# Capstone Project - Binary Classification on ‘Customer\_Churn’using Keras

BY NOAHCHRISTY

Table of Contents

1. Problem Statement
2. Project Objective
3. Data Description
4. Data Pre-processing Steps and Inspiration
5. Choosing the Algorithm for the Project
6. Motivation and Reasons For Choosing the Algorithm
7. Assumptions
8. Model Evaluation and Techniques
9. Inferences from the Same
10. Future Possibilities of the Project
11. Conclusion
12. References

# Problem Statement

You are the Data Scientist at a telecom company “Leo” whose customers are churning out to its competitors. You have to analyse the data of your company and find insights and stop your customers from churning out to other telecom companies.

# Project Objective

**Feature Engineering and Data Preprocessing:**

* Conduct comprehensive feature engineering to extract meaningful insights from the customer dataset.
* Apply appropriate data preprocessing techniques, such as handling missing values, scaling numerical features, and encoding categorical variables.

**Develop a Customer Churn Prediction Mode**l:

* Design and implement a robust deep learning model using Keras to predict customer churn.
* The model should be capable of analyzing historical customer data and identifying patterns indicative of potential churn.

**Model Training and Validation:**

* Train the Customer Churn Prediction Model on historical data with a focus on achieving high accuracy and reliability.
* Utilize techniques such as cross-validation and hyperparameter tuning to optimize the model's performance.

# Data Description

**Overview:**

The Telco Customer Churn dataset is a collection of data related to telecommunications customers and their churn behavior. Churn, in this context, refers to customers ending their association with the telecommunications service. This dataset is valuable for understanding the factors influencing customer churn and developing predictive models to identify potential churners.

**Dataset Information:**

**Source: Kaggle**

**Number of Instances:** 7043

**Number of Features:** 21

# **Features:**

# The dataset contains various features that provide insights into customer behavior and interactions with the telecommunications service. Key features include:

# CustomerID: Unique identifier for each customer.

# Gender: The gender of the customer (e.g., Male, Female).

# SeniorCitizen: Whether the customer is a senior citizen (1) or not (0).

# Partner: Whether the customer has a partner (Yes/No).

# Dependents: Whether the customer has dependents (Yes/No).

# Tenure: The number of months the customer has been with the company.

# PhoneService: Whether the customer has phone service (Yes/No).

# MultipleLines: Whether the customer has multiple lines (Yes/No).

# InternetService: Type of internet service subscribed by the customer (DSL, Fiber optic, No).

# OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport: Various online services subscribed by the customer (Yes/No/No internet service).

# StreamingTV, StreamingMovies: Whether the customer subscribes to streaming services (Yes/No/No internet service).

# Contract: Type of contract the customer has (Month-to-month, One year, Two years).

# PaperlessBilling: Whether the customer uses paperless billing (Yes/No).

# PaymentMethod: The payment method used by the customer.

# MonthlyCharges: The amount charged to the customer monthly.

# TotalCharges: The total amount charged to the customer over their tenure

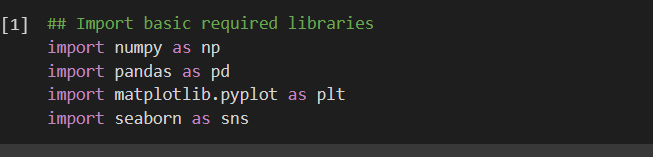
# **Target Variable:**

# Churn: Whether the customer has churned (Yes/No).

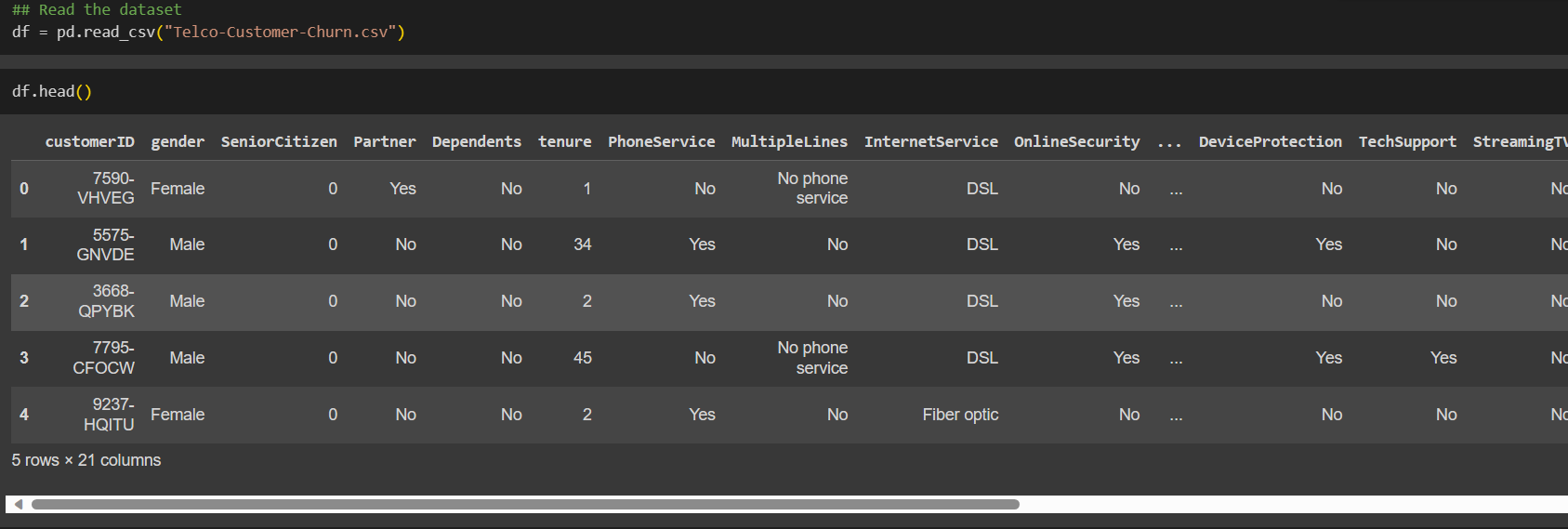
# Data Preprocessing Steps And Inspiration

The preprocessing of the data included the following steps:

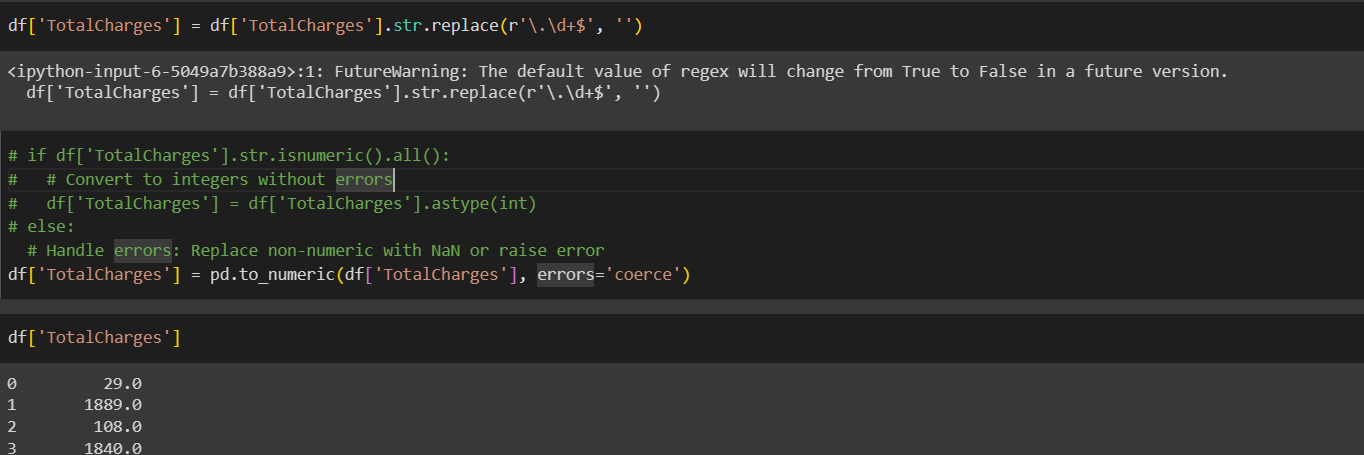
1. Import the basic required libraries:



1. Download the telco customer churn dataset from kaggle and readf it



1. The code df['TotalCharges'] = df['TotalCharges'].str.replace(r'\.\d+$', '') is a pandas operation used to modify the 'TotalCharges' column in a DataFrame (df).



# Choosing the Algorithm For the Project.

It is focused on building neural network models using the Keras library for a Telco customer churn prediction project. The choice of this algorithm, specifically neural networks, might be driven by several considerations:

1. Complex Relationships: Neural networks, especially deep neural networks, are known for their ability to capture complex relationships within data. If the Telco customer churn dataset has intricate patterns and dependencies that may not be well-modeled by simpler algorithms, a neural network can learn and represent these patterns effectively.
2. Feature Learning: Neural networks excel at automatically learning hierarchical features from raw data. In this project, the models are built using features like 'tenure,' 'Monthly Charges,' and 'Total Charges.' Neural networks can autonomously learn relevant features from these inputs, potentially uncovering non-linear relationships.
3. Feature Learning: Neural networks excel at automatically learning hierarchical features from raw data. In this project, the models are built using features like 'tenure,' 'Monthly Charges,' and 'Total Charges.' Neural networks can autonomously learn relevant features from these inputs, potentially uncovering non-linear relationships

It's important to note that the choice of algorithm depends on various factors such as the nature of the data, the problem at hand, the size of the dataset, computational resources, and the specific goals of the project. While neural networks offer powerful capabilities, they may also require careful tuning, consideration of overfitting, and sufficient data for training

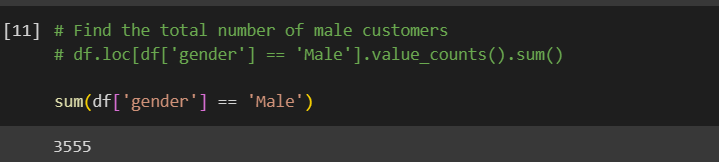
# Assumptions

The following assumptions were made in order to create the model for Customer Churn project.

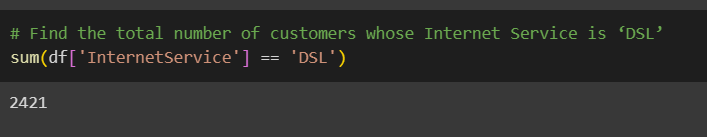
**Data Manipulation:**

1. Total Number of Male Customers: Count the number of male customers in the dataset

.

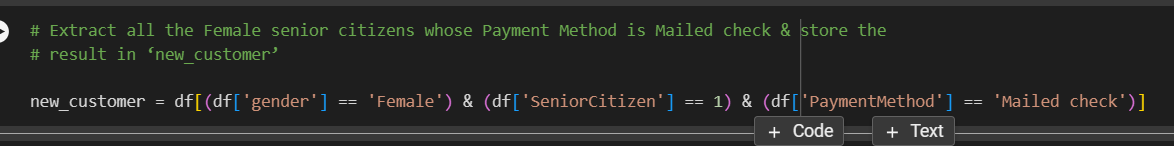


1. Total Number of Customers with DSL Internet Service: Count the number of customers whose Internet Service is 'DSL'.



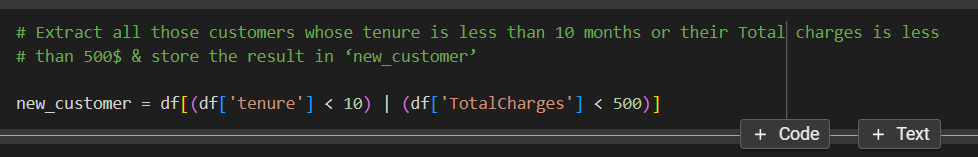
1. Extract Female Senior Citizens with Mailed Check Payment Method:

Create a new DataFrame (new\_customer) by extracting female senior citizens with the payment method 'Mailed check'.



1. Extract Customers with Tenure Less Than 10 Months or Total Charges Less Than $500:

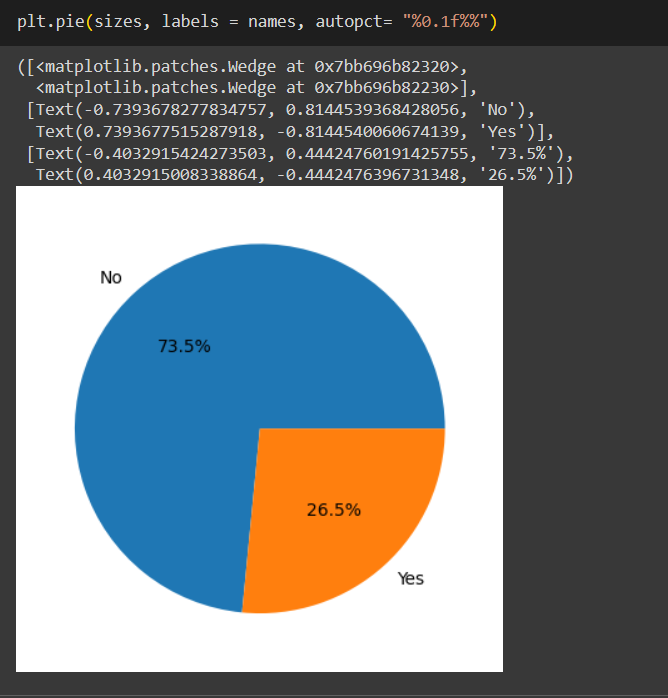
Update the new\_customer DataFrame by adding customers with tenure less than 10 months or total charges less than $500.



**Data Visualization:**

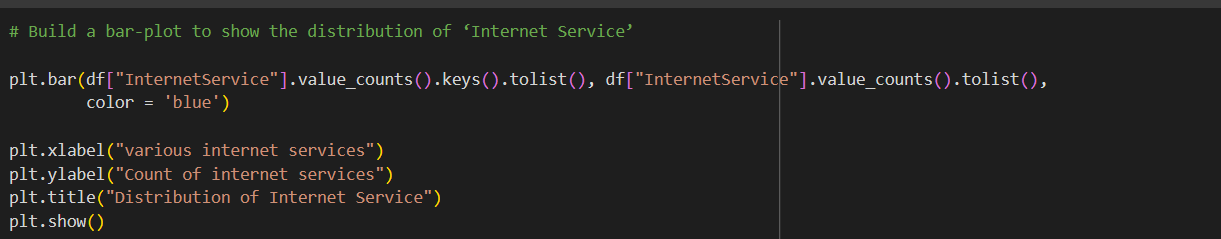
1. Pie Chart for Churn Distribution:

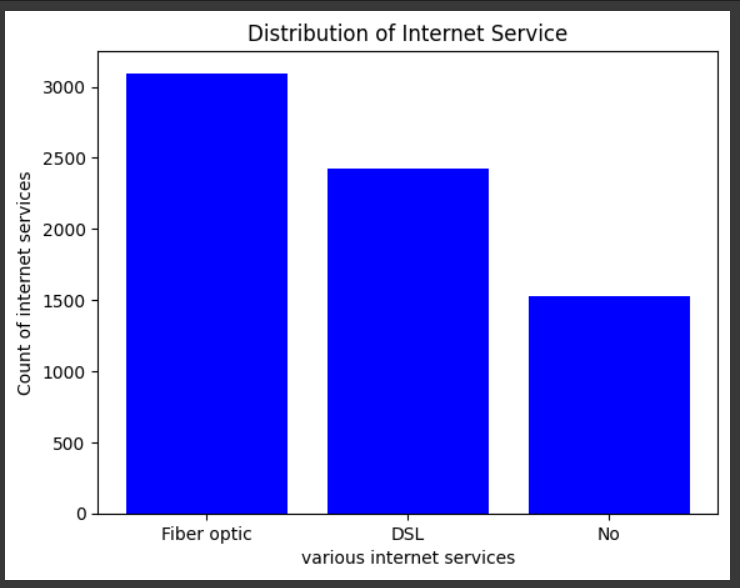
Visualize the distribution of churn (Yes/No) using a pie chart.



1. Bar Plot for Internet Service Distribution:

Visualize the distribution of Internet Service types using a bar plot.





# Model Evaluation and Technique

The following techniques and steps were involved in the evaluation of the model

Model 1:

* Use Keras to build a sequential model.
* Input layer with 12 nodes and ReLU activation.
* Hidden layer with 8 nodes and ReLU activation.
* Output layer with 1 node and Sigmoid activation.
* Compile the model with the Adam optimizer and binary cross-entropy loss.
* Train the model on the 'tenure' feature for predicting churn.
* Print the confusion matrix and plot the accuracy vs epochs graph.

Model 2:

* Build a similar model as Model 1 with additional dropout layers.
* Add a dropout layer after the input layer (dropout rate = 0.3) and after the hidden layer (dropout rate = 0.2).
* Compile, train, print confusion matrix, and plot accuracy vs epochs.

Model 3:

* Build Sequential Model with Multiple Features (3rd Model):
* Build a model with 'tenure', 'Monthly Charges', and 'Total Charges' as features.
* Input layer with 12 nodes and ReLU activation.
* Hidden layer with 8 nodes and ReLU activation.
* Output layer with 1 node and Sigmoid activation.
* Compile, train, print confusion matrix, and plot accuracy vs epochs.

# Inferences from the Project

From Model 1:

* Here, we are setting the test\_size to be 0.25, which means 25% of the records go into the test set, while 70% of the records go into the train set.
* Going ahead we will add the input layer to our model. This input layer would comprise of 12 nodes and would have ‘relu’ as the activation function. After that we’ll add a hidden layer with 8 nodes and ‘relu’ as activation function. Finally, we’ll add the output layer which would comprise of just one node and ‘sigmoid’ as activation function.
* We are using ‘sigmoid’ here because this is a binary classification problem and ‘sigmoid’ gives us a probability between 0 & 1.
* Optimizer used is ‘adam’ and we would want to calculate the accuracy
* Going ahead, we will fit the model on the train set and evaluate it on top of the test set. The number of epochs given over here is 150.
* So, the mean accuracy comes out to be 74.59%
* Further, we will, predict the values on ‘x\_test’ and build a confusion matrix with the actual values and the predicted values.

From Model 2:

* Now, we are building our 2nd model, where we are adding a drop-out layer after the input layer and the hidden layer.
* Drop-out value of 0.3 means that 70% of the nodes in the input layer will be dropped out.
* Drop-out value of 0.2 means that 80% of the nodes in the hidden layer will be dropped out.
* So, the mean accuracy comes out to be 73.68%.

From Model 3:

* This time, we are taking ‘Monthly Charges’, ‘Total Charges’ and ‘Tenure’ as the features and ‘Churn’ as the target
* After this, we divide the data into train and test sets and build the model on train test and predict the values on the test set
* this gives a mean validation accuracy of 73.73%

# Future Possibilities

When discussing the future possibilities of a project, you can explore various directions and potential enhancements

Integration of Additional Features:

Explore the inclusion of new relevant features that could provide more insights into customer behavior and churn prediction. This may involve collaborating with other departments to gather additional data.

Utilizing Advanced Machine Learning Techniques:

Experiment with more advanced machine learning techniques, such as ensemble methods, hyperparameter tuning, or automated machine learning (AutoML), to optimize model performance and robustness.

Customer Segmentation:

Implement customer segmentation strategies to tailor models for different customer segments. This could lead to more personalized churn prediction models, recognizing that different customer groups may exhibit distinct churn behaviors.

Customer Satisfaction Analysis:

Integrate customer satisfaction metrics and sentiment analysis to better understand the factors influencing customer decisions. This could involve mining feedback data from customer interactions and surveys.

# Conclusion

The first model gave us a mean validation accuracy of 74.59%, the second model had accuracy of 73.68 and the third model had a mean validation accuracy of 73.73%.

The second model gave us the least accuracy because we added two dropout layers with high probabilities of dropout.

Now, there could be many factors why third model’s accuracy was less than that of first model.

Most probably one or more of the features used during the model building could be of less significance leading to the reduction in accuracy. It should also be kept in mind that these accuracy values are very specific to the hyperparameters used during the model building process such as optimizers, activation functions and number of epochs. If we were to tweak these hyperparameters we would get completely different accuracy values for all the three models.

# References