TELCO CUSTOMER

CHURN PREDICTION

Joey Christian S

BUSINESS CONTEXT

A company engaging in telecommunications would like to acquire customers that are loyal to the company (stayed for long term). From a list of customers, the company would like to figure out whether a customer would stay for a long period of time or for just a short period of time.

In maximizing the company's profit, Customer Relationship Management (CRM) division would like to evaluate the data on a customer's churn possibilities based on a few features, which would provide beneficial decisions in detecting whether a customer would churn or not.

So far, the company predicts all customer to churn, therefore many financial losses are formed. It is also known that the cost of obtaining a new customer are 5 times more than the cost of maintaining a customer.

Cost of obtaining new customers also known as Retention Cost, while cost of acquiring new customers are known acquisition cost.

WRONG PREDICTION CONSEQUENCES

False Positive

CHURN, BUT STAYED

A retention cost must be issued to maintain customer from churning out, such as discounts, added services, etc.

However, the customer stayed which results in a waste of money.

False Negative

STAYED, BUT CHURN

Company did no treatment, knowing that customers would stay.

Unexpectedly, customer churns out. Therefore the company must issue an acquisition cost for new customers.

Based from the loss possibilities, the best option would be to reduce the value of False Positive. Therefore f2 score would be used as the best metric.

01 DATA

Understanding data used in this analysis

DATASET USED

TELCO CUSTOMERS

Rows in this dataset are considered to be the company's customers.

CUSTOMER TENURE

Months on how long a customer stayed within the company.

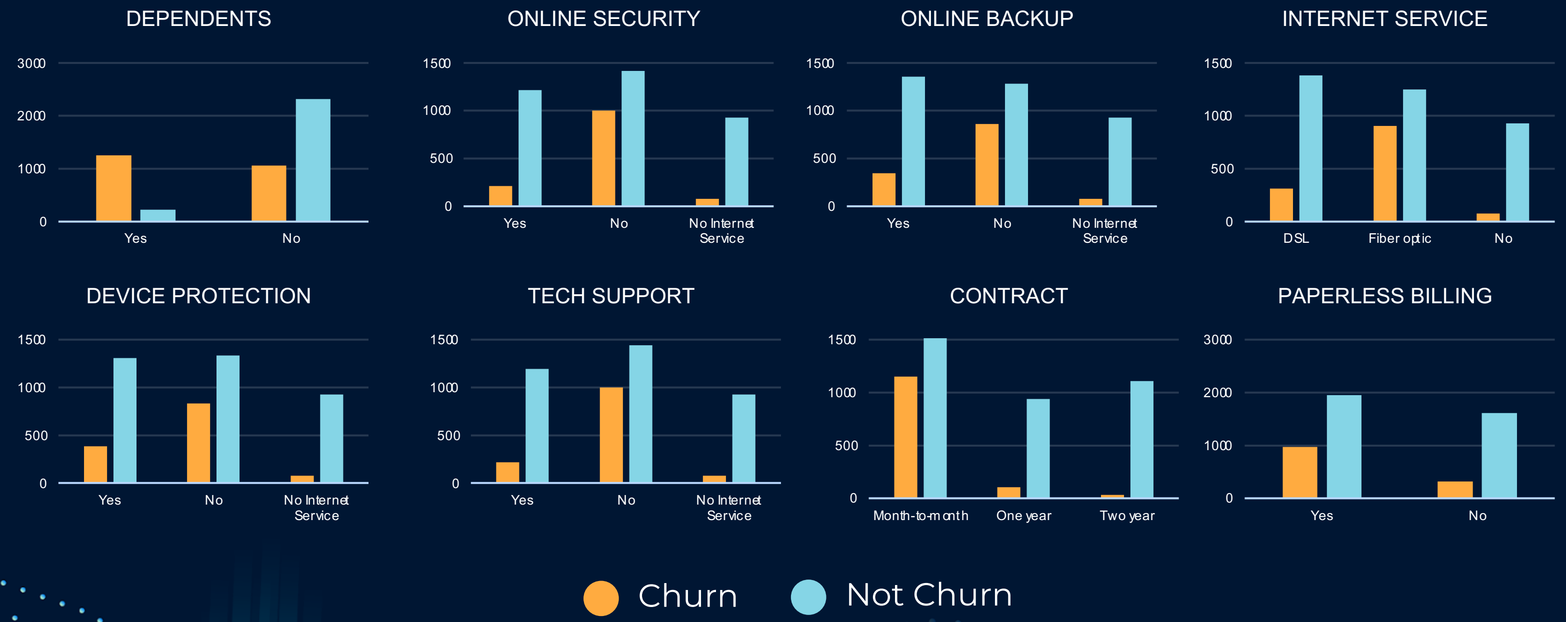
TELCO SERVICES

Online Security, Online Backup, Internet Service, Device Protection, Tech Support, Contract, Paperless Billing

MONTHLY CHARGES

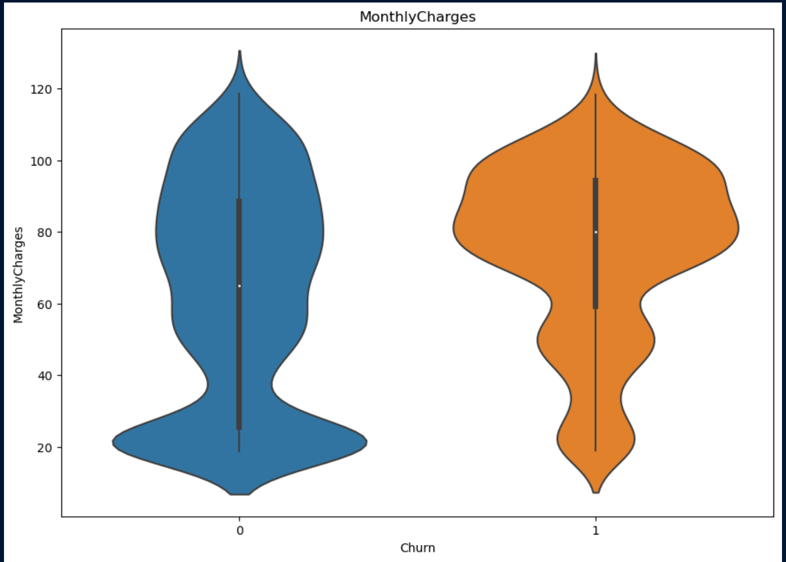
Amount of charges for services on a monthly basis.

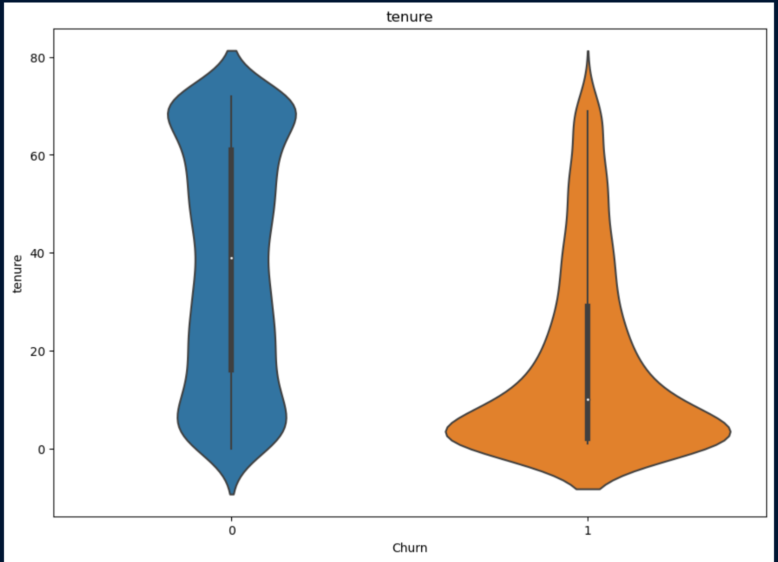
EDA



Based on the graphs above, customers with no dependents, no online security, no online backup, fiber optic internet service, no device protection, no tech support, month-to-month contract and paperless billing shows a tendency to churn out.

EDA





Graphs above shows that customers who churn out are the ones with low tenure or new customers and customers that pay a high amount for their monthly charges.

02

PROCESSING

Data cleaning and preprocessing

DATA HANDLING

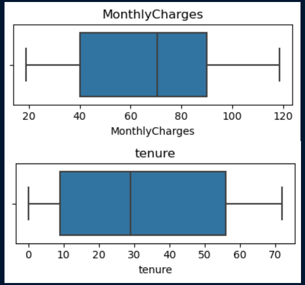
DUPLICATES

```
1  df.duplicated().sum()
✓ 0.3s
77
```

77 duplicated rows.

4853 rows left

OUTLIERS



MISSING VALUES

```
Dependents          0
tenure              0
OnlineSecurity      0
OnlineBackup        0
InternetService     0
DeviceProtection    0
TechSupport         0
Contract            0
PaperlessBilling    0
MonthlyCharges      0
Churn               0
dtype: int64
```

DATA PREPARATION

1. **OneHotEncoder()**
   Dependents, Paperless Billings, Internet Service, Contract

2. **OrdinalEncoder()**
   Online Security, Online Backup, Device Protection, Tech Support

3. **RobustScaler()**
   All numerical and ordinal columns

RESAMPLING

```
0    0.734597
1    0.265403
Name: Churn, dtype: float64
```

IMBALANCE

Customers who doesn’t churn are considered to be the majority class, with a proportion of 73% against 26%.

OVERSAMPLING

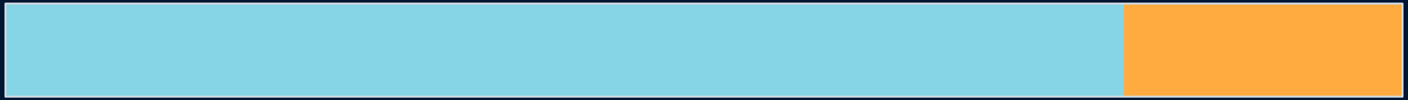
Oversampling would duplicate rows that are considered to be the minority class randomly.

03

MODELING

Finding Algorithms and Analytics

DATA SPLITTING



80%

Data used for training

20%

Data used for testing

For this case (classification), splitting would require the parameter stratify which is useful in maintaining the proportion of customers who churn and who doesn't.

CROSS VALIDATION

ALGO	MEAN	STD
AdaBoostClassifier	0.728	0.015
DecisionTreeClasssifier	0.720	0.013
VotingClassifier	0.720	0.013
LogisticRegression	0.710	0.018
GradientBoosingClassifier	0.700	0.018
StackingClassifier	0.674	0.024
KNeighbors	0.640	0.023
XGBClassifier	0.620	0.009
BaggingClassifier	0.618	0.012
RandomForestClassifier	0.552	0,012

Based on the table, the best model would be adaboost with the highest score and decisiontree with the lower std.

ADABOOST HYPERPARAMS

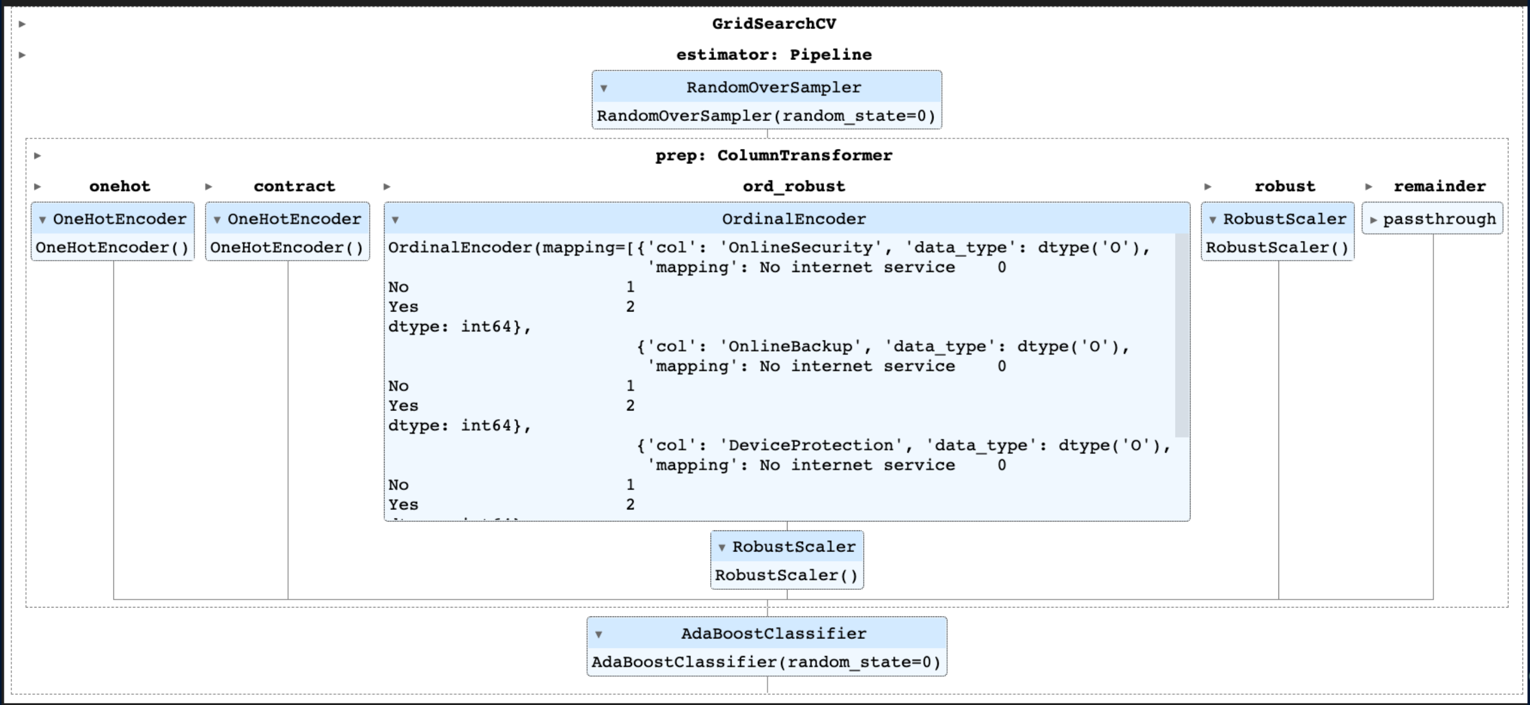
Hyperparamters that are tuned:

- Resampling: oversampler, udnersampler, SMOTE, nearmiss, None

- Ordinal encoding map: ordinal_map1, ordinal_map2, ordinal_map3
- Scaling for ordinal encoding: robust, minmax, standard
- Contract column: pipe_contract1, pipe_contract2
- Onehot encoding: onehotencoder, binaryencoder
- Scaler: robust, minmax, standard

- n_estimators: 100, 200, 300, 400, 500,
- Learning rate: 0.05, 0.25, 0.5, 1, 2

BEST ADABOOST MODEL



DECISION TREE HYPERPARAMS

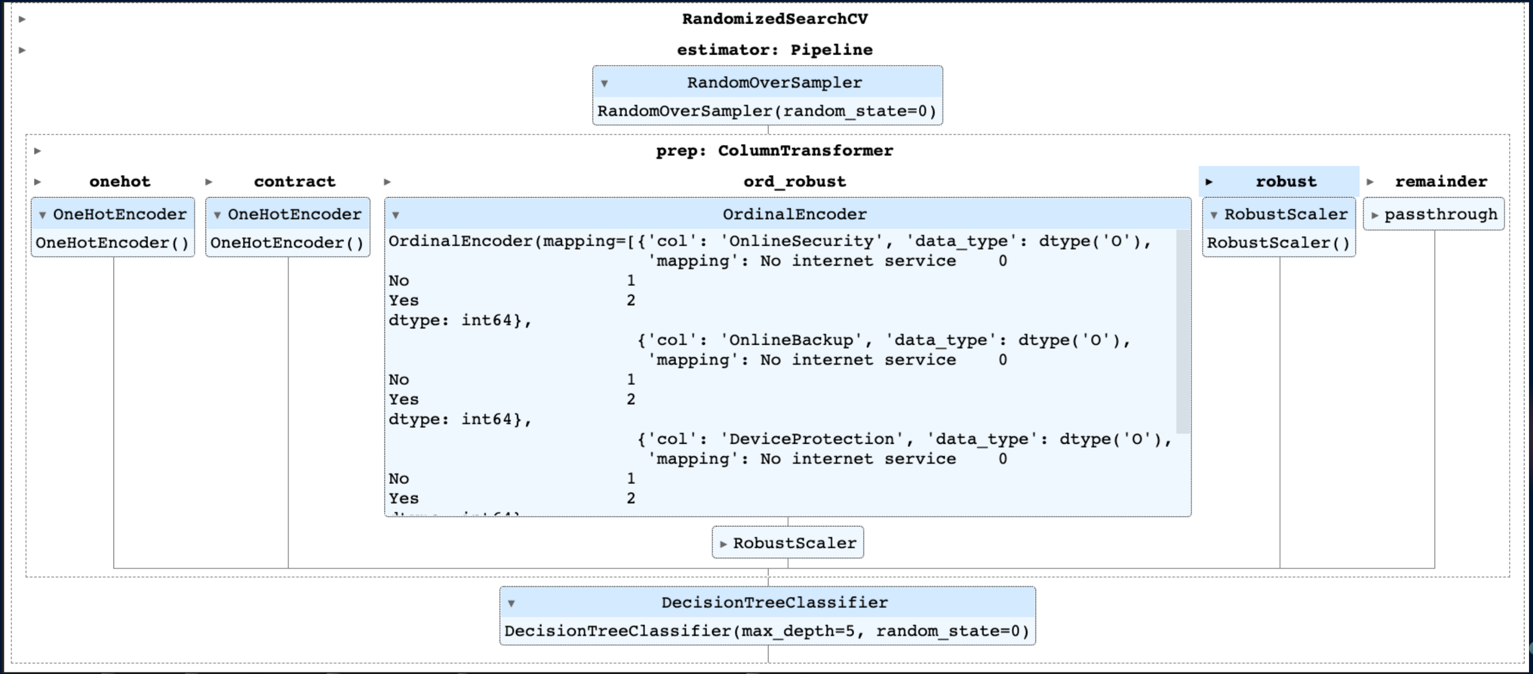
Hyperparamters that are tuned:

- Resampling: oversampler, udnersampler, SMOTE, nearmiss, None

- Ordinal encoding map: ordinal_map1, ordinal_map2, ordinal_map3
- Scaling for ordinal encoding: robust, minmax, standard
- Contract column: pipe_contract1, pipe_contract2
- Onehot encoding: onehotencoder, binaryencoder
- Scaler: robust, minmax, standard

- Criterion: fgini, entropy
- Min_samples_leaf: range(2, 50, 2)
- Min_samples_split : range(2,100,2)
- Max_depth: range(2,50,2)

BEST DECISIONTREE MODEL



TUNING RESULTS

0.733

Adaboost best score

0.730

Decisiontree best score

Therefore, the model that will be used for this machine learning would be adaboost.

Adaboost is an algorithm that uses decision tree with just one splitting. Those underfitting models would then be used in another decision tree splitting. This iteration happens until the number of estimators is reached. In short adaboost would study data from obtaining weak models and turning them into strong ones.

PREDICTING TO TEST SET

0.545

Model before tuning

0.752

Model after tuning

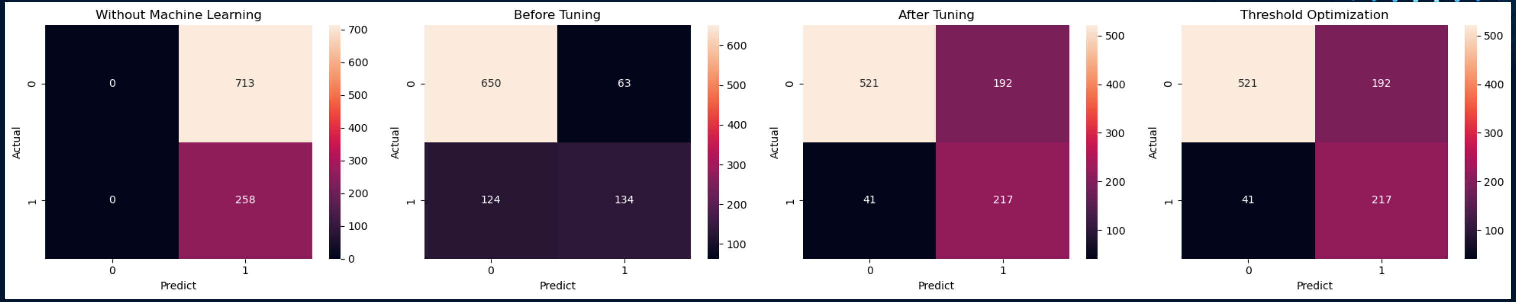
Scores are obtained from f2 score, which means the higher the number, the better that model is In predicting results. Based from the two scores above, it can be concluded that the model's performance would be better after tuning.

04

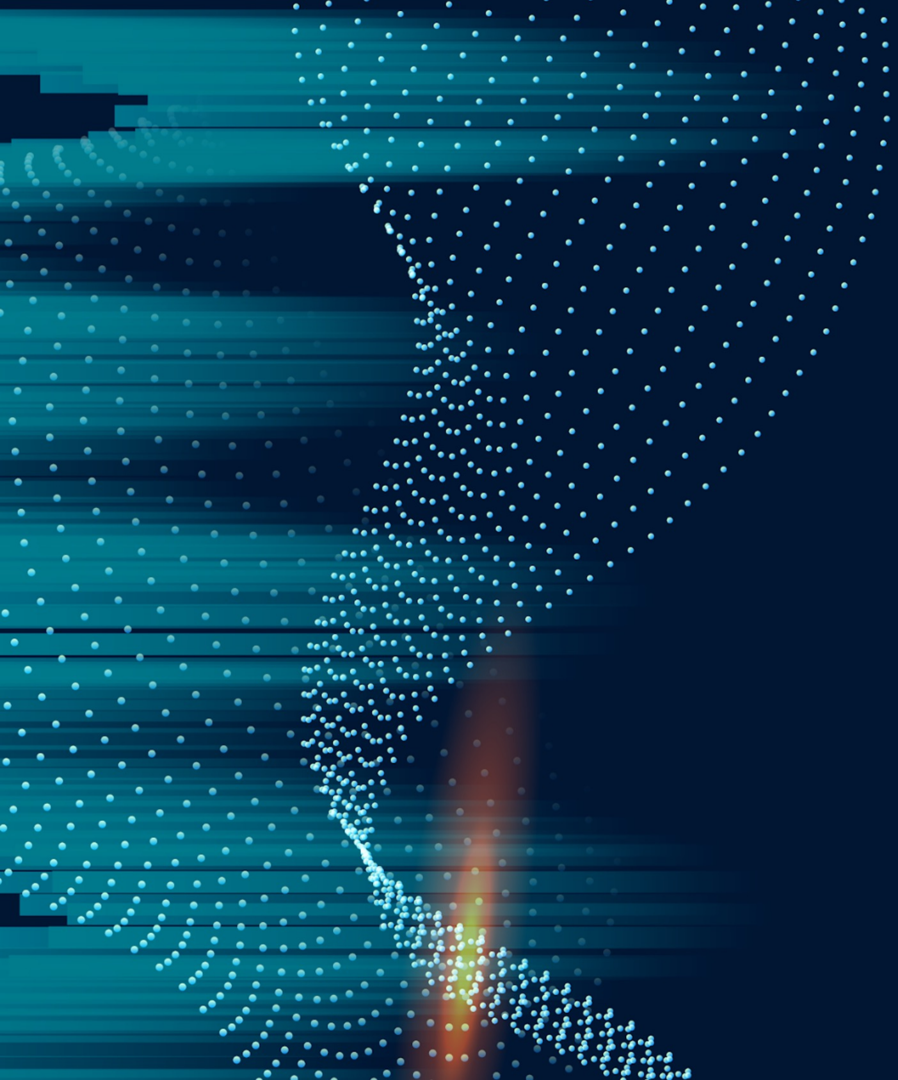
CONCLUSION

Best results and business loss calculations

TOTAL LOSS



Based on the graph, the total loss before the existence of machine learning would be very high, as this is caused by the amount of False Positive that are more than 700. After using machine learning, the amount of false positives experience a drop. However after tuning the machine learning algorithm, a more beneficial results are achieved.



142.600

Total loss without machine learning

79.400

Total loss with machine learning

44%

Total loss prevented with machine learning

05

RECCOMENDATIONS

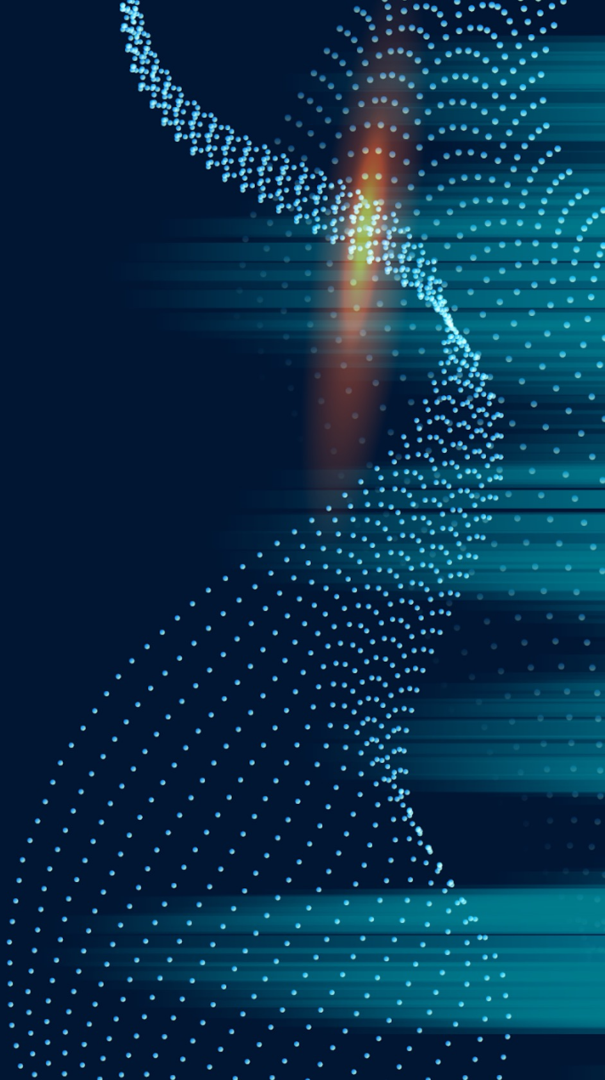
For both the model and the business.

MODEL RECOMENDATION

1. **Gridsearch instead of Randomsearch**
   Due to the limits of the hardware used, the optimal model might not be achieved.

2. **Dropping dependents column**
   Whether a customer have a dependents or not, it doesn’t affects the prediction results significantly.

3. **Adding joined date column**
   This column would aid in analyzing does a customer joined because of a certain quarter or that customer wants to join without a certain quarter.



BUSINESS RECOMENDATION

01

MONTHLY CHARGES

Customers with high monthly charges tends to churn out. One of the ways to ensure customer don’t churn out would be to give discounts, budlings, and so on.

02

TENURE

Customers with low tenure values tends to churn out. Therefore to prevent customers churning out would be to lower subscription after a certain period of time, create a membership, and so on.

THANKYOU

For your attention