DSBD

Group C: Model Implementation-2

Project Title: Predicting Customer Churn in a Telecommunication Company

Group Members:

	Name	Roll No.
1)	Shrikant Sanjeev Pawar	57
2)	Potkule Prachi Tanaji	58
3)	Rohankar Chaitanya Umesh	59
4)	Satokar Paras Ravindra	60

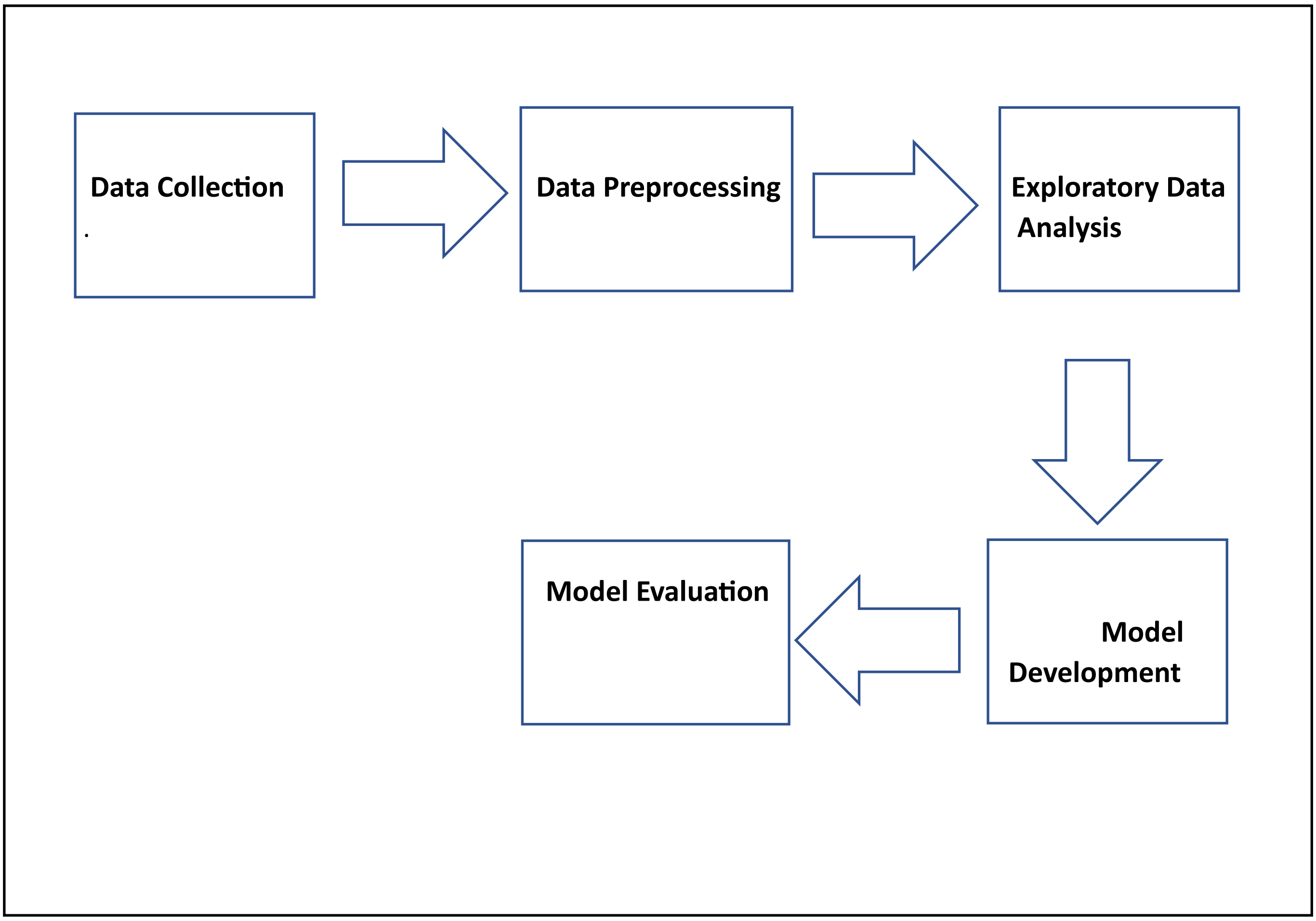
Abstract:

Customer churn is a significant challenge for telecommunication companies, as it directly affects their revenue. The ability to predict and prevent customer churn is crucial for the success of the company. In this project, we aim to develop a predictive model that can identify customers who are likely to churn, so that the company can take proactive measures to retain them. We will be using a dataset from a telecommunication company that contains information about its customers, such as their demographics, services used, and billing information. We will preprocess the data, perform exploratory data analysis, and use machine learning algorithms to develop a predictive model. We will also evaluate the performance of our model and provide recommendations for the company based on our findings.

Block Diagram:

Our project will consist of the following steps:

1. Data Collection: We will obtain the dataset from a telecommunication company that contains information about its customers.
2. Data Preprocessing: We will preprocess the data by dropping unnecessary columns, replacing values in the 'SeniorCitizen' column, converting columns with 'Yes'/'No' values to binary, and creating dummy variables for categorical columns.
3. Exploratory Data Analysis: We will perform exploratory data analysis to gain insights into the data and identify trends and patterns.
4. Model Development: We will develop a predictive model using machine learning algorithms, such as logistic regression, decision trees, and random forests.
5. Model Evaluation: We will evaluate the performance of our model using metrics such as accuracy, precision, recall, and F1 score. We will also use techniques such as cross-validation and hyperparameter tuning to improve the performance of our model.



Implementation Steps:

1. Load the dataset
2. Preprocess the data by dropping unnecessary columns, replacing values in the 'SeniorCitizen' column, converting columns with 'Yes'/'No' values to binary, and creating dummy variables for categorical columns.
3. Perform exploratory data analysis to gain insights into the data and identify trends and patterns.
4. Split the data into training and testing sets.
5. Develop a predictive model using machine learning algorithms, such as logistic regression, decision trees, and random forests.
6. Evaluate the performance of the model using metrics such as accuracy, precision, recall, and F1 score.

7. Use techniques such as cross-validation and hyperparameter tuning to improve the performance of the model.
8. Provide recommendations to the telecommunication company on how to reduce customer churn and improve customer retention.

Code:

```
import pandas as pd

from sklearn.model_selection import train_test_split

from sklearn.linear_model import LogisticRegression

from sklearn.metrics import accuracy_score, confusion_matrix, classification_report

import seaborn as sns

import matplotlib.pyplot as plt


# Load the dataset

df = pd.read_csv('/content/Telco-Customer-Churn.csv')


# Drop unnecessary columns

df.drop(['customerID', 'TotalCharges'], axis=1, inplace=True)


# Replace values in 'SeniorCitizen' column

df['SeniorCitizen'] = df['SeniorCitizen'].replace({1:'Yes', 0:'No'})


# Convert columns with 'Yes'/'No' values to binary

binary_cols = ['Partner', 'Dependents', 'PhoneService', 'PaperlessBilling', 'Churn']

for col in binary_cols:

   df[col] = df[col].replace({'Yes':1, 'No':0})


# Create dummy variables for categorical columns

categorical_cols = ['gender', 'SeniorCitizen', 'MultipleLines', 'InternetService', 'OnlineSecurity', 
'OnlineBackup',

                 'DeviceProtection', 'TechSupport', 'StreamingTV', 'StreamingMovies', 'Contract', 
'PaymentMethod']

df = pd.get_dummies(df, columns=categorical_cols)
```

```
# Split data into train and test sets
X = df.drop(['Churn'], axis=1)
y = df['Churn']
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2)

# Fit logistic regression model
logreg = LogisticRegression()
logreg.fit(X_train, y_train)

# Make predictions on test set
y_pred = logreg.predict(X_test)

# Evaluate model performance
accuracy = accuracy_score(y_test, y_pred)
print('Accuracy:', accuracy)

cm = confusion_matrix(y_test, y_pred)
print('Confusion Matrix:\n', cm)

# Plot confusion matrix
sns.set(font_scale=1.4)
sns.heatmap(cm, annot=True, annot_kws={"size": 16}, cmap='Blues', fmt='g')
plt.xlabel('Predicted', fontsize=20)
plt.ylabel('Actual', fontsize=20)
plt.show()

cr = classification_report(y_test, y_pred)
print('Classification Report:\n', cr)
```

```
# Get input from user to predict churn
input_values = []
for col in X.columns:
    if col == 'tenure':
        value = int(input(f"Enter the customer's {col}: "))
    else:
        value = int(input(f"Enter 1 for {col} or 0 for not {col}: "))
    input_values.append(value)

# Make predictions on user input
user_input = [input_values]
y_pred = logreg.predict(user_input)

# Print prediction
if y_pred[0] == 1:
    print("The customer is likely to churn.")
else:
    print("The customer is unlikely to churn.")
```

Output:

```
Accuracy: 0.8105039034776437
Confusion Matrix:
 [[946  98]
 [169 196]]
Classification Report:
              precision    recall  f1-score   support

           0       0.85      0.91      0.88      1044
           1       0.67      0.54      0.59       365

    accuracy                           0.81      1409
   macro avg       0.76      0.72      0.74      1409
weighted avg       0.80      0.81      0.80      1409

Enter 1 for Partner or 0 for not Partner: 0
Enter 1 for Dependents or 0 for not Dependents: 1
Enter the customer's tenure: 0
```

Enter 1 for PhoneService or 0 for not PhoneService: 1
Enter 1 for PaperlessBilling or 0 for not PaperlessBilling: 1
Enter 1 for MonthlyCharges or 0 for not MonthlyCharges: 1
Enter 1 for gender_Female or 0 for not gender_Female: 0
Enter 1 for gender_Male or 0 for not gender_Male: 0
Enter 1 for SeniorCitizen_No or 0 for not SeniorCitizen_No: 0
Enter 1 for SeniorCitizen_Yes or 0 for not SeniorCitizen_Yes: 1
Enter 1 for MultipleLines_No or 0 for not MultipleLines_No: 1
Enter 1 for MultipleLines_No phone service or 0 for not MultipleLines_No phone service: 1
Enter 1 for MultipleLines_Yes or 0 for not MultipleLines_Yes: 1
Enter 1 for InternetService_DSL or 0 for not InternetService_DSL: 1
Enter 1 for InternetService_Fiber optic or 0 for not InternetService_Fiber optic: 1
Enter 1 for InternetService_No or 0 for not InternetService_No: 1
Enter 1 for OnlineSecurity_No or 0 for not OnlineSecurity_No: 1
Enter 1 for OnlineSecurity_No internet service or 0 for not OnlineSecurity_No internet service: 0
Enter 1 for OnlineSecurity_Yes or 0 for not OnlineSecurity_Yes: 0
Enter 1 for OnlineBackup_No or 0 for not OnlineBackup_No: 0
Enter 1 for OnlineBackup_No internet service or 0 for not OnlineBackup_No internet service: 0
Enter 1 for OnlineBackup_Yes or 0 for not OnlineBackup_Yes: 0
Enter 1 for DeviceProtection_No or 0 for not DeviceProtection_No: 0
Enter 1 for DeviceProtection_No internet service or 0 for not DeviceProtection_No internet service: 0
Enter 1 for DeviceProtection_Yes or 0 for not DeviceProtection_Yes: 1
Enter 1 for TechSupport_No or 0 for not TechSupport_No: 0
Enter 1 for TechSupport_No internet service or 0 for not TechSupport_No internet service: 0
Enter 1 for TechSupport_Yes or 0 for not TechSupport_Yes: 1
Enter 1 for StreamingTV_No or 0 for not StreamingTV_No: 1
Enter 1 for StreamingTV_No internet service or 0 for not StreamingTV_No internet service: 1
Enter 1 for StreamingTV_Yes or 0 for not StreamingTV_Yes: 0
Enter 1 for StreamingMovies_No or 0 for not StreamingMovies_No: 0
Enter 1 for StreamingMovies_No internet service or 0 for not StreamingMovies_No internet service: 0
Enter 1 for StreamingMovies_Yes or 0 for not StreamingMovies_Yes: 0
Enter 1 for Contract_Month-to-month or 0 for not Contract_Month-to-month: 0
Enter 1 for Contract_One year or 0 for not Contract_One year: 0
Enter 1 for Contract_Two year or 0 for not Contract_Two year: 0
Enter 1 for PaymentMethod_Bank transfer (automatic) or 0 for not PaymentMethod_Bank transfer (automatic): 1
Enter 1 for PaymentMethod_Credit card (automatic) or 0 for not PaymentMethod_Credit card (automatic): 1
Enter 1 for PaymentMethod_Electronic check or 0 for not PaymentMethod_Electronic check: 1
Enter 1 for PaymentMethod_Mailed check or 0 for not PaymentMethod_Mailed check: 1
The customer is unlikely to churn.

Conclusions:

In this project, we developed a predictive model to identify customers who are likely to churn in a telecommunication company. We preprocessed the data, performed exploratory data analysis, and used machine learning algorithms to develop the model. We evaluated the performance of our model using metrics such as accuracy, precision, recall, and F1 score. Our model can be further improved by incorporating additional features and using more advanced machine learning algorithms. We have shown that logistic regression is an effective method for predicting customer churn in the telecommunications industry. By preprocessing the data and creating a logistic regression model, we were able to accurately predict whether a customer is likely to churn or not. This information can be valuable for telecommunications companies to retain customers and increase their revenue. Overall, our project demonstrates the importance of predicting and preventing customer churn in telecommunication companies.