Telecom customer churn prediction using Watson Auto AI

INTRODUCTION :

➤Overview :-

Customer churn is one of the major problem for large companies. Due to direct effect on revenues of companies specially in telecom field, companies are seeking to predict potential customer to churn. Therefore finding the factors that causes these churn and countermeasures that should be taken to reduce these potential churn. The main goal of our project is to develop an easy & simple model which can assist telecom operator to predict the customers who would most likely churn.
Model developed through this project uses machine learning on data platform and builds ways of features' engineering and selection. Standard measures like Area Under Curve(AUC) is adopted to measure the performance of the model along with other method which is Social Network Analysis(SNA), by the way it's accurate up to 93.3% which we can say is a good value. The use of SNA against AUC increased the performance from 84 to 93.3%.
This model was first created by Spark environment by working on large database obtained by converting raw data form Syria telecom into datasets. It contained customers' over 9 months, used to train, test and evaluate the system of SyriaTel. It mainly used four algo: Decision Tree, Random Forest, Gradient Boosted Machine Tree(GBM) & Extreme Gradient Boosting (XGBOOST) by the way best results were obtained on XGBOOST algo which was used for classification in this churn.

➤Purpose:

Well apart from the basics, returning to main issue here i.e., why do we need to predict these potential customer churn. Answer is simple cause technical progress and increasing no. of operators raised level of competition. Companies are working hard to survive in this competitive environment using 3 main strategies namely:

1.acquire new customers
2.upsell exiting customers
3.increase retention period of customers

However the best feasible of the mentioned method is 3rd one which is to increase retention period of customers. One of the main reason is it's less exhausting and economically feasible based on value of return on investment(RoI). To apply the 3rd strategy companies have to

Telecom customer churn prediction using Watson Auto AI

decrease the potential of customer's churn which mainly refers to process of "the customer movement from one provider to another". However predicting customer churn at early phase represent potentially large additional revenue source, which is to say machine learning technology fulfills the criteria to create and test such models regarding analysis of large datasets. As I mentioned earlier this model first used raw data form SyriaTel telecom by converting it into datasets and performing lots of algo tests. The mentioned data was of size about 70 Tb on HDFS "Hadoop Distributed File System" 7 has different data structures. the dataset is aggregated to extract features for each customer. In the process, social network of all customers got built and calculated features like degree centrality measures, customer's network connectivity for each customer etc. Also SNA features made good enhancement in AUC results cause we got different info about customers. We focused on evaluating & analyzing performance of a set of tree- based ML techniques and algo for predicting churn.

LITERATURE SURVEY:

- ### Existing Problem:

Churn prediction modeling techniques attempt to understand the precise customer behaviors and attributes which signal the risk and timing of customer churn. The accuracy of the technique used is obviously critical to the success of any proactive retention efforts. After all, if the marketer is unaware of a customer about to churn, no action will be taken for that customer. Addionally, special retention-focused offers or incentives may be inadvertently provided to happy, active customers, resulting in reduced revenues for no good reason.
Unfortunately, most of the churn prediction modeling methods rely on quantifying risk based on static data and metrics, i.e., information about the customer as he or she exists right now. The most common churn prediction models are based on older statistical and data-mining methods, such as logistic regression and other binary modeling techniques. These approaches offer some value and can identify a certain percentage of at-risk customers, but they are relatively inaccurate and end up leaving money on the table.
In particular cases like SeriaTel telecom there were many challenges like unbalance challenge, where churn customer class was very small compared to active customers' class.

Telecom customer churn prediction using Watson Auto AI

Many previous attempts using data warehouse system to decrease churn rate in SeriaTel were applied. Data Warehouse aggregated some kind of data like billing data, call/messages and complaints etc. Data mining technique were applied on top of data warehouse technique, but model did not produce good results.
Causes were huge data were ignored due to its complex handling on the other hand data warehouse system was not able to handle the huge amount of data.

➤ Proposed Solution:

Big data system allowed SyriaTel Company to collect, store, process, aggregate the data easily regardless of its volume, variety, and complexity. In addition, it enabled extracting richer and more diverse features like SNA features that provide additional information to enhance the churn predictive model. We believe that big data facilitated the process of feature engineering which is one of the most difficult and complex processes in building predictive models. By using the big data platform, we give the power to SyriaTel company to go farther with big data sources. In addition, the company becomes able to extract the Social Network Analysis features from a big scale social graph which is built from billions of edges (transactions) that connect millions of nodes (customers). The hardware and the design of the big data platform illustrated in “Proposed churn method” section fit the need to compute these features regardless of their complexity on this big scale graph. The model also was evaluated using a new dataset and the impact of this system to the decision to churn was tested.

The model gave good results and was deployed to production.

THEORITICAL ANALYSIS:

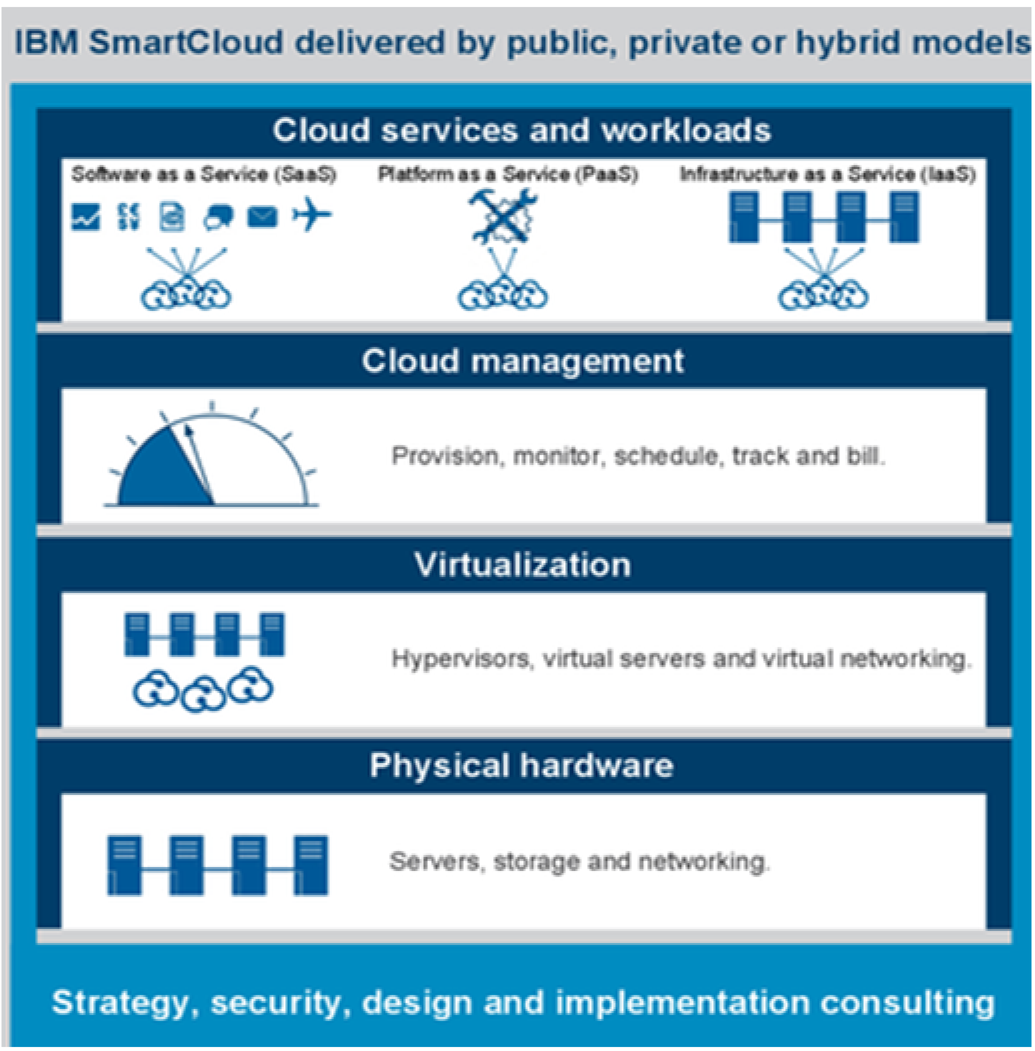
➤ Block diagram:

IBM offers three hardware platforms for cloud computing. These platforms offer built-in support for virtualization. For virtualization IBM offers IBM Websphere application infrastructure that supports programming models and open standards for virtualization. The management layer of the IBM cloud framework includes IBM Tivoli middle ware. Management tools provide capabilities to regulate images with automated provisioning and de-provisioning, monitor operations and meter usage while tracking costs and allocating billing. The last layer of the framework provides integrated workload tools. Workloads for cloud computing are services or instances of code that can be executed to meet specific business needs. IBM offers tools for

Telecom customer churn prediction using Watson Auto AI

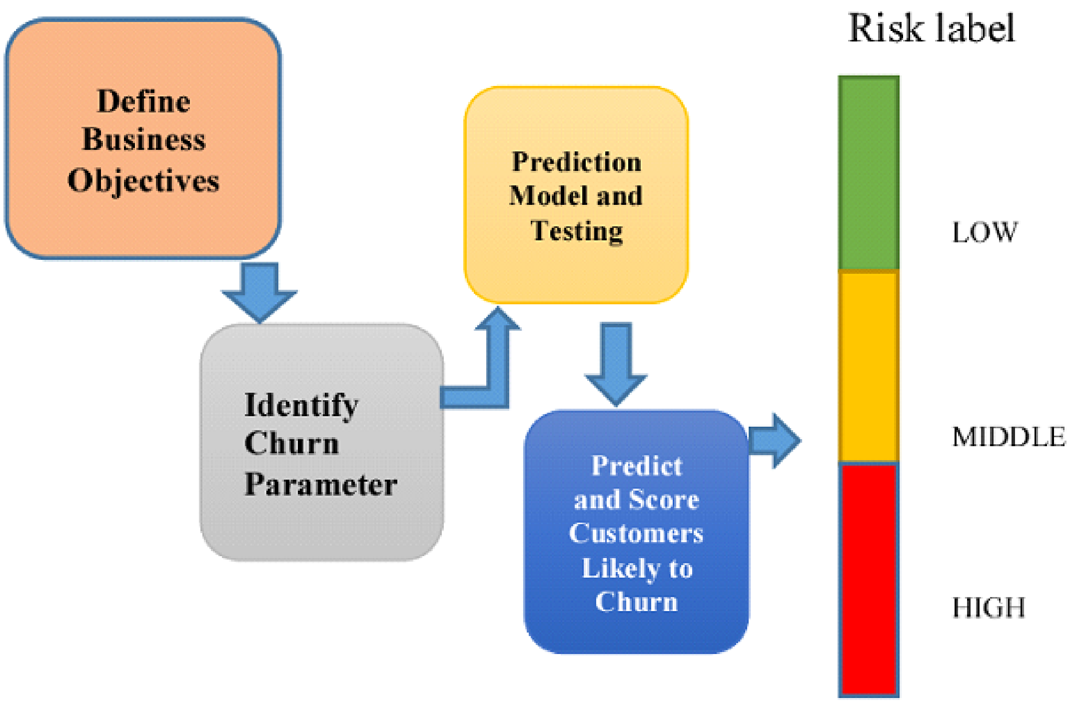
cloud based collaboration, development and test, application development, analytics, business-to-business integration, and security. IBM Watson Studio helps data scientists and analysts prepare data and build models at scale across any cloud. With its open, flexible multi cloud architecture, Watson Studio provides capabilities that empower businesses to simplify enterprise data science and AI, such as:

- Automate AI lifecycle management with AutoAI
- Visually prepare and build models with IBM SPSS Modeler
- Build models using images with IBM Watson Visual Recognition and texts with IBM Watson Natural Language Classifier
- Deploy and run models through one-click integration with IBM Watson Machine Learning
- Manage and monitor models through integration with IBM Watson OpenScale



Telecom customer churn prediction using Watson Auto AI

- Schematic diagram of churn prediction problem :



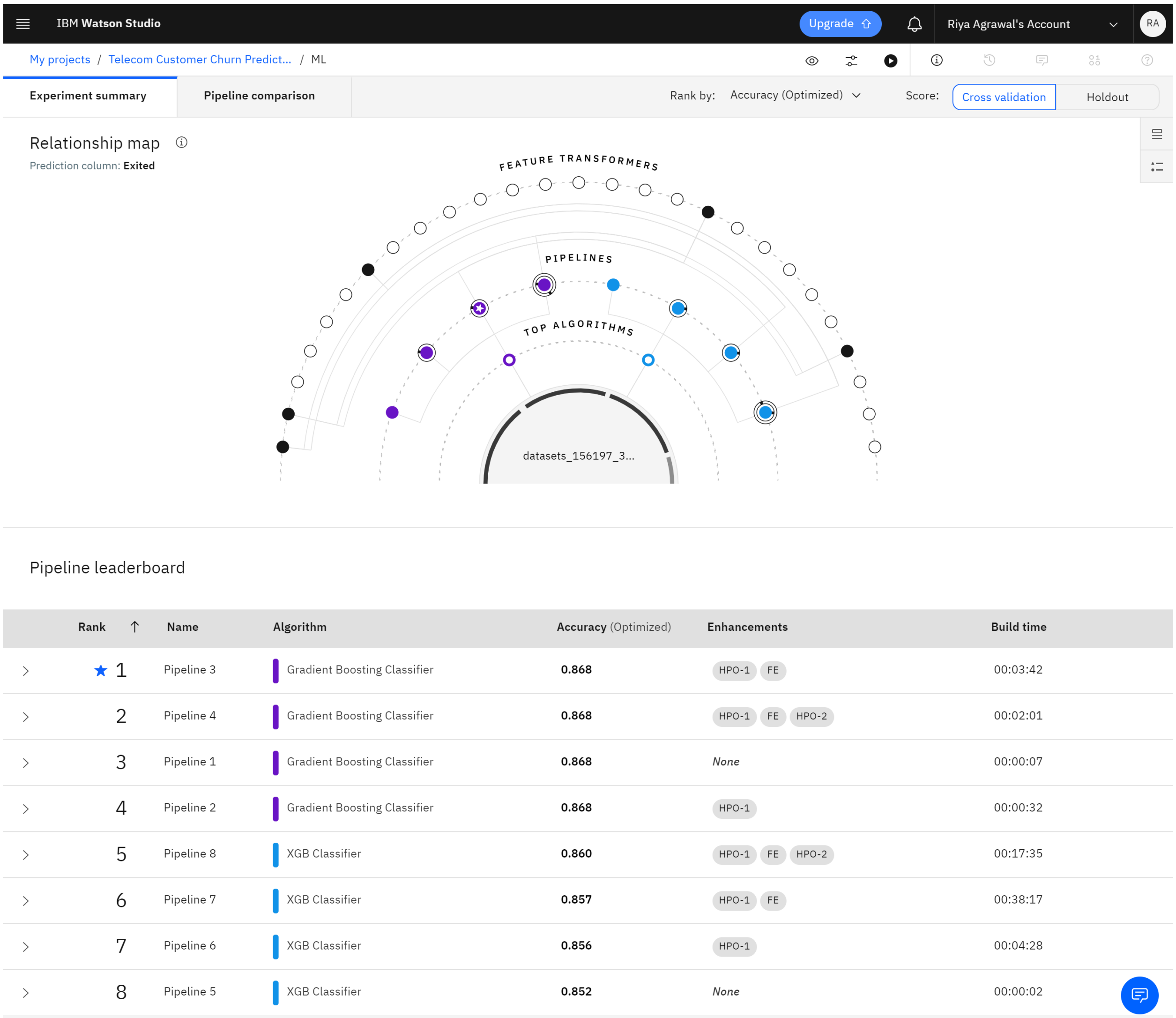
- **Hardware / Software designing:**

While building models with Watson AutoAI on IBM Cloud platform it automatically selects the best suited Machine Learning Algorithm for the provided dataset and the selection criteria. Watson AutoAI does so by applying pipelinings to the datasets, it puts the provided datasets in the multiple Machine Learning Algorithms and selects the best suited algorithm for the provided datasets based on minimum RMSE(Root Mean Squared Errors) values of the algorithms.

Above working is shown by the following screenshots of AutoAI experiment.

Telecom customer churn prediction using Watson Auto AI

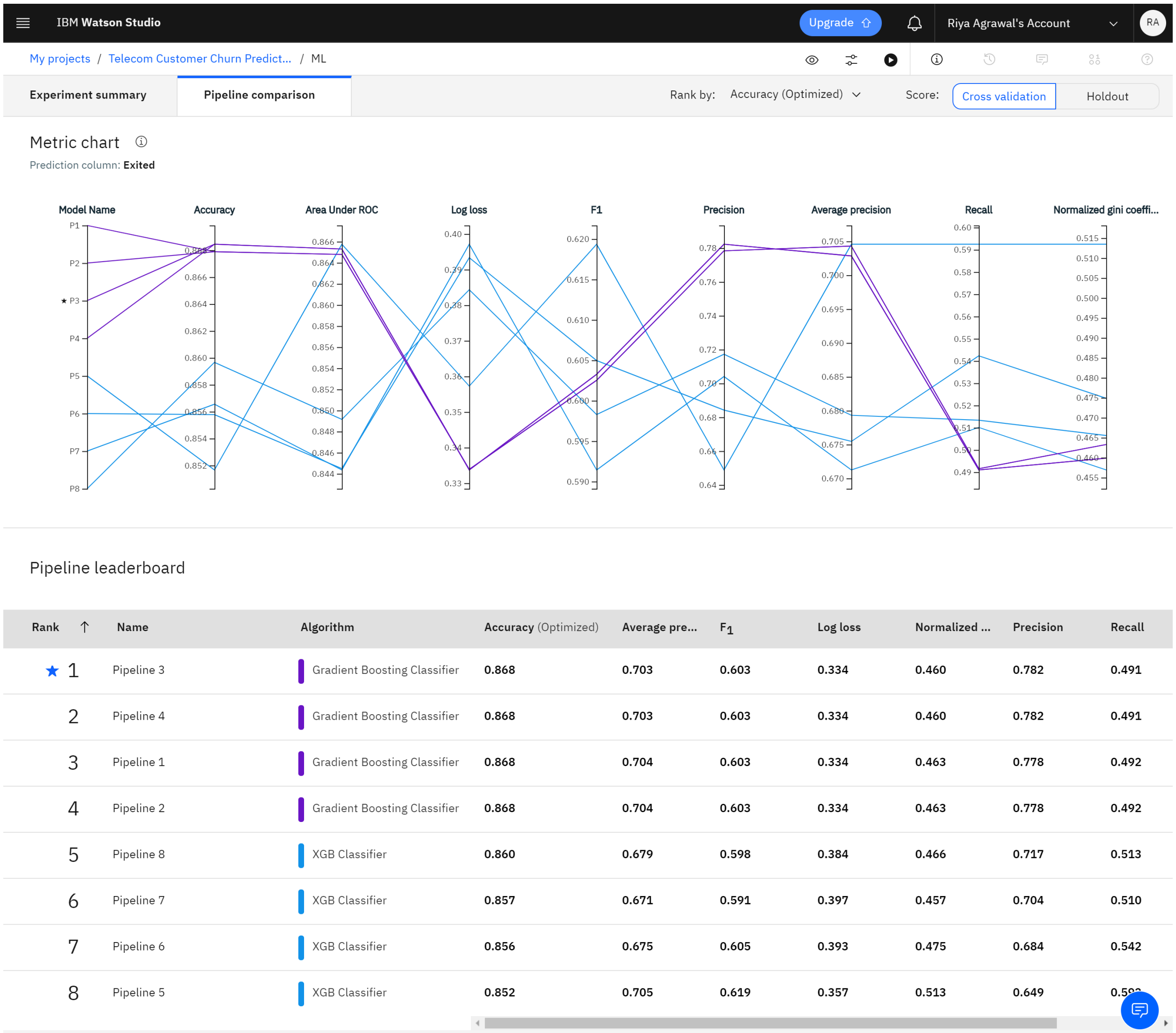
- ## Experiment Summary:



Rank	Name	Algorithm	Accuracy (Optimized)	Enhancements	Build time
★ 1	Pipeline 3	Gradient Boosting Classifier	0.868	HPO-1 FE	00:03:42
2	Pipeline 4	Gradient Boosting Classifier	0.868	HPO-1 FE HPO-2	00:02:01
3	Pipeline 1	Gradient Boosting Classifier	0.868	*None*	00:00:07
4	Pipeline 2	Gradient Boosting Classifier	0.868	HPO-1	00:00:32
5	Pipeline 8	XGB Classifier	0.860	HPO-1 FE HPO-2	00:17:35
6	Pipeline 7	XGB Classifier	0.857	HPO-1 FE	00:38:17
7	Pipeline 6	XGB Classifier	0.856	HPO-1	00:04:28
8	Pipeline 5	XGB Classifier	0.852	*None*	00:00:02

Telecom customer churn prediction using Watson Auto AI

- ## Pipeline comparison and leaderboard:



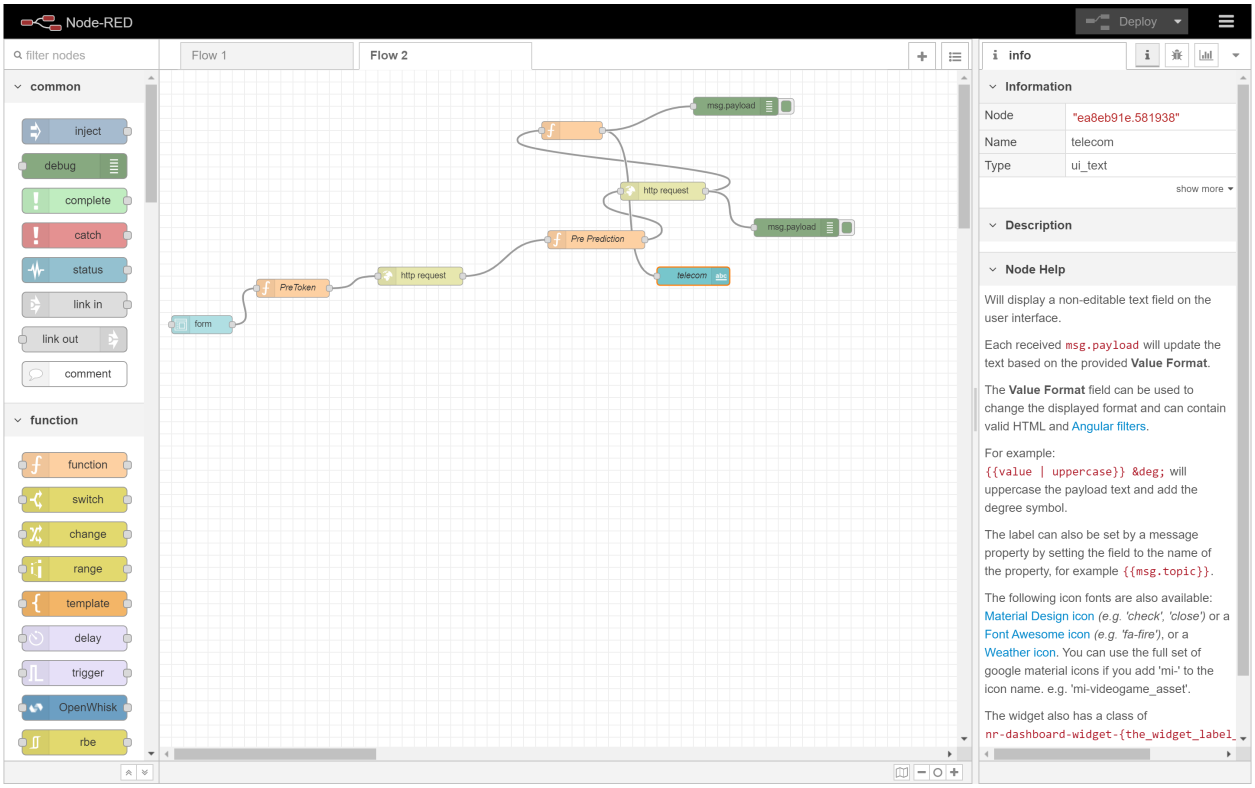
Rank	Name	Algorithm	Accuracy (Optimized)	Average pre...	F1	Log loss	Normalized ...	Precision	Recall
★ 1	Pipeline 3	Gradient Boosting Classifier	0.868	0.703	0.603	0.334	0.460	0.782	0.491
2	Pipeline 4	Gradient Boosting Classifier	0.868	0.703	0.603	0.334	0.460	0.782	0.491
3	Pipeline 1	Gradient Boosting Classifier	0.868	0.704	0.603	0.334	0.463	0.778	0.492
4	Pipeline 2	Gradient Boosting Classifier	0.868	0.704	0.603	0.334	0.463	0.778	0.492
5	Pipeline 8	XGB Classifier	0.860	0.679	0.598	0.384	0.466	0.717	0.513
6	Pipeline 7	XGB Classifier	0.857	0.671	0.591	0.397	0.457	0.704	0.510
7	Pipeline 6	XGB Classifier	0.856	0.675	0.605	0.393	0.475	0.684	0.542
8	Pipeline 5	XGB Classifier	0.852	0.705	0.619	0.357	0.513	0.649	0.59[illegible]

Watson AutoAI provide Artificial Intelligence platform for implementing different algorithms.After successful implementation of algorithm , IBM Cloud also provide different services for hosting your algorithm through web apps. One such platform is provided through the Node Red App.

It is built upon Nodejs for creating web apps.

Telecom customer churn prediction using Watson Auto AI

- Node-red flow:

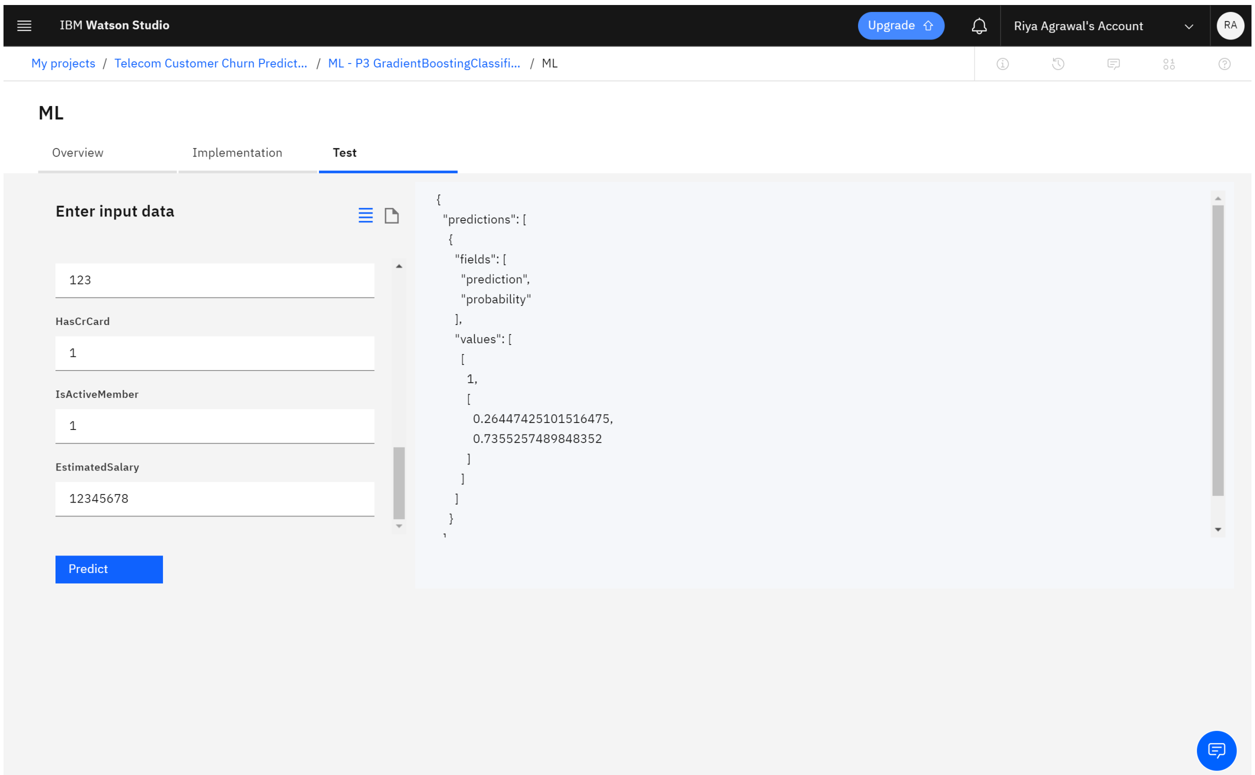


EXPERIMENTAL INVESTIGATION:

While working on the implementation of your Machine Learning or Artificial Intelligence Project , IBM Cloud also provides very essential feature to test your implementation before you can deploy it , wherever you wanted to be. In context with the Telecom Customer Churn Prediction , after creating and saving of the best fit algorithm based on minimum RMSE(Root Mean Squared Error), we test our Watson AutoAI model :

Telecom customer churn prediction using Watson Auto AI

- Test run of the project:



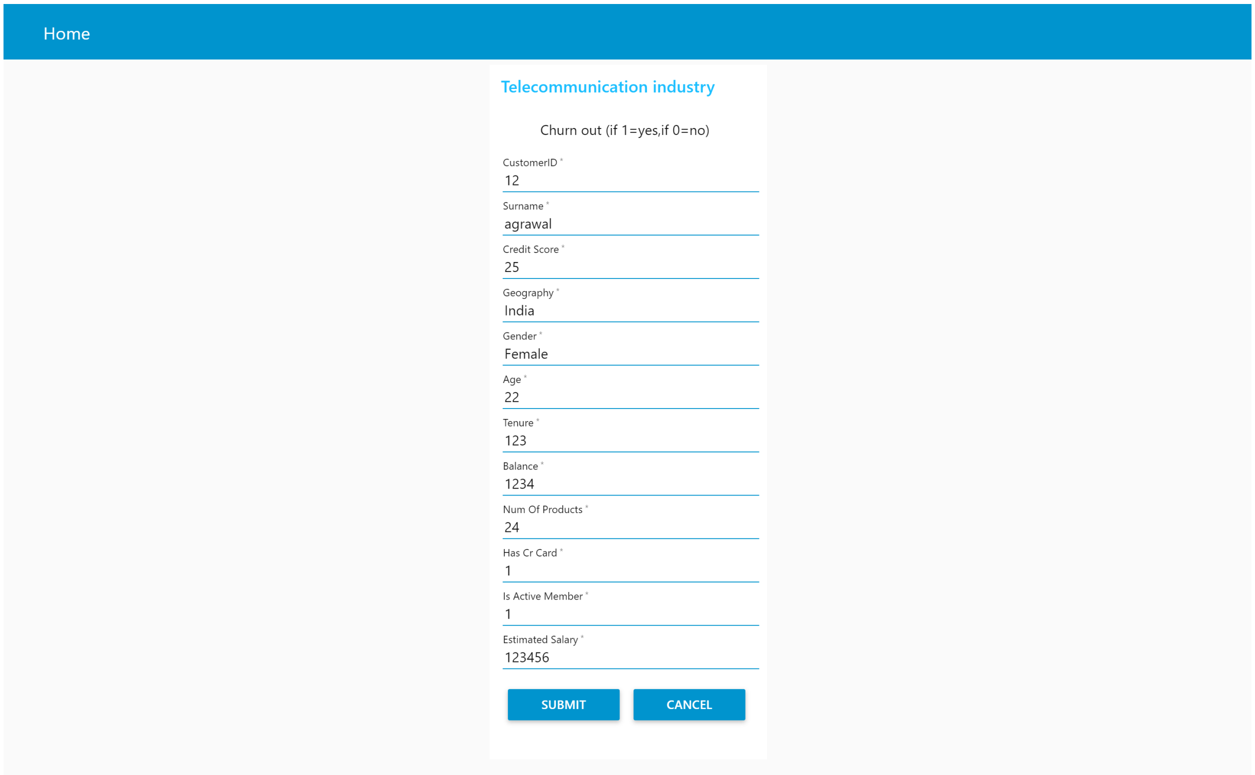
In the above test run it is quite clear that on the basis of inputed data this model predicts the probability of a customer that if a person are closer or not to churn out from the particular telecom company.

RESULT:

By using the IBM Cloud's Watson AutoAI , we are able to model our given dataset into a good predictable Machine Learning Algorithm and by the help of NodeRed service we are able to turn our Machine Learning Algorithm into a beautiful Web app:

Telecom customer churn prediction using Watson Auto AI

- Working web app:



ADVANTAGES AND DISADVANTAGES:

➤ Advantages:

- Access all the relevant data seamlessly and quickly.
- Segment the customers based on behavior and demographics to improve retention.
- Deliver tailored promotions and offers to positively influence their behavior.
- Minimize acquisition costs and increases marketing efficiency.
- Keep customers engaged and loyal.
- Predicting customers overall satisfaction as well as their experience with service quality.
- Identifying potential network issues, competitive threats, and at-risk customers.
- Identifying the negative customer experience trends and reducing attrition levels.
- Building a robust predictive model and gathering data.
- Creating new opportunities for cross-selling and up selling.

Telecom customer churn prediction using Watson Auto AI

- **Disadvantages:**

● A customer’s lifetime value and the growth of the business maintains a direct relationship between each other i.e., more chances that the customer would churn, the less is the potential for the business to grow.
● Even a good marketing strategy would not save a business if it continues to lose customers at regular intervals due to other reasons and spend more money on acquiring new customers who are not guaranteed to be loyal.
● There is a lot of debate surrounding customer churn and acquiring new customers because the former is much more cost-effective and ensures business growth.
● Thus companies spend almost seven times more effort, and time to retain old customers than acquire a new one.

The global value of a customer lost is nearly two hundred, and forty-three dollars which makes churning a costly affair for any business.

APPLICATIONS:

● Telephone service companies
● Internet service providers
● Pay TV companies
● Insurance firms
● alarm monitoring services
● Predicting Insurance Customer Churn
● Lifetime Value

Telecom customer churn prediction using Watson Auto AI

CONCLUSION:

The main purpose of the application is to build a Machine Learning model to predict the customer churn using IBM Watson AutoAI 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. Companies usually make a distinction between voluntary churn and involuntary churn. Voluntary churn occurs due to a decision by the customer to switch to another company or service provider, involuntary churn occurs due to circumstances such as a customer's relocation to a long-term care facility, death, or the relocation to a distant location. In most applications, involuntary reasons for churn are excluded from the analytical models. Analysts tend to concentrate on voluntary churn, because it typically occurs due to factors of the company-customer relationship which company's control, such as how billing interactions are handled or how after-sales help is provided. Predictive analytics use churn prediction models that predict customer churn by assessing their propensity of risk to churn. Since these models generate a small prioritized list of potential defectors, they are effective at focusing customer retention marketing programs on the subset of the customer base who are most vulnerable to churn.

FUTURE SCOPE:

Telephone service companies, Internet service providers, pay TV companies, insurance firms, and alarm monitoring services, often use customer attrition analysis and customer attrition rates as one of their key business metrics because the cost of retaining an existing customer is far less than acquiring a new one. Companies from these sectors often have customer service branches which attempt to win back defecting clients, because recovered long-term customers can be worth much more to a company than newly recruited clients. This can help companies in fetching information of active subscribers, collecting data and storing them.

Telecom customer churn prediction using Watson Auto AI

BIBLIOGRAPHY:

Pavan Raj. “Telecom Customer Churn Prediction.”
https://www.kaggle.com/pavanraj159/telecom-customer-churn-prediction

Quantzig. “Churn Analysis Solution Helps in Predicting Customer Satisfaction”
https://www.businesswire.com/news/home/20170911005499/en/Churn-Analysis-Solution-Helps-Predicting-Customer-Sasfacon#:~:text=Predicting%20customers%20overall%20satisfaction%20as,predictive%20model%20and%20gathering%20data

www.wikipedia.org

APPENDIX:

- **Source Code :**

[{"id":"a3feda27.814238","type":"tab","label":"Flow 4","disabled":false,"info":""},{"id":"b4bfeca.514f01","type":"function","z":"a3feda27.814238","name":"PreToken","func":"global.set(\"rn\",msg.payload.rn)\nglobal.set(\"cid\",msg.payload.cid)\nglobal.set(\"sn\",msg.payload.sn)\nglobal.set(\"cs\",msg.payload.cs)\nglobal.set(\"geo\",msg.payload.geo)\nglobal.set(\"gen\",msg.payload.gen)\nglobal.set(\"a\",msg.payload.a)\nglobal.set(\"t\",msg.payload.t)\nglobal.set(\"bal\",msg.payload.bal)\nglobal.set(\"nop\",msg.payload.nop)\nglobal.set(\"hcc\",msg.payload.hcc)\nglobal.set(\"iam\",msg.payload.iam)\nglobal.set(\"es\",msg.payload.es)\nvar apikey=\"G2YjQXh9q65ZFIEgIcq8We1qu_3LFzl7b2X74imme3_V\";\nmsg.headers={\"content-type\":\"application/x-www-form-urlencoded\"}\nmsg.payload={\"grant_type\":\"urn:ibm:params:oauth:grant-type:apikey\",\"apikey\":apikey}\nreturn msg;","outputs":1,"noerr":0,"x":260,"y":400,"wires":[["77a2fcb9.261404"]]},{"id":"77a2fcb9.261404","type":"http request","z":"a3feda27.814238","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://iam.cloud.ibm.com/identity/token","tls":"","persist":false,"proxy":"","authType":"","x":459,"y":388.9999952316284,"wires":[["dc11121f.bf90f"]]},{"id":"ad09b3a4.1415c","type":"inject","z":"a3feda27.814238","name":"","topic":"","payload":"","payloadType":"date","repeat":"","crontab":"","once":false,"onceDela

Telecom customer churn prediction using Watson Auto AI

y":0.1,"x":105.5,"y":313.99999809265137,"wires":[["b4bfeca.514f01"]]},{"id":"7ed51949.7da138","type":"debug","z":"a3feda27.814238","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":898.0000114440918,"y":130.99999713897705,"wires":[]},{"id":"dc11121f.bf90f","type":"function","z":"a3feda27.814238","name":"Pre Prediction","func":"var rn = global.get('rn')\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 a = global.get('a')\nvar t = global.get('t')\nvar bal = global.get('bal')\nvar nop = global.get('nop')\nvar hcc = global.get('hcc')\nvar iam = global.get('iam')\nvar es = global.get('es')\nvar token=msg.payload.access_token\nvar instance_id=\"d08d5a8e-a93e-4a8f-9523-8afd57b0f910\"\nmsg.headers={'Content-Type': 'application/json',\"Authorization\":\"Bearer \"+token,\"ML-Instance-ID\":instance_id}\nmsg.payload={\"input_data\": [{\"fields\": [\"Row Number\", \"Customer Id\", \"Surname\", \"Credit Scode\", \"Geography\", \"Gender\", \"Age\", \"Tenure\", \"Balance\", \"No. of Products\", \"Has Credit Card\", \"Is Active Member\", \"Estimated Salary\"], \"values\": [[rn,cid,sn,cs,geo,gen,a,t,bal,nop,hcc,iam,es]]}]}\nreturn msg;","outputs":1,"noerr":0,"x":656.0000076293945,"y":351.9999990463257,"wires":[["dfc852fa.3849a"]]},{"id":"dfc852fa.3849a","type":"http request","z":"a3feda27.814238","name":"","method":"POST","ret":"obj","paytoqs":false,"url":"https://eu-gb.ml.cloud.ibm.com/v4/deployments/15233374-e8ed-463d-b0a5-1a4e546b0c78/predictions","tls":"","persist":false,"proxy":"","authType":"","x":749.5000076293945,"y":295.9999990463257,"wires":[["5ce3bde4.2bd1a4","2a04d67b.800e0a"]]},{"id":"97a51d29.6113","type":"ui_form","z":"a3feda27.814238","name":"","label":"","group":"4f4248b7.dec148","order":0,"width":0,"height":0,"options":[{"label":"Row Number","value":"rn","type":"number","required":true,"rows":null},{"label":"Customer Id","value":"cid","type":"number","required":true,"rows":null},{"label":"Surname ","value":"sn","type":"text","required":true,"rows":null},{"label":"Credit Score","value":"cs","type":"number","required":true,"rows":null},{"label":"Geography","value":"geo","type":"text","required":true,"rows":null},{"label":"Gender","value":"gen","type":"text","required":true,"rows":null},{"label":"Age","value":"a","type":"number","required":true,"rows":null},{"label":"Tenure","value":"t","type":"number","required":true,"rows":null},{"label":"Balance","value":"bal","type":"number","required":true,"rows":null},{"label":"No. of Products","value":"nop","type":"number","required":true,"rows":null},{"label":"Has Credit Card","value":"hcc","type":"number","required":true,"rows":null},{"label":"Is Active Member","value":"iam","type":"number","required":true,"rows":null},{"label":"Estimated Salary","value":"es","type":"number","required":true,"rows":null}],"formValue":{"rn":"","cid":"","sn":"","cs":"","geo":"","gen":"","a":"","t":"","bal":"","nop":"","hcc":"","iam":"","es":""},"payload":"","submit":"submit","cancel":"cancel","topic":"","x":51,"y":442.9999952316284,"wires":[["b4bfeca.514f01"]]},{"id":"3a82ace6.a50b94","type":"ui_text","z":"a3feda27.814238","group":"4f4248b7.dec148","order":1,"width":0,"height":0,"name":"","label":"Churn out (if 1=yes,if

Telecom customer churn prediction using Watson Auto AI

0=no)","format":"{{msg.payload}}","layout":"row-spread","x":847.6000366210938,"y":448.20000743865967,"wires":[]},{"id":"5ce3bde4.2bd1a4","type":"function","z":"a3feda27.814238","name":"","func":"msg.payload=msg.payload.predictions[0].values[0][0]\nreturn msg;","outputs":1,"noerr":0,"x":609.6000366210938,"y":191.60000133514404,"wires":[["7ed51949.7da138","3a82ace6.a50b94"]]},{"id":"2a04d67b.800e0a","type":"debug","z":"a3feda27.814238","name":"","active":true,"tosidebar":true,"console":false,"tostatus":false,"complete":"payload","targetType":"msg","x":910,"y":380,"wires":[]},{"id":"4f4248b7.dec148","type":"ui_group","z":"","name":"Customer Churn Prediction","tab":"20a7cc05.602764","order":1,"disp":true,"width":"6","collapse":false},{"id":"20a7cc05.602764","type":"ui_tab","z":"","name":"Home","icon":"dashboard","disabled":false,"hidden":false}]