

Telecom customer Churn Documentation

About Dataset

Context

"Predict behavior to retain customers. You can analyze all relevant customer data and develop focused customer retention programs." [IBM Sample Data Sets]

Content

Each row represents a customer, and each column contains the customer's attributes described in the column Metadata.

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

New version from IBM:

<https://community.ibm.com/community/user/businessanalytics/blogs/steven-macko/2019/07/11/telco-customer-churn-1113>

And here is the link to the dataset on Kaggle:

<https://www.kaggle.com/datasets/blatchar/telco-customer-churn>

Approach

First of all the aim is to reduce customer churn.

The approach is as follows:

- After downloading the dataset from Kaggle we perform all the necessary Exploratory Data Analysis (EDA), which is an analysis approach that identifies general patterns in the data. These

patterns include outliers and features of the data that might be unexpected. EDA is an important first step in any data analysis.

- From EDA we found:
 - TotalCharges feature(column) is missing 11 besides that dataset is not Null.
- EDA Graphs Shows results:

```
* **Total Charges:**  
  * Increases with higher tenure and monthly charges.  
  * This implies longer stays and higher monthly charges lead to  
    higher overall spending.  
  
* **Monthly Charges:**  
  * Shows some increase with both senior citizen status and  
    tenure.  
  * This suggests potentially higher monthly plans or increased  
    usage for senior citizens who tend to stay longer.  
  
* **Senior Citizen and Contract Preference:**  
  * The data doesn't definitively show a preference for long-term  
    contracts by senior citizens.  
  * Further analysis into contract types and tenure specifically  
    for senior citizens might be needed.
```

-
- Data is not equally distributed about 75% of the dataset represents 0 class (not Churns)
 - Which is solved by using the SMOTE algorithm (which uses semantic over-sampling.)
- EDA Shows that the dataset has many outliers and anomalies.
 - This is identified by using two techniques Plotting Box Plot (which is used for this purpose) and Isolation Forest (which is a very good model for finding and removing the anomalies in the dataset).
- Once the dataset is clean we use MinMax scalar from sklearn for Scaling our dataset.
- Here is an important thing to mention: after the data is cleaned and scaled between (0 and 1) we save the file in CSV format using the pandas function pd.to_csv.
 - it is the better approach for model training because we do not have to go through the same process again.
- All the Models in this task are trained using Randomized Search CV and Grid Searched CV.

In Randomized Search, a fixed number of parameter settings is sampled from the specified distributions. Python sci-kit-learn library implements Randomized Search in its RandomizedSearchCV function. This function needs to be used along with its parameters, such as estimator, param_distributions, scoring, n_iter, cv, etc.

What is the difference between randomized search CV and GridSearchCV?

The advantage of RandomizedSearchCV over GridSearchCV is that it allows you to explore a wider range of hyperparameter values more efficiently. Since it randomly samples combinations, it can be useful when you have a large search space or when the search space is not well-defined

About the Models

Here is the list of models used in this task.

1. Linear Regression
 - MSE: 0.1555317076877575 (because it works on continuous values).
 - In advance of Linear Regression SGD Regression is also used from sklearn.
 - The reason for using SGDRegressor: because Linear Regression doesn't have parameters of learning rate and Regularization.
2. Logistic Regression
 - Accuracy: 0.78 %
3. - AUC-ROC Score: 0.7778729912195355
4. Neural Network.
 - Test accuracy: 0.8985074758529663
 - Test loss: 0.25862714648246765
 - But this model is overfitted and other models are much better at predicting on unseen data.
5. Decision Tree
 - Accuracy: 0.8333333333333334
 - AUC-ROC Score: 0.8314379907583453
6. Random Forest
 - Accuracy: 0.8333333333333334
 - AUC-ROC Score: 0.8314379907583453
7. XGBoost
 - Accuracy: 0.8606965174129353

- AUC-ROC Score: 0.859578025927810
- This model is our highest-performing model.

key decisions

The Main Key Decisions made during this task are

- Use of SMOTE for class Imbalance
- Use of Isolated Forest Model For Anomaly Detection.
- Use XGBoost for our main model because of its Accuracy.
- Using Linear Regression for Feature Engineering.
 - Add a churn Risk column in our dataset.
- Use of Randomized and Grid Search Cross Validation Techniques
 - Because these techniques are proven by experts for best parameter finding.

Key Inside

This part is about the feature that affects the customer churn in our dataset:

The values against each feature tell us how much it affects the customer churn the higher the value the more it affects.

- | | |
|----------------------------------|----------|
| • InternetService_DSL | 0.021925 |
| • OnlineSecurity_Yes | 0.021933 |
| • OnlineBackup_Yes | 0.022390 |
| • StreamingTV_No | 0.022815 |
| • TechSupport_No | 0.024897 |
| • DeviceProtection_Yes | 0.025075 |
| • OnlineSecurity_No | 0.026525 |
| • PaperlessBilling_No | 0.029010 |
| • InternetService_Fiber optic | 0.032341 |
| • SeniorCitizen | 0.038273 |
| • Contract_One year | 0.043479 |
| • StreamingMovies_Yes | 0.044757 |
| • PaymentMethod_Electronic check | 0.047256 |
| • churn_risk | 0.058811 |

- Contract_Two year 0.059854
- Contract_Month-to-month 0.221449

How to improve the Churn

According to my analysis based on this dataset.

The customer is involved in a long contract one or two-year stay with the company longer.

Whereas the customer who acts month-to-month subscription tends to leave the company.

The monthly packet price increase causes the long-term package price to be high its any directly proportional relation.

The senior citizens tend to leave the company more than young people.

The people who buy internet and online security have more chances to stay longer.

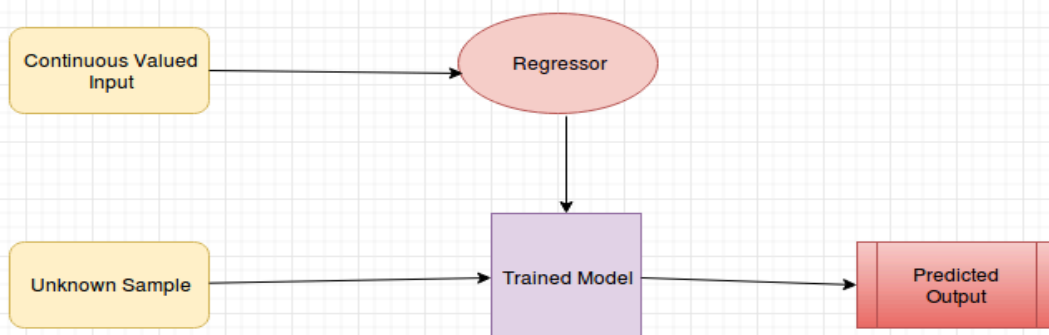
Now My suggestion on how to improve the Churn rate:

- First introduce a package that focuses on older people.
- Encourage customers to buy one or two packages/ contracts.
- Focus on decreasing the monthly subscription price.
- Encourage people to use the internet and online security

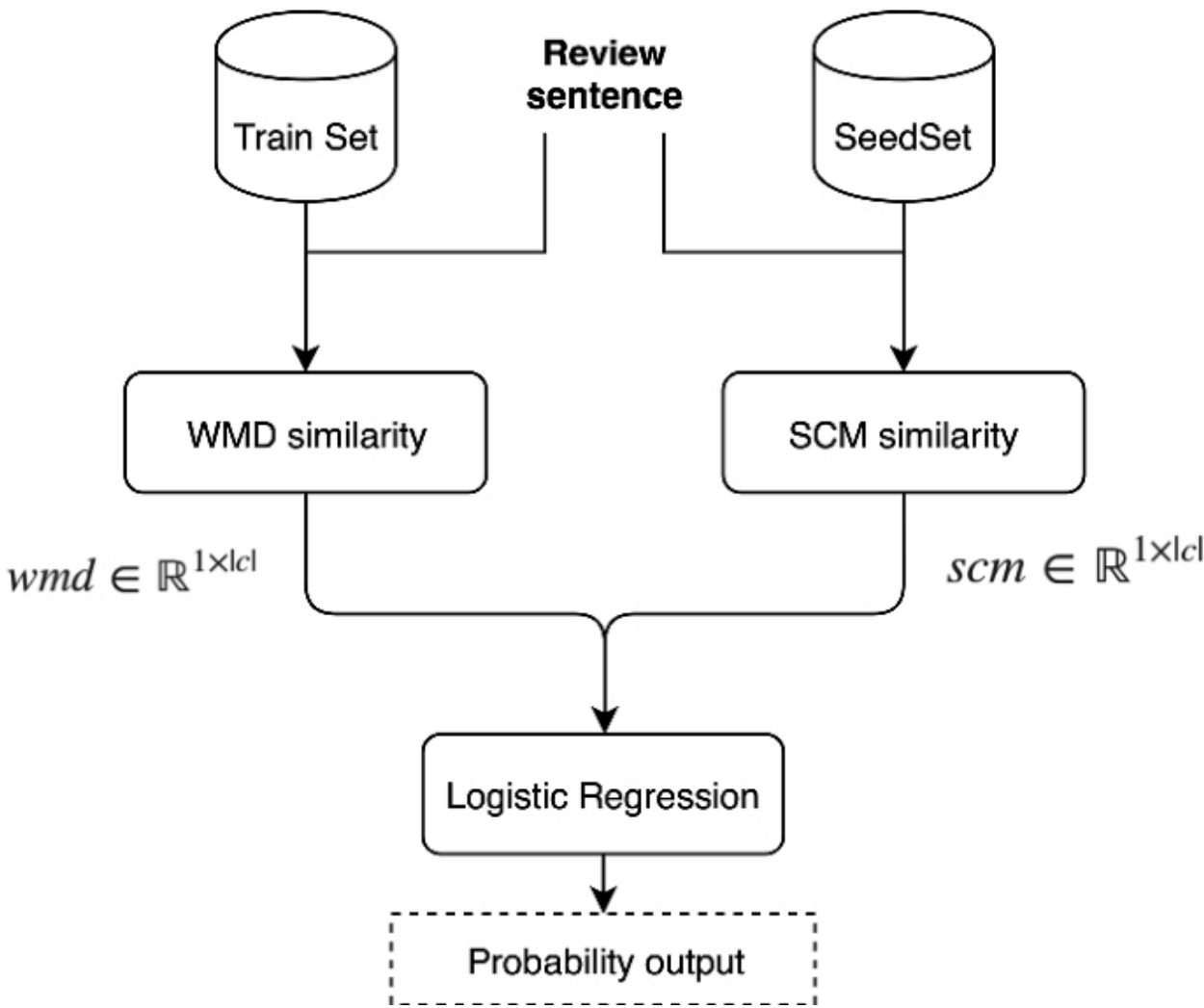
An interesting fact is that gender those not play any role in this Churn rate meaning males and females behave the same.

Model Architecture

Linear Regression and Logistic Regression: have the same Architecture except Logistic Regression uses the Sigmoid Function on its output.



The above diagram shows the architecture of linear Regression.



The above diagram shows the logistic regression model. When the sigmoid function is applied to Linear Regression it gives the output in probability of classes.

Neural Network:

Here is the Model Architecture Summary of the model. I used this task.

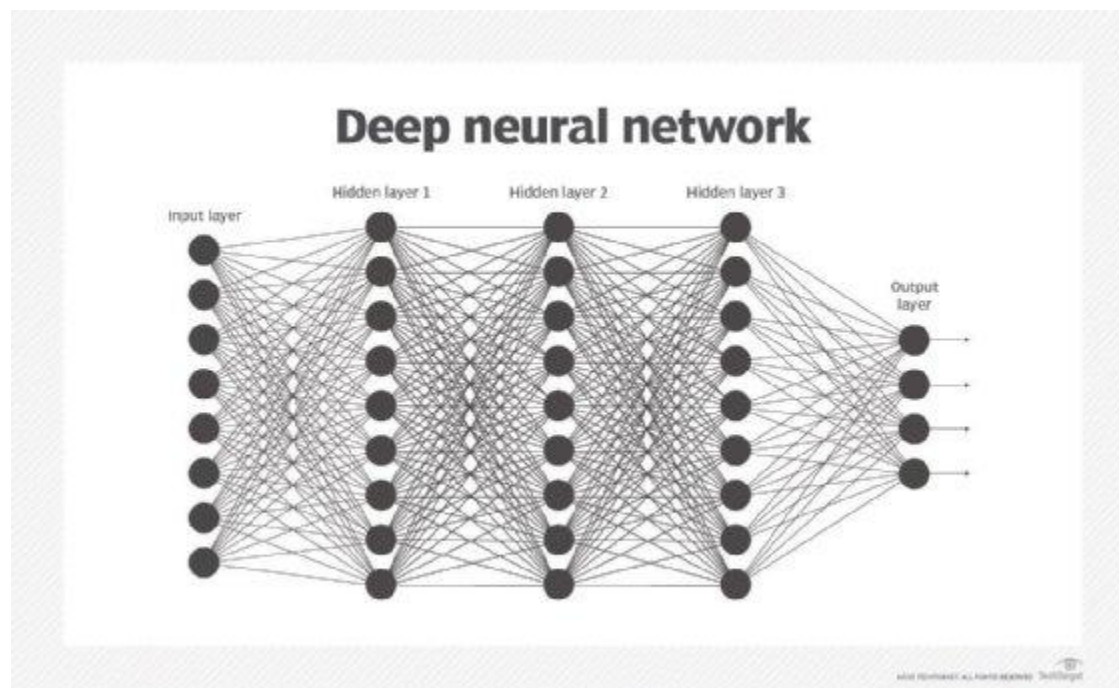
Model: "sequential_138"

Layer (type)	Output Shape	Param #
=====		
dense_414 (Dense)	(None, 128)	6016
dense_415 (Dense)	(None, 128)	16512
dropout_276 (Dropout)	(None, 128)	0

dropout_277 (Dropout)	(None, 128)	0
dense_416 (Dense)	(None, 1)	129

=====

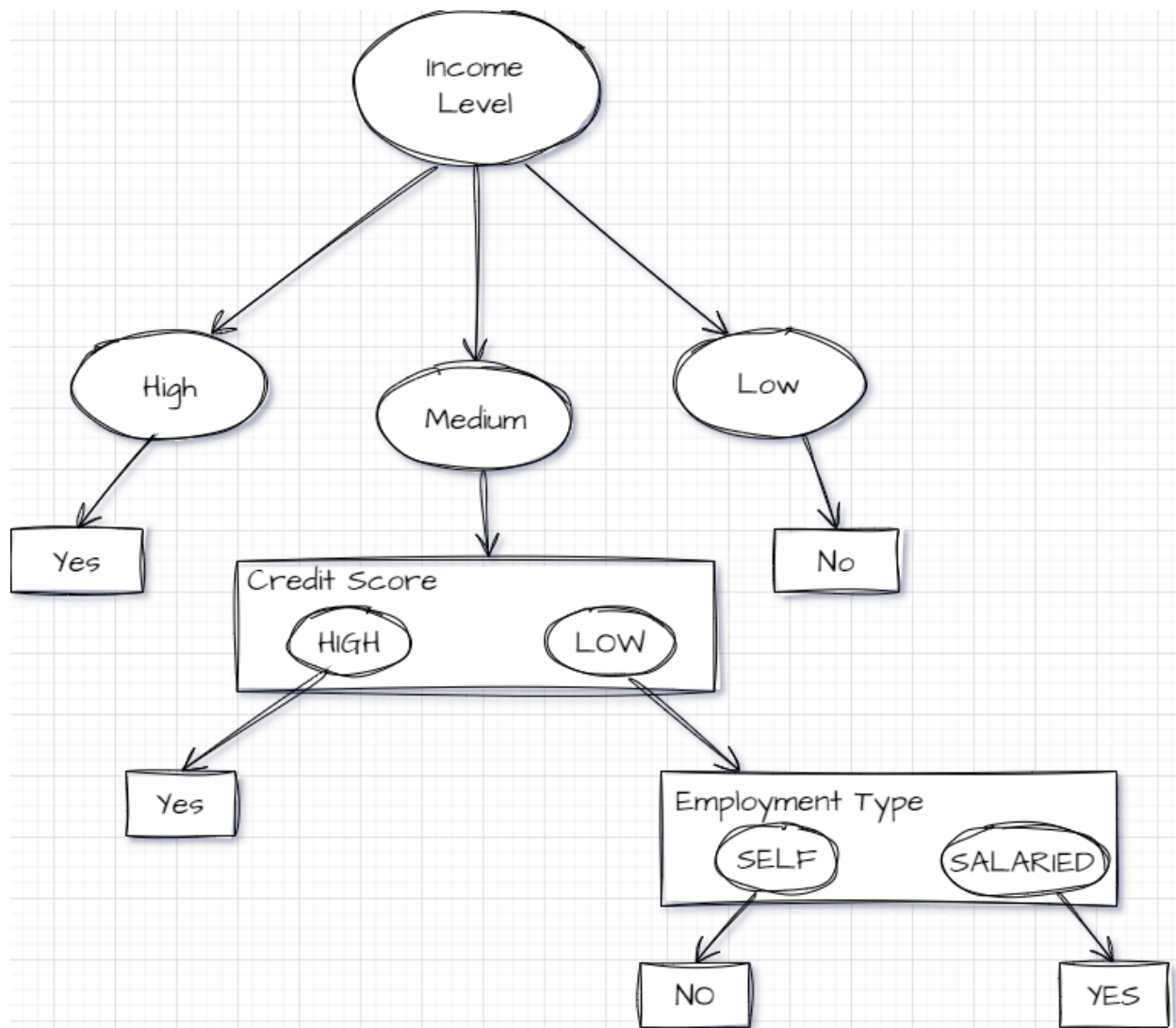
Total params: 22657 (88.50 KB)
Trainable params: 22657 (88.50 KB)
Non-trainable params: 0 (0.00 Byte)



Decision Tree:

Decision Tree works on the principle of entropy and and the information gain the higher the feature information gain more the chance for the feature to become root node.

Decision Tree used a recursive approach for making a decision tree because, on each child node, there is another Decision Tree.



The above the the example of a decision tree which tells us the prediction of loan Eligibility.

The other two are tree ensemble models which are the collection of many decision trees.

Random Forest:

The random Forest randomly creates nearly 50 to 100 decision trees and the newly created decision trees make predictions on the test set. The output is decided using a vote system the highest voted class is predicted.

XGBoost:

XGBoost is a powerful model similar to Random Forest but it does not create trees Randomly. XGBoost creates trees based on previously wrong decisions made by trees.