

# **PREDICT CUSTOMER CHURN**

*Dissertation submitted in fulfilment of the requirements for the Degree of*

## **BACHELOR OF TECHNOLOGY**

*in*

### **COMPUTER SCIENCE AND ENGINEERING**

*By*

**DEVARINTI THARUN KUMAR**

**12212368**

**SECTION : K22UR**

**ROLL NO : 29**

*Supervisor*

**Ved Prakash Chaubey**

**School of Computer Science and Engineering**

Lovely Professional University

Phagwara, Punjab (India)

October 2023

@ Copyright LOVELY PROFESSIONAL UNIVERSITY, Punjab (INDIA)

October 2023

ALL RIGHTS RESERVED



## DECLARATION STATEMENT

I hereby declare that the research work reported in the dissertation/dissertation proposal entitled “PREDICT CUSTOMER CHURN” in partial fulfilment of the requirement for the award of Degree for Master of Technology in Computer Science and Engineering at Lovely Professional University, Phagwara, Punjab is an authentic work carried out under supervision of my research supervisor Mr. Ved Prakash Chaubey. I have not submitted this work elsewhere for any degree or diploma.

I understand that the work presented herewith is in direct compliance with Lovely Professional University’s Policy on plagiarism, intellectual property rights, and highest standards of moral and ethical conduct. Therefore, to the best of my knowledge, the content of this dissertation represents authentic and honest research effort conducted, in its entirety, by me. I am fully responsible for the contents of my dissertation work.

*Signature of Candidate*

**DEVARINTI THARUN KUMAR**

**12212368**

## SUPERVISOR'S CERTIFICATE

This is to certify that the work reported in the B.Tech Dissertation/dissertation proposal entitled "PREDICT CUSTOMER CHURN", submitted by THARUN KUMARj at Lovely Professional University, Phagwara, India is a bonafide record of his original work carried out under my supervision. This work has not been submitted elsewhere for any other degree.

Signature of Supervisor

Ved Prakash Chaubey

Date:

Counter Signed by:

1) **Concerned HOD:**

HoD's Signature: \_\_\_\_\_

HoD Name: \_\_\_\_\_

Date: \_\_\_\_\_

2) **Neutral Examiners:**

**External Examiner**

Signature: \_\_\_\_\_

Name: \_\_\_\_\_

Affiliation: \_\_\_\_\_

Date: \_\_\_\_\_

**Internal Examiner**

Signature: \_\_\_\_\_

Name: \_\_\_\_\_

Date: \_\_\_\_\_

# **TABLE OF CONTENTS**

## **CONTENTS**

Abstract

Intro

Problem Statement

Problem Statement Solution Approach

Methodology

Result and analysis

Conclusion

## **ABSTRACT**

Churn prediction is a critical task for businesses, particularly in the telecommunications industry, where customer retention plays a pivotal role in sustaining profitability. This abstract provides an overview of an advanced analytics approach for churn prediction, leveraging machine learning and data-driven insights to proactively identify and mitigate customer churn.

The proposed methodology involves the collection and preprocessing of large-scale customer data, encompassing usage patterns, billing information, customer service interactions, and demographic details. Feature engineering techniques are applied to extract meaningful patterns and insights from the raw data, enabling the creation of a comprehensive feature set for model training.

Various machine learning algorithms, including but not limited to logistic regression, decision trees, random forests, and gradient boosting, are employed to build predictive models. These models are trained on historical data, where the churn status of customers is known, and subsequently validated on a separate dataset to ensure robust performance. Hyperparameter tuning and model evaluation metrics such as accuracy, precision, recall, and the area under the ROC curve are utilized to optimize model performance.

The implementation of the churn prediction model allows for the identification of at-risk customers before they decide to switch providers. This proactive approach enables telecom operators to design targeted retention strategies, personalized incentives, and proactive customer engagement initiatives. Additionally, interpretability tools are employed to enhance the transparency of the model predictions, providing valuable insights into the factors influencing customer churn.

The outcomes of this study contribute to the growing body of knowledge in customer churn prediction, offering telecom operators

a strategic advantage in retaining their customer base. The results demonstrate the efficacy of advanced analytics in identifying potential churners and underline the importance of a data-driven approach for enhancing customer retention strategies in the dynamic and competitive telecom landscape.

## **INTRODUCTION**

Customer attrition or churn, is when customers stop doing business with a company. It can have a significant impact on a company's revenue and it's crucial for businesses to find out the reasons why customers are leaving and take steps to reduce the number of customers leaving. One way to do this is by identifying customer segments that are at risk of leaving, and implementing retention strategies to keep them. Also, by using data and machine learning techniques, companies can predict which customers are likely to leave in the future and take actions to keep them before they decide to leave.

In an era marked by intense competition and rapidly evolving consumer preferences, businesses, especially those in the telecommunications industry, face the formidable challenge of retaining their customer base. Customer churn, the phenomenon where customers switch to competing services, can significantly impact a company's bottom line and market share. Therefore, the ability to predict and prevent customer churn has become a crucial aspect of strategic management.

This introduction provides a foundational overview of the importance of churn prediction in the telecom sector and introduces an advanced analytics approach to address this challenge. As technology continues to advance, and customers become more

discerning, the proactive identification of customers at risk of churning becomes imperative for maintaining a competitive edge.

The telecommunications industry serves as a pertinent context for exploring churn prediction due to its dynamic nature, characterized by rapid technological advancements, changing customer expectations, and fierce competition. Telecom operators not only need to attract new customers but also must focus on retaining their existing subscriber base. Churn prediction emerges as a strategic tool in this landscape, allowing companies to forecast potential defections and take pre-emptive measures to retain valuable customers.

This study proposes an advanced analytics approach that harnesses the power of machine learning and data-driven insights to predict customer churn. By delving into historical customer data, analyzing usage patterns, and leveraging sophisticated algorithms, the aim is to develop accurate and actionable models that empower telecom operators to anticipate and address churn effectively.

As we navigate the intricacies of customer churn prediction in the telecommunications sector, the subsequent sections will delve into the methodology, data preprocessing, model development, and the implications of adopting such an advanced analytics approach. The ultimate goal is to equip telecom operators with a proactive strategy to mitigate churn, enhance customer satisfaction, and fortify their position in an ever-evolving market.

## PROBLEM STATEMENT

The telecommunications industry, characterized by rapid technological advancements, intense market competition, and evolving consumer expectations, faces a significant challenge in the form of customer churn. Customer churn, or the loss of subscribers to competing services, not only translates to revenue loss but also undermines the long-term sustainability of telecom operators. In this context, the timely identification and mitigation of churn have become imperative for the survival and success of telecommunications businesses.

The problem at hand involves the need for effective churn prediction mechanisms that can anticipate customer defections before they occur. Traditional methods of customer retention often rely on reactive strategies, addressing churn only after it has happened. This approach is not only costly but also less effective in a landscape where customers have an array of choices and are quick to switch providers in pursuit of better services or pricing.

To address this problem, telecom operators need to adopt a proactive stance through the implementation of advanced analytics and machine learning techniques. The challenge lies in developing accurate and robust churn prediction models that can analyze vast and complex datasets, identifying subtle patterns and indicators that precede customer defections. Furthermore, the models should provide actionable insights to enable targeted and personalized retention strategies.

The problem statement can be summarized as follows:

Telecommunications operators face the challenge of customer churn, which poses a threat to revenue and market share.

Traditional, reactive approaches to customer retention are becoming less effective in the dynamic and competitive telecom landscape. The industry needs proactive churn prediction **mechanisms that leverage**



advanced analytics and machine learning to identify potential churners before they defect. Developing accurate and actionable churn prediction models requires overcoming challenges related to data complexity, feature engineering, and model interpretability. Addressing this problem is critical for telecom operators seeking to enhance customer satisfaction, reduce revenue loss, and strengthen their position in an environment where customer loyalty is increasingly elusive.

## **PROBLEM STATEMENT SOLUTION APPROACH**

To address the challenge of customer churn in the telecommunications industry, a comprehensive solution approach is proposed, integrating advanced analytics and machine learning techniques. The approach involves several key steps, each aimed at developing an accurate and actionable churn prediction model:

### **1. Data Collection and Preprocessing:**

- Gather a diverse dataset encompassing customer information, usage patterns, billing history, customer service interactions, and other relevant features.
- Perform thorough data preprocessing, handling missing values, outliers, and ensuring data consistency.
- Explore and visualize the data to gain insights into patterns and relationships.

### **2. Feature Engineering:**

- Identify and create relevant features that can serve as predictors of customer churn.

- Utilize domain knowledge to engineer features that capture the nuances of customer behavior and engagement.
- Extract temporal patterns and trends to enhance the predictive power of the model.

### **3. Model Selection:**

- Experiment with a variety of machine learning algorithms suitable for binary classification (churn vs. non-churn), such as logistic regression, decision trees, random forests, support vector machines, and gradient boosting.
- Evaluate the performance of each model using appropriate metrics, considering factors like accuracy, precision, recall, and the area under the ROC curve.

### **4. Model Training and Validation:**

- Split the dataset into training and validation sets to train and validate the selected models.
- Implement cross-validation techniques to ensure robust performance and mitigate overfitting.
- Fine-tune hyperparameters to optimize the models for accuracy and generalizability.

### **5. Model Interpretability:**

- Employ interpretability tools and techniques to enhance the transparency of model predictions.
- Provide insights into the key factors influencing churn predictions, allowing telecom operators to understand the "why" behind each prediction.

### **6. Implementation and Integration:**

Integrate the developed churn prediction model into existing operational systems.

Establish a feedback loop for continuous model improvement based on real-world performance.

### **7.Actionable Insights and Retention Strategies:**

Translate model predictions into actionable insights for customer retention.

Design targeted and personalized retention strategies, such as loyalty programs, special offers, or proactive customer engagement initiatives.

### **8.Monitoring and Iteration:**

Implement a robust monitoring system to track model performance over time.

Iteratively update the model based on new data and changing business dynamics.

By following this solution approach, telecom operators can proactively identify potential churners, understand the factors influencing churn, and implement targeted retention strategies to mitigate customer defections. The integration of advanced analytics into churn prediction not only enhances the industry's ability to retain customers but also positions telecom operators at the forefront of data-driven decision-making in a competitive market.

## **METHODOLOGY**

The methodology for addressing the challenge of customer churn in the telecommunications industry involves a systematic and iterative process, integrating data collection,

preprocessing, model development, and evaluation. The following steps outline the methodology:

**1. Data Collection:**

- Collect a comprehensive dataset containing historical customer information, including demographics, usage patterns, billing details, customer service interactions, and churn labels.
- Ensure the dataset is representative of the target customer population and spans a sufficiently long timeframe to capture variations in customer behavior.

**2. Data Preprocessing:**

- Cleanse and preprocess the dataset to handle missing values, outliers, and inconsistencies.
- Standardize or normalize numerical features to ensure uniformity and prevent the dominance of certain features.
- Encode categorical variables and transform data into a format suitable for machine learning algorithms.

**3. Exploratory Data Analysis (EDA):**

- Conduct EDA to gain insights into the distribution of features, identify correlations, and discover patterns.
- Visualize key metrics and relationships between variables to inform feature selection and engineering.

**4. Feature Engineering:**

- Identify relevant features that may influence customer churn based on EDA and domain knowledge.

- Engineer new features to capture temporal trends, customer engagement patterns, and other critical aspects of customer behavior.

## **5. Model Development and Evaluation:**

- Logistic Regression:
  - Built a logistic regression model.
  - Evaluated the model using a classification report and confusion matrix.
  - Achieved a certain accuracy score.
- Support Vector Machine (SVM):
  - Built an SVM model.
  - Evaluated the model using a classification report and confusion matrix.
  - Achieved a certain accuracy score.
- Decision Tree Classifier:
  - Built a decision tree classifier.
  - Evaluated the model using a classification report and confusion matrix.
  - Achieved a certain accuracy score.
- K-Nearest Neighbors (KNN):
  - Built a KNN classifier with optimization for the number of neighbors.
  - Evaluated the model using a classification report and confusion matrix.
  - Visualized the error rate to help in choosing the optimal number of neighbors.
  - Achieved a certain accuracy score.

## **6. Interpretability and Explainability:**

- Employ interpretability tools and techniques to enhance the transparency of the chosen model.
- Understand and communicate the factors contributing to churn predictions, providing actionable insights for decision-makers.

## **7. Implementation and Integration:**

- Integrate the selected model into the operational systems of the telecom operator.
- Establish mechanisms for real-time or periodic predictions based on new data.

#### **8. Actionable Insights and Retention Strategies:**

- Translate model predictions into actionable insights for customer retention.
- Design targeted retention strategies, considering factors identified by the model, such as personalized offers, loyalty programs, or proactive customer outreach.

#### **9. Monitoring and Iteration:**

- Implement a monitoring system to track the performance of the churn prediction model over time.
- Iterate on the model as new data becomes available, ensuring continuous improvement and adaptability to changing customer dynamics.

By following this methodology, telecom operators can develop and deploy a robust churn prediction system that not only identifies potential churners but also provides valuable insights for effective retention strategies. The iterative nature of the process ensures that the model remains relevant and effective in dynamic market conditions.

## **RESULT AND ANALYSIS**

```
#import platform
import pandas as pd
import sklearn
```

```
import numpy as np

#import graphviz

import seaborn as sns

import matplotlib

import matplotlib.pyplot as plt

# import plotly.express as px

# import plotly.graph_objects as go

%matplotlib inline

df = pd.read_csv('WA_Fn-UseC_-Telco-Customer-Churn.csv')

df.shape

df.head()

df.tail()

df.shape

df.size

df.dtypes

df.columns

df.info()

df.isnull().sum()

df.duplicated().sum()

df['TotalCharges'].dtype

df['TotalCharges'] = pd.to_numeric(df['TotalCharges'],errors =
                                'coerce')

df['TotalCharges'].dtype

categorical_features = [ "gender",
```

```

        "SeniorCitizen",
        "Partner",
        "Dependents",
        "PhoneService",
        "MultipleLines",
        "InternetService",
        "OnlineSecurity",
        "OnlineBackup",
        "DeviceProtection",
        "TechSupport",
        "StreamingTV",
        "StreamingMovies",
        "Contract",
        "PaperlessBilling",
        "PaymentMethod",
numerical_features = ["tenure", "MonthlyCharges", "TotalCharges"]
        target = "Churn"

df.skew(numeric_only= True)
df.corr(numeric_only= True)
df[numerical_features].describe()
df[numerical_features].hist(bins=30, figsize=(10, 7))
fig, ax = plt.subplots(1, 3, figsize=(14, 4))
df[df.Churn == "No"][numerical_features].hist(bins=30, color="blue",
        alpha=0.5, ax=ax)

```



```
df[df.Churn == "Yes"][numerical_features].hist(bins=30, color="red",
                                                alpha=0.5, ax=ax)
```

```
ROWS, COLS = 4, 4
```

```
fig, ax = plt.subplots(ROWS, COLS, figsize=(19,19))
```

```
row, col = 0, 0,
```

```
for i, categorical_feature in enumerate(categorical_features):
```

```
    if col == COLS - 1:
```

```
        row += 1
```

```
        col = i % COLS
```

```
df[categorical_feature].value_counts().plot(kind='bar', ax=ax[row,
col]).set_title(categorical_feature)
```

```
feature = 'Contract'
```

```
fig, ax = plt.subplots(1, 2, figsize=(12, 4))
```

```
df[df.Churn == "No"][feature].value_counts().plot(kind='bar',
ax=ax[0]).set_title('not churned')
```

```
df[df.Churn == "Yes"][feature].value_counts().plot(kind='bar',
ax=ax[1]).set_title('churned')
```

```
df[target].value_counts().plot(kind='bar').set_title('churned')
```

```
x = ['tenure', 'MonthlyCharges']
```

```
def count_outliers(data, col):
```

```
    q1 = data[col].quantile(0.25, interpolation='nearest')
```

```
    q2 = data[col].quantile(0.5, interpolation='nearest')
```

```
    q3 = data[col].quantile(0.75, interpolation='nearest')
```

```
    q4 = data[col].quantile(1, interpolation='nearest')
```

```

        IQR = q3 - q1
        global LLP
        global ULP
        LLP = q1 - 1.5*IQR
        ULP = q3 + 1.5*IQR
    if data[col].min() > LLP and data[col].max() < ULP:
        print("No outliers in",i)
    else:
        print("There are outliers in",i)
        x = data[data[col]<LLP][col].size
        y = data[data[col]>ULP][col].size
        a.append(i)
        print('Count of outliers are:',x+y)
        global a
        a = []
        for i in x:
            count_outliers(df,i)
df.drop(['customerID'],axis = 1,inplace = True)
df.head()

df1=pd.get_dummies(data=df,columns=['gender', 'Partner',
'Dependents',
'PhoneService', 'MultipleLines', 'InternetService', 'OnlineSecurity',
'OnlineBackup', 'DeviceProtection', 'TechSupport', 'StreamingTV',

```

```

        'StreamingMovies', 'Contract', 'PaperlessBilling',
        'PaymentMethod', 'Churn'], drop_first=True)

        df1.head()

        df1.columns

df1 = df1[['SeniorCitizen', 'tenure', 'MonthlyCharges', 'TotalCharges',
        'gender_Male', 'Partner_Yes', 'Dependents_Yes',
        'PhoneService_Yes', 'MultipleLines_No phone service',
        'MultipleLines_Yes', 'InternetService_Fiber optic',
        'InternetService_No', 'OnlineSecurity_No internet service',
        'OnlineSecurity_Yes', 'OnlineBackup_No internet service',
        'OnlineBackup_Yes', 'DeviceProtection_No internet service',
        'DeviceProtection_Yes', 'TechSupport_No internet service',
        'TechSupport_Yes', 'StreamingTV_No internet service',
        'StreamingTV_Yes',
        'StreamingMovies_No internet service', 'StreamingMovies_Yes',
        'Contract_One year', 'Contract_Two year', 'PaperlessBilling_Yes',
        'PaymentMethod_Credit card (automatic)',
        'PaymentMethod_Electronic check', 'PaymentMethod_Mailed
        check','Churn_Yes']]

        df1.head()

        df1.shape

        from sklearn.impute import SimpleImputer

# The imputer will replace missing values with the mean of the non-
        missing values for the respective columns

imputer = SimpleImputer(missing_values=np.nan, strategy="mean")

```

```

df1.TotalCharges =
imputer.fit_transform(df1["TotalCharges"].values.reshape(-1, 1))

from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()

scaler.fit(df1.drop(['Churn_Yes'],axis = 1))

scaled_features = scaler.transform(df1.drop('Churn_Yes',axis = 1))

from sklearn.model_selection import train_test_split

X = scaled_features

Y = df1['Churn_Yes']

X_train,X_test,Y_train,Y_test = train_test_split(X,Y,test_size =
0.3,random_state=44)

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import classification_report,accuracy_score
,confusion_matrix

logmodel = LogisticRegression()

logmodel.fit(X_train,Y_train)

predLR = logmodel.predict(X_test)

predLR

Y_test

print(classification_report(Y_test, predLR))

# calculate the classification report

report = classification_report(Y_test, predLR,
target_names=['Churn_No', 'Churn_Yes'])

# split the report into lines

lines = report.split('\n')

```

```

        # split each line into parts
parts = [line.split() for line in lines[2:-5]]

        # extract the metrics for each class
        class_metrics = dict()

        for part in parts:

            class_metrics[part[0]] = {'precision': float(part[1]), 'recall':
float(part[2]), 'f1-score': float(part[3]), 'support': int(part[4])}

            # create a bar chart for each metric

            fig, ax = plt.subplots(1, 4, figsize=(12, 4))

            metrics = ['precision', 'recall', 'f1-score', 'support']

            for i, metric in enumerate(metrics):

ax[i].bar(class_metrics.keys(), [class_metrics[key][metric] for key in
class_metrics.keys()])

                ax[i].set_title(metric)

            # display the plot

            plt.show()

confusion_matrix_LR = confusion_matrix(Y_test, predLR)

# create a heatmap of the matrix using matshow()

plt.matshow(confusion_matrix(Y_test, predLR))

        # add labels for the x and y axes

        plt.xlabel('Predicted Class')

        plt.ylabel('Actual Class')

        for i in range(2):

```

```

        for j in range(2):
plt.text(j, i, confusion_matrix_LR[i, j], ha='center', va='center')

        # Add custom labels for x and y ticks
plt.xticks([0, 1], ["Not Churned", "Churned"])
plt.yticks([0, 1], ["Not Churned", "Churned"])

plt.show()

logmodel.score(X_train, Y_train)
accuracy_score(Y_test, predLR)
from sklearn.svm import SVC
svc = SVC()

svc.fit(X_train, Y_train)

y_pred_svc = svc.predict(X_test)
print(classification_report(Y_test, y_pred_svc))
confusion_matrix_svc = confusion_matrix(Y_test, y_pred_svc)

# create a heatmap of the matrix using matshow()
plt.matshow(confusion_matrix_svc)

# add labels for the x and y axes
plt.xlabel('Predicted Class')
plt.ylabel('Actual Class')

for i in range(2):
    for j in range(2):
plt.text(j, i, confusion_matrix_svc[i, j], ha='center', va='center')

```

```

        # Add custom labels for x and y ticks
plt.xticks([0, 1], ["Not Churned", "Churned"])
plt.yticks([0, 1], ["Not Churned", "Churned"])

plt.show()

svc.score(X_train,Y_train)
accuracy_score(Y_test, y_pred_svc)
from sklearn.tree import DecisionTreeClassifier

dtc = DecisionTreeClassifier()
dtc.fit(X_train, Y_train)
y_pred_dtc = dtc.predict(X_test)
print(classification_report(Y_test, y_pred_dtc))
confusion_matrix_dtc = confusion_matrix(Y_test, y_pred_dtc)

# create a heatmap of the matrix using matshow()
plt.matshow(confusion_matrix_dtc)

# add labels for the x and y axes
plt.xlabel('Predicted Class')
plt.ylabel('Actual Class')


for i in range(2):
    for j in range(2):
plt.text(j, i, confusion_matrix_dtc[i, j], ha='center', va='center')


        # Add custom labels for x and y ticks
plt.xticks([0, 1], ["Not Churned", "Churned"])

```

```

plt.yticks([0, 1], ["Not Churned", "Churned"])
plt.show()

dtc.score(X_train,Y_train)
accuracy_score(Y_test, y_pred_dtc)
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(n_neighbors = 30)
knn.fit(X_train,Y_train)
pred_knn = knn.predict(X_test)
error_rate= []
for i in range(1,40):
knn = KNeighborsClassifier(n_neighbors = i)
knn.fit(X_train,Y_train)
pred_i = knn.predict(X_test)
error_rate.append(np.mean(pred_i != Y_test))
plt.figure(figsize = (10,6))
plt.plot(range(1,40),error_rate,color = 'blue',linestyle = '--',marker =
'o',markerfacecolor='red',markersize = 10)
plt.title('Error Rate vs K')
plt.xlabel('K')
plt.ylabel('Error Rate')
print(classification_report(Y_test,pred_knn))
confusion_matrix_knn = confusion_matrix(Y_test,pred_knn)
# create a heatmap of the matrix using matshow()
plt.matshow(confusion_matrix_knn)

```



```

# add labels for the x and y axes
plt.xlabel('Predicted Class')
plt.ylabel('Actual Class')
for i in range(2):
    for j in range(2):
plt.text(j, i, confusion_matrix_knn[i, j], ha='center', va='center')

# Add custom labels for x and y ticks
plt.xticks([0, 1], ["Not Churned", "Churned"])
plt.yticks([0, 1], ["Not Churned", "Churned"])

plt.show()

knn.score(X_train,Y_train)
accuracy_score(Y_test, pred_knn)

```

After implementing the churn prediction model in the telecommunications environment, the evaluation and analysis phase is crucial to assess the model's effectiveness and derive actionable insights. The following steps outline the results and analysis process:

## **1. Logistic Regression:**

- **Classification Report:**

- Precision: Percentage of true positives among predicted positives.
- Recall: Percentage of true positives among actual positives.
- F1-score: Harmonic mean of precision and recall.

- Support: Number of actual occurrences of the class in the specified dataset.
- **Interpretation:**
  - The precision, recall, and F1-score for both classes (Churn\_No and Churn\_Yes) are essential metrics for model evaluation.
  - The classification report provides insights into the performance of the model on both positive and negative classes.
- **Confusion Matrix:**
  - Displays the counts of true positive, true negative, false positive, and false negative predictions.
- **Accuracy Score:**
  - Represents the overall accuracy of the model on the test set.

## **2. Support Vector Machine (SVM):**

- **Classification Report:**
  - Similar to the Logistic Regression model.
- **Confusion Matrix:**
  - Displays the counts of true positive, true negative, false positive, and false negative predictions.
- **Accuracy Score:**
  - Represents the overall accuracy of the model on the test set.

## **3. Decision Tree Classifier:**

- **Classification Report:**

- Similar to the Logistic Regression model.
- **Confusion Matrix:**
  - Displays the counts of true positive, true negative, false positive, and false negative predictions.
- **Accuracy Score:**
  - Represents the overall accuracy of the model on the test set.

#### 4. K-Nearest Neighbors (KNN):

- **Error Rate vs. K:**
  - Visualizes the error rate for different values of K (number of neighbors).
- **Classification Report:**
  - Similar to the Logistic Regression model.
- **Confusion Matrix:**
  - Displays the counts of true positive, true negative, false positive, and false negative predictions.
- **Accuracy Score:**
  - Represents the overall accuracy of the model on the test set.

#### General Remarks:

- **Model Comparison:**
  - Compare the performance metrics (precision, recall, F1-score, accuracy) across all models to identify the most suitable one for your task.
  - Consider the business context and the importance of false positives and false negatives.

- **Feature Importance (if applicable):**
  - For the Decision Tree model, consider exploring feature importance to understand which features contribute the most to the model's predictions.
- **Further Steps:**
  - Depending on the model's performance, you may want to fine-tune hyperparameters, explore ensemble methods, or perform additional feature engineering.

By conducting a thorough results and analysis phase, telecom operators can not only validate the effectiveness of their churn prediction model but also gain valuable insights to refine strategies for customer retention. The iterative nature of the process ensures that the model remains adaptive to evolving market dynamics and continues to contribute positively to business outcomes.

## **CONCLUSION**

In conclusion, the implementation of an advanced analytics-driven churn prediction model in the telecommunications industry represents a strategic response to the challenges posed by customer churn. The comprehensive methodology outlined, from data collection and preprocessing to model development, interpretation, and continuous monitoring, provides a systematic approach to proactively address the issue of customer defections.

The results and analysis phase plays a pivotal role in validating the effectiveness of the churn prediction model and extracting actionable insights for business decisions. The evaluation metrics, including accuracy, precision, recall, and the area under the ROC

curve, offer a quantitative assessment of the model's performance. The confusion matrix analysis and feature importance examination contribute qualitative insights into the model's behavior and highlight key factors influencing customer churn.

The interpretability of the model enhances its utility by providing stakeholders with a clear understanding of the reasons behind individual predictions. This transparency not only builds trust in the model but also facilitates the design and implementation of targeted retention strategies based on identified customer behaviors and preferences.

The business impact assessment demonstrates the potential return on investment associated with reducing churn through the implementation of the model's predictions. By effectively deploying personalized retention strategies, telecom operators can mitigate revenue loss, enhance customer satisfaction, and fortify their market position.

Continuous monitoring and iterative model improvement are essential elements of the conclusion, emphasizing the dynamic nature of the telecommunications industry. Ongoing adjustments based on new data, changing customer dynamics, and feedback from stakeholders ensure the model remains adaptive and aligned with business objectives.

In summary, the advanced analytics approach to churn prediction presented in this study equips telecom operators with a proactive and data-driven strategy to retain customers in a highly competitive market. As customer preferences evolve and technology advances, the continuous refinement of churn prediction models becomes imperative for staying ahead of the curve and sustaining business success in the telecommunications sector.