Use case 4 :

Customer Churn : Telecom Dataset

A telecommunications company is concerned about the number of customers leaving their landline business for cable competitors. They need to understand who is leaving. Imagine that you’re an analyst at this company and you have to find out who is leaving and why.

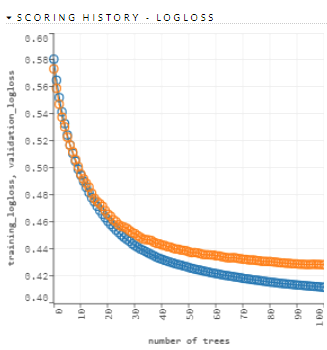
This data set provides info to help you predict behavior to retain customers. We can analyze all relevant customer data and develop focused customer retention programs.

The data set includes information about:

* Customers who left within the last month – the column is called Churn
* Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies
* Customer account information – how long they’ve been a customer, contract, payment method, paperless billing, monthly charges, and total charges
* Demographic info about customers – gender, age range, and if they have partners and dependents

The data was split into 3 parts : Training set(50%), Validation set(25%) and Testing set(25%)

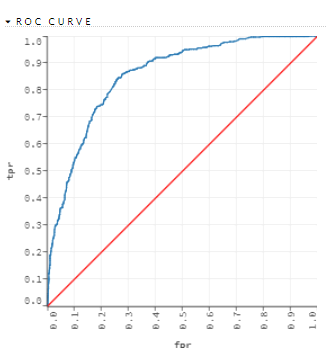
Scoring History shows the improvement of the data considering the error rate. The figure below compares the training and the validation set's error rate(logloss).



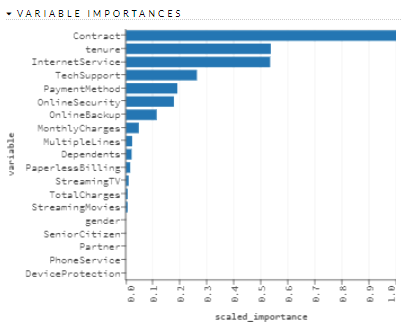
Here, the error rate decreased fro 58.07% to 41.19% , I.e., by 17% which is a really good improvement in the modelling of the data.

The AUC is also increased from 50% to 85.38%, which means the model became soundly predictable.

ROC curve :

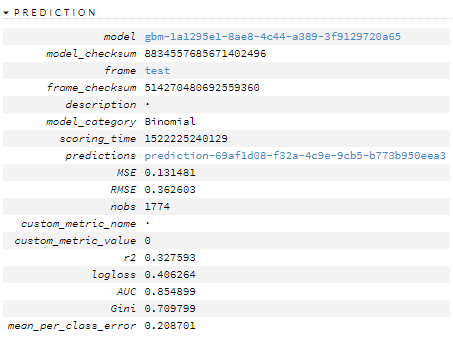


The ROC curve represents the true positive and false positive rate in the prediction result, which gives us an idea about how far is the prediction result agreeing to our original data.

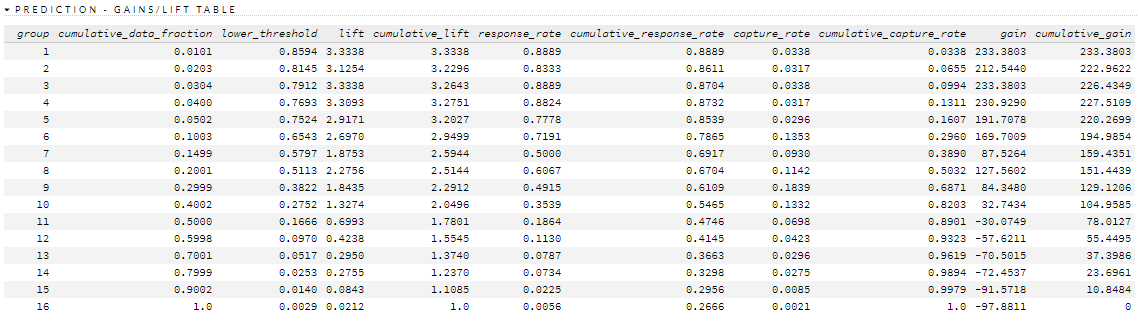


The variable importance gives us the knowledge about which all columns are useful for the prediction and mostly used for the purpose of feature engineering.

We can clearly see Contract, Tenure, Online Security are highly driving the churn rate.



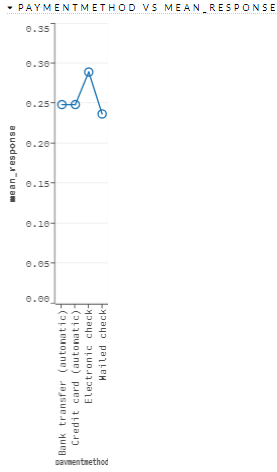
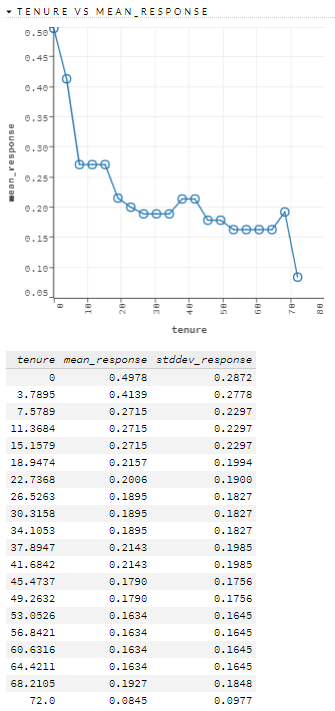
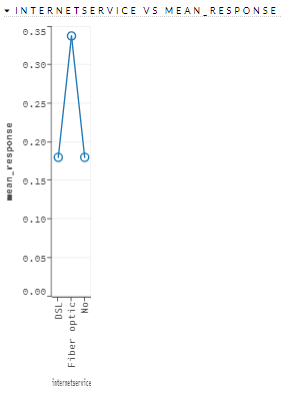
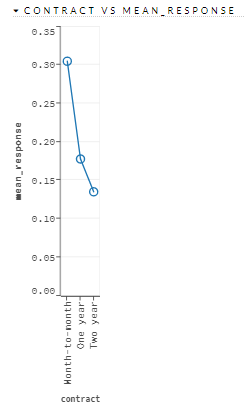
In the above stated figure, we are able to get a summary of the mean square error which is very low, which shows the predicted data has less error. R2 is 32.75% which shows the data can be improved by 32.75% and the high AUC percent shows the model is highly predictable.

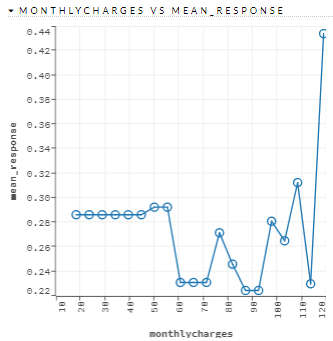
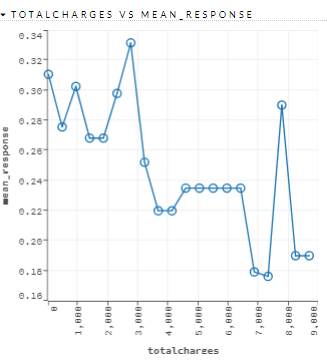


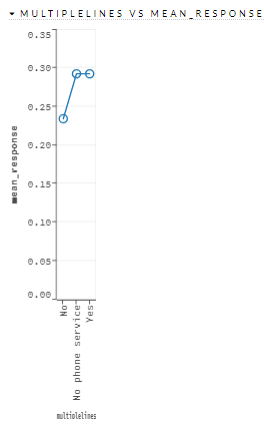
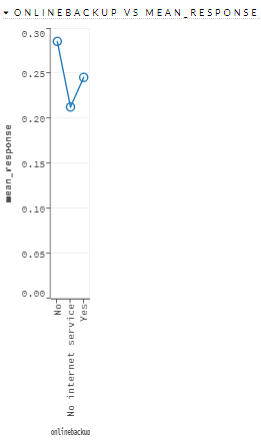
The above table states us the gains/lift table which shows us which groups are more likely to churn out of the service. More the lift, more is the variance, more is the group likely to churn. Here, we may observe that groups 1 to 5 have a high lift rate which shows they are more likely to churn out. Groups 6 to 10 are moderately likely and groups 11 to 16 are least likely to churn.

Hence, people under groups 1 to 5 should be kept under high focus to save them from churning out.

Now, we shall visualize each factors of the data that shall give us an insight about where the problem is, which is resulting in the churning out of the users.







In the above, visualization, we can clearly see, Contract, Internet services, Tenure, Total Charges are driving the Churn rate of the users.