z":"394fa569.c5ac2a","name":"","active":true,"tosideba
r":true,"console":false,"tostatus":false,"complete":"false","statusVa
l":"","statusType":"auto","x":730,"y":540,"wires":[]},{"id":"c2385056
.a288e","type":"function","z":"394fa569.c5ac2a","name":"Pre
Prediction","func":"var r = global.get(\"r\")\nvar cid =
global.get(\"cid\")\nvar sn = global.get(\"sn\")\nvar cs =
global.get(\"cs\")\nvar geo = global.get(\"geo\")\nvar gen =
global.get(\"gen\")\nvar age = global.get(\"age\")\nvar ten =
global.get(\"ten\")\nvar bal= global.get(\"bal\")\nvar nop =
global.get(\"nop\")\nvar cr = global.get(\"cr\")\nvar ac =
global.get(\"ac\")\nvar es = global.get(\"es\")\nvar
token=msg.payload.access_token\nmsg.headers={'Content-Type':
'application/json',\"Authorization\":\"Bearer
\"+token,\"Accept\":\"application/json\"}\nmsg.payload={\"input_data\
":[{\"fields\":[\"RowNumber\",\"CustomerId\",\"Surname\",\"CreditScor
e\",\"Geography\",\"Gender\",\"Age\",\"Tenure\",\"Balance\",\"NumOfPr
oducts\",\"HasCrCard\",\"IsActiveMember\",\"EstimatedSalary\"],\"valu
es\": [[r,cid,sn,cs,geo,gen,age,ten,bal,nop,cr,ac,es]]}]}\nreturn
msg;","outputs":1,"noerr":0,"initialize":"","finalize":"","x":940,"y"
:440,"wires":[["864f1f21.6c298"]]},{"id":"864f1f21.6c298","type":"htt
p
request","z":"394fa569.c5ac2a","name":"","method":"POST","ret":"obj",
"paytoqs":"ignore","url":"https://us-
south.ml.cloud.ibm.com/ml/v4/deployments/492b62c5-8625-4fcb-8bc9-
b637b0bee058/predictions?version=2020-09-
01","tls":"","persist":false,"proxy":"","authType":"","x":1230,"y":40
0,"wires":[["f0698968.4d5e78","2ab592e8.c8d81e"]]},{"id":"f0698968.4d
5e78","type":"debug","z":"394fa569.c5ac2a","name":"","active":true,"t
osidebar":true,"console":false,"tostatus":false,"complete":"false","s
tatusVal":"","statusType":"auto","x":1530,"y":400,"wires":[]},{"id":"
8a13ddf8.e0cd5","type":"inject","z":"394fa569.c5ac2a","name":"","prop
s":[{"p":"payload"},{"p":"topic","vt":"str"}],"repeat":"","crontab":"
","once":false,"onceDelay":0.1,"topic":"","payload":"","payloadType":
"date","x":160,"y":240,"wires":[["ecd2a7cc.c04708"]]},{"id":"d35fa552
.783a08","type":"debug","z":"394fa569.c5ac2a","name":"","active":true
,"tosidebar":true,"console":false,"tostatus":false,"complete":"false"
,"statusVal":"","statusType":"auto","x":1230,"y":260,"wires":[]},{"id
":"2ab592e8.c8d81e","type":"function","z":"394fa569.c5ac2a","name":"P
arsing","func":"msg.payload=msg.payload.predictions[0].values[0][0]\n
return
msg;","outputs":1,"noerr":0,"initialize":"","finalize":"","x":940,"y"
:260,"wires":[["d35fa552.783a08","35d66199.427f3e"]]},{"id":"8d612992
.f537c8","type":"ui_form","z":"394fa569.c5ac2a","name":"","label":"",
"group":"618c85b2.0ac14c","order":1,"width":0,"height":0,"options":[{
"label":"RowNumber","value":"r","type":"number","required":true,"rows
":null},{"label":"CustomerID","value":"cid","type":"number","required
":true,"rows":null},{"label":"Surname","value":"sn","type":"text","re
quired":true,"rows":null},{"label":"CreditScore","value":"cs","type":
"number","required":true,"rows":null},{"label":"Geography","value":"g
eo","type":"text","required":true,"rows":null},{"label":"Gender","val
ue":"gen","type":"text","required":true,"rows":null},{"label":"Age","
value":"age","type":"number","required":true,"rows":null},{"label":"T
enure","value":"ten","type":"number","required":true,"rows":null},{"l
abel":"Balance","value":"bal","type":"number","required":true,"rows":
null},{"label":"NumOfProducts","value":"nop","type":"number","require
d":true,"rows":null},{"label":"HasCrCard","value":"cr","type":"number
","required":true,"rows":null},{"label":"IsActiveMember","value":"ac"
,"type":"number","required":true,"rows":null},{"label":"EstimatedSala
ry","value":"es","type":"number","required":true,"rows":null}],"formV
alue":{"r":"","cid":"","sn":"","cs":"","geo":"","gen":"","age":"","te
n":"","bal":"","nop":"","cr":"","ac":"","es":""},"payload":"","submit
":"submit","cancel":"cancel","topic":"","x":110,"y":460,"wires":[["ea

5c8f27.ca778","ecd2a7cc.c04708"]],"icon":"font-awesome/fa-window-minimize"},{"id":"35d66199.427f3e","type":"ui_text","z":"394fa569.c5ac2a","group":"618c85b2.0ac14c","order":2,"width":0,"height":0,"name":"","label":"Prediction","format":"{{msg.payload}}","layout":"row-left","x":1340,"y":340,"wires":[]}]