

TELCO CHURN BI ANALYTICS MODEL REPORT

(Business Intelligence)

(TM3426)

Sixth Semester

Thapar Institute of

Engineering & Technology, Patiala

May 2023



ti

 **tslas**

Prepared By:
Jatin Bansal
142000052



01

Report Timeline

.....

TOPIC	PAGE NO
-------	---------

INTRODUCTION 04-05

RESEARCH METHODOLOGY 06

OBJECTIVES 07-08

METHODOLOGY 09-12

02

Project Timeline

o o o o

TOPIC	PAGE NO
BENEFITS AND IMPACT	13-14
CHANLLENGES AND CONSIDERATIONS	15-16
CONCLUSION	17-18
REFERENCES	19
APPENDIX	20-22
GITHUB REPOSITORY	23

Introduction



A little bit about the project

.....

This report aims to provide an overview of the Telco Churn Business Intelligence (BI) Model. Telco churn refers to the rate customers switch their telecommunication service provider. Understanding and predicting churn is crucial for telco companies as it directly impacts customer retention, revenue, and overall business performance.

The Telco Churn BI Model aims to leverage business intelligence techniques to analyze customer data and identify potential churners, allowing the telco company to retain customers and mitigate churn proactively. Through this model, the business seeks to increase efficiency by reducing resource use and unnecessary marketing expenses while focusing on customers who are leaving.

Research Methodology.....



The field of Churning is well-known in the Business Intelligence field. Thus, previous work is available on churn management; the piece tells how to handle churn and how it will be helpful in the formulation of business strategy, enhancing strategic value (Samalia et al., 2014). Because of public policies and standardization, people can easily switch to other mobile communication competitors. Thus, mobile carriers now focus on customer retention rather than acquisition (Qureshi et al., 2014).

03

OBJECTIVES....

The main objectives of the Telco Churn BI Model are as follows:

- a. Identify Churn Patterns: Analyze historical customer data to identify patterns and indicators associated with churn. This involves exploring customer demographics, usage patterns, service plans, billing information, and other relevant data.

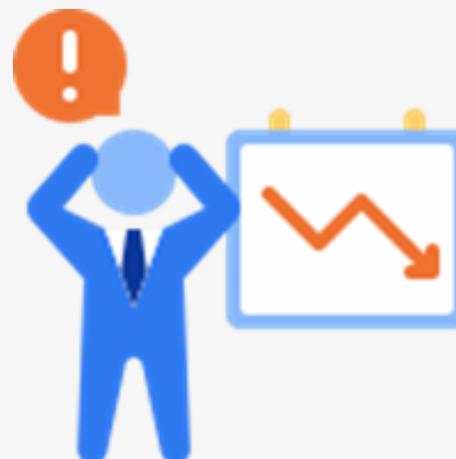


Figure 2: Identifying churn patterns

https://www.google.com/url?sa=i&url=https://www.flaticon.com/free-icons/churn&psig=AOvVaw0Ulq31zWD0kek nZ2u4kN9&ust=1684644699200000&source =images&cd=vfe&ved=0CBEQjRxqFwoTCMjx 6feUg_8CFQAAAAAdAAAAABD

- b. Predict Churns: Develop a predictive model to forecast the likelihood of a customer churning within a specific timeframe. This involves applying machine learning algorithms to historical customer data and training the model to predict churn probabilities based on various independent variables.

04

OBJECTIVES

.....

- c. Customer Segmentation: Segment customers based on their churn probability and other relevant characteristics. This allows the telco company to prioritize retention efforts and tailor marketing strategies based on different customer segments' specific needs and preferences.
- d. Actionable Insights: Provide actionable insights and recommendations to the telco company based on churn patterns and predictive model analysis. These insights can guide decision-making processes, such as implementing targeted marketing campaigns, improving customer service, or offering personalized incentives to reduce churn.



Figure 3: Targeting the mass in accordance with the churn prediction

Methodology

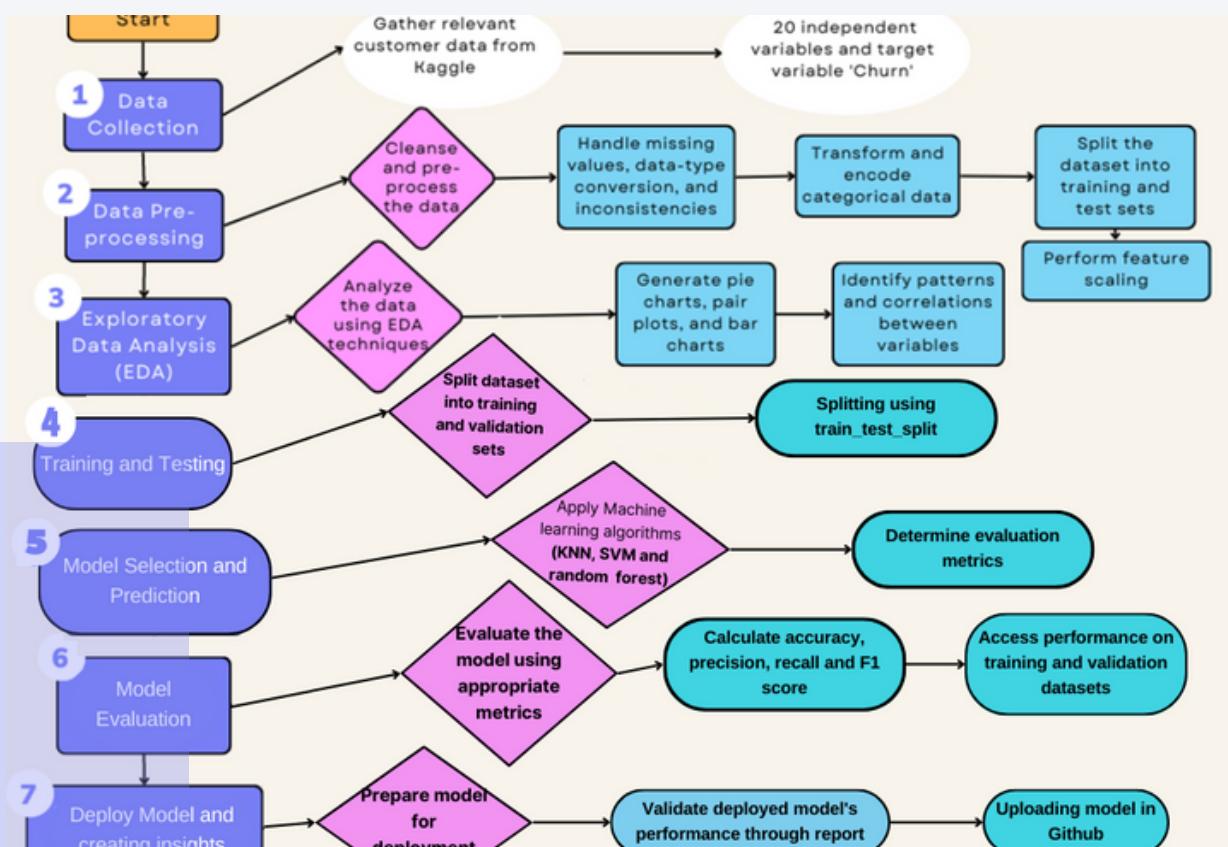


Figure 4: Methodology Flowchart



The Telco Churn BI Model follows a systematic methodology to achieve its objective:

a. Data Collection: Gathered relevant customer data from Kaggle, the world's largest data science community. The data collected for the business model consisted of 20 independent variables such as customerID, Gender, SeniorCitizen, Partner, Dependents, tenure, PhoneService, MultipleLines, InternetService, OnlineSecurity, OnlineBackup, DeviceProtection, TechSupport, StreamingTV, StreamingMovies, Contract, PaperlessBilling, PaymentMethod, MonthlyCharges and TotalCharges, and a target variable that's churning value named as 'Churn.'

The data can clearly define that we will be making a supervised learning model as the data consists of the labeled set having variables with individual identities. It is a 'Classification' model having two groups of target variables; one group consists of people who are leaving the telco company, and the other are not leaving the company.

b. Data Pre-processing: Cleanse and pre-process the collected data by importing the data set to handle missing values, data-type conversion, and inconsistencies. Transform, and encode categorical data, splitting the data set into training and test set, and feature scaling as required for analysis.



I checked the missing values in the BI model, but they were not there in the data set as they were hidden in the 'string' data type. Thus, I removed those rows with blank space, converted their data type into 'int,' and dropped a column named 'customerID' as it was unnecessary. Inconsistencies were tackled by converting various groups into Boolean values and handling categorical data by doing 'one-hot encoding.'

First, the data set was divided into two parts, 'X' consisting of independent variables and 'Y' consisting of dependent variables. It further got split into training and test set. The model also standardizes the value of the Independent variable using the 'Normalization' method.

c. Exploratory Data Analysis (EDA): Perform exploratory data analysis to gain insights into the data, identify patterns, and uncover correlations between variables using pie charts, pair plots, and bar charts. The above analytical techniques help in understanding the underlying factors contributing to churn.

d. Training and Testing: The data divided into parts of 'X' and 'Y' is further split into training and test data using `sklearn.model_selection` module.



- e. Model Selection and Prediction: Apply appropriate machine learning algorithms such as KNN, Support Vector Machine, and random forests to train the churn prediction model. Thus, predictions are made by taking previous training data as the reference that helped to train the model.
- f. Model Evaluation: Evaluate the trained model using appropriate evaluation metrics such as accuracy, precision, recall, and F1 score. Could you assess the model's performance on training and validation datasets to ensure reliability and generalization?
- g. Deploying Model and creating Insights: Generating reports and visualization to present the findings and insights derived from the Telco Churn BI Model. And, thus, uploading the insights on a platform like Github.

Furthermore, the appendix includes sample code for the document's main body. I have made the model in Jupyter Notebook using Python.

Benefits and Impacts

06

Implementing a Telco Churn BI Model can provide several benefits and have a significant impact on the telco company:

- a. Improved Customer Retention: The telco company can proactively improve customer retention by identifying potential churners in advance. This may include targeted marketing campaigns, personalized offers, improved customer service, or loyalty programs designed to address customers' specific needs and concerns at risk of churn.
- b. Increased Revenue: Retaining customers is more cost-effective than acquiring new ones. The telco company can stabilize its customer base and increase revenue by reducing churn rates. Additionally, identifying cross-selling or upselling opportunities for existing customers can contribute to revenue growth.
- c. Enhanced Customer Experience: Understanding churn patterns and customer preferences allows the telco company to tailor its products and services to meet customer expectations better. By providing a personalized and seamless experience, customer satisfaction and loyalty can be significantly improved.

d. Competitive Advantage: A Telco Churn BI Model can provide a competitive edge in the highly competitive telecommunications industry. The ability to anticipate and prevent churn puts the telco company in a better position to retain customers and attract new ones, ultimately gaining an advantage over its competitors.

e. Data-Driven Decision Making: Implementing a Telco Churn BI Model establishes a data-driven approach to decision-making. The insights and recommendations derived from the model can guide strategic initiatives, marketing campaigns, and operational improvements, leading to more informed and effective decision-making across the organization.

BENEFITS OF DATA-DRIVEN DECISION MAKING

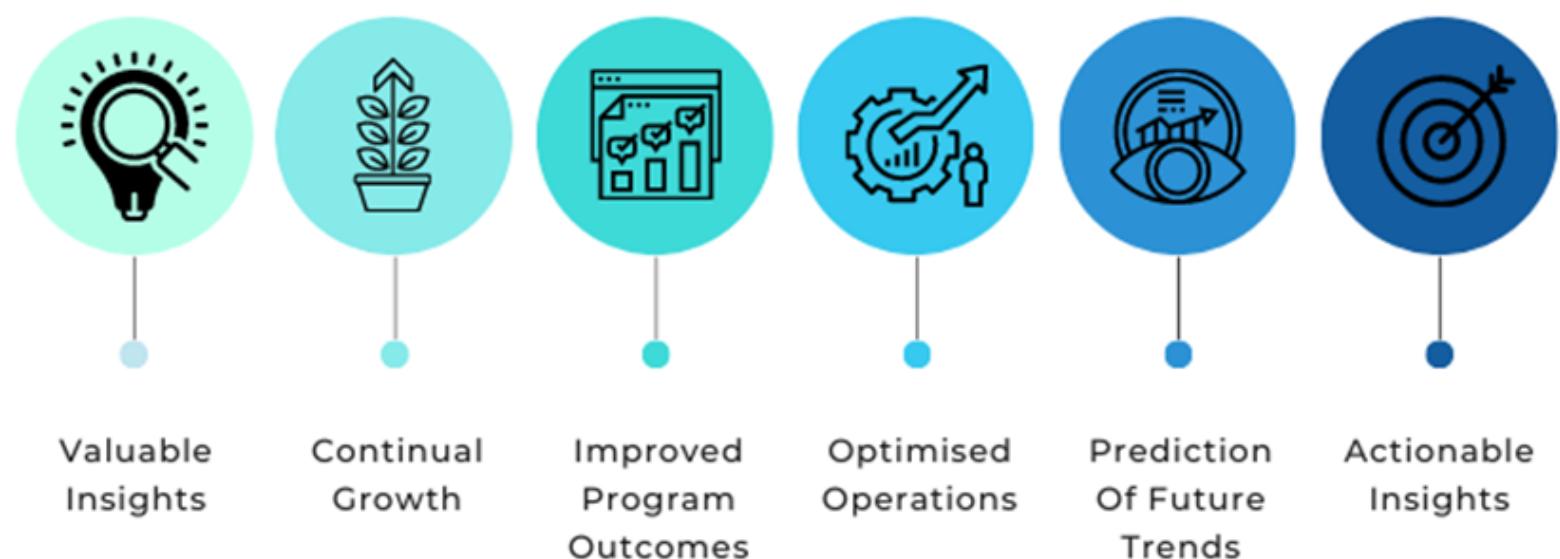


Figure 5: Data-Driven Decision Making

<https://www.altisconsulting.com/au/wp-content/uploads/sites/4/2020/09/Jasons-Blog-Image.png>

Challenges and Considerations

07

While the Telco Churn BI Model offers significant benefits, some specific challenges and considerations need to be addressed:

- a. Data Quality and Availability: The model's success relies on high-quality and comprehensive customer data availability. Ensuring data accuracy, consistency, and completeness can be challenging, requiring data governance processes and integration across different systems. The data available on Kaggle can have a series of data accuracy issues.

- b. Model Interpretability: While machine learning models can provide accurate predictions, they may need more interpretability. It is essential to strike a balance between model accuracy and the ability to explain the factors driving churn predictions. This can help gain the trust of stakeholders and facilitate the implementation of actionable recommendations.

c. Scalability: As the telco customer base grows, the model should be able to handle increasing volumes of data and adapt to changing customer dynamics. Scalability considerations should be considered during model development to ensure effectiveness as the business scales. The data-set that I took is having data of 7043 customers and measuring 20-21 variables is really difficult to analyse. Thus, model working issues when data-set scales up can be there.



Figure 6

https://upload.wikimedia.org/wikipedia/commons/7/7c/Kaggle_logo.png

CONCLUSION



0.8

EXIT

"Data-driven decisions driven by our BI analytics churn model reduce churn and fuel business growth." ns on this project.

The Telco Churn Business Intelligence Model offers a powerful solution for telco companies to manage customer churn proactively. By leveraging data analytics and machine learning techniques, telcos can gain valuable insights into churn patterns, predict churn probability, segment customers, and take proactive measures to retain valuable customers. Implementing such a model can improve customer retention, increase revenue, enhance customer experience, and a competitive advantage in the telecommunications industry. However, addressing data quality, interpretability, scalability, and ethical considerations is crucial for successfully implementing and deploying the Telco Churn BI Model. And use the Support Vector Machine algorithm as it has the highest accuracy level, which is 80%. Thus, the model can be considered successful.



The code demonstrates the visualization of different attributes for churners and non-churners. The plotted details provide valuable insights into customer behavior and preferences. For example, comparing characteristics like partner status, dependents, and paperless billing can indicate the influence of personal relationships and billing preferences on churn rates. Thus, the visual comparison of attributes can inform business strategies to reduce churn rates.

"Data-driven decisions driven by our BI analytics churn model reduce churn and fuel business growth." ns on this project.



REFERENCES 09

Qureshi, Rehman, Qamar, Kamal and Rehman. Telecommunication subscribers' churn prediction model using machine learning. Eighth International Conference on Digital Information Management (ICDIM 2013), September 2013.

Singh and Samalia. A Business Intelligence Perspective for Churn Management. 2nd World Conference On Business, Economics And Management - WCBEM 2013.

codebasics. (2018, December 23). Machine Learning Tutorial Python - 11 Random Forest [Video]. YouTube. Retrieved May 18, 2023, from <https://youtu.be/ok2s1vV9XWO>

codebasics. (2019, December 19). Machine Learning Tutorial Python - 10 Support Vector Machine (SVM) [Video]. YouTube. Retrieved May 18, 2023, from <https://youtu.be/FB5EdxAGxQg>

codebasics. (2020, September 07). Customer churn prediction using ANN | Deep Learning Tutorial 18 (Tensorflow2.0, Keras & Python) [Video]. YouTube. Retrieved May 08, 2023, from <https://youtu.be/MSBY28IJ47U>

APPENDIX

10

Data Pre-Processing

```
In [120]: df.info()
Out[120]: 
customerID    int64
gender        object
SeniorCitizen   int64
Partner        object
Dependents     int64
tenure         int64
PhoneService    object
MultipleLines   object
InternetService object
OnlineSecurity   object
OnlineBackup    object
DeviceProtection object
TechSupport     object
StreamingTV    object
StreamingMovies object
Contract        object
PaperlessBilling int64
PaymentMethod    object
MonthlyCharges  float64
TotalCharges    float64

In [121]: df.drop('customerID',axis=1,inplace=True)

In [122]: df.head()

Out[122]:
   gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection TechSupport
0   Female            0       0           0      1      No phone service        DSL          0           0           1           0           0
1     Male            0       0           0      34      Yes              0          DSL          1           0           1           0
2     Male            0       0           0       2      Yes              0          DSL          1           1           1           0
3     Male            0       0           0      45      No phone service        DSL          0           1           0           0           1
4   Female            0       0           0       2      Yes              0      Fiber optic          0           0           0           0           0
```

```
In [123]: df=df[df['TotalCharges']!=' ']

In [124]: df.head()

Out[124]:
   gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection TechSupport
0   Female            0       0           0      1      No phone service        DSL          0           0           1           0           0
1     Male            0       0           0      34      Yes              0          DSL          1           0           1           0
2     Male            0       0           0       2      Yes              0          DSL          1           1           1           0
3     Male            0       0           0      45      No phone service        DSL          0           1           0           0           1
```

Data-visualization

```
# Set the common title for the figure
fig.suptitle('Churners vs Non-Churners', fontweight='bold', color='black', y=1.00)

# Add the subtitle separately
fig.text(0.5, 1.02, 'Comparison of different attributes between Churners and Non-Churners', fontweight='normal', color='black', ha='center', va='bottom', size=12, style='italic')

# Ensure tight layout to prevent overlapping labels
plt.tight_layout()

# Show the plots
plt.show()
```

```
In [534]: #Dependent and Independent variables
X = df.iloc[:,1:13]
Y = df.iloc[:,13:14]

In [535]: X['gender'].replace({'Female':1, 'Male':0}, inplace=True)

In [536]: # handling catg data through one-hot encoding
X=pd.get_dummies(X)

In [537]: X.head()

Out[537]:
   gender SeniorCitizen Partner Dependents tenure PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection TechSupport
0   Female            0       0           0      1      No phone service        DSL          0           0           1           0           0
1     Male            0       0           0      34      Yes              0          DSL          1           0           1           0
2     Male            0       0           0       2      Yes              0          DSL          1           1           1           0
3     Male            0       0           0      45      No phone service        DSL          0           1           0           0           1
```

Training and Testing the Model

```

jupyter Customer Churn Business Intelligence Model Last Checkpoint: 05/23/2023 (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) Logout
In [189]: from sklearn.model_selection import train_test_split
X_train, Y_train, X_test, Y_test = train_test_split(X, Y, test_size=0.25, random_state=1)

In [190]: from sklearn.model_selection import train_test_split
X1_train, X1_test, Y1_train, Y1_test = train_test_split(X1, Y1, test_size=0.25, random_state=1)

In [191]: X_train
Out[191]:
   gender SeniorCitizen Partner Dependents Tenure PhoneService MultipleLines InternetService OnlineSecurity OnlineBackup DeviceProtection TechSupport
4533    1            0       1        0     71          1         1      Fiber optic           0        1        1
5074    1            0       0        1     49          1         0           0        0        0        0
2693    0            0       0        0     72          1         0           0        0        0        0
6095    0            0       0        0     4          1         0           0        0        0        0
4552    1            0       0        0     9          1         0           0        0        0        0
... ...
907     0            0       0        0     19          1         0           0        0        1        0
5200    1            0       1        1     70          1         1      Fiber optic           0        1        1
3987    1            0       0        0     11          1         1      Fiber optic           0        0        0
235     0            0       0        0     2          1         1           0        1        0        0
5165    0            1       0        0     1          1         0           0        0        0        0
5274 rows × 19 columns

In [192]: X_test

```

KNN-Classification Algorithm

```

jupyter Customer Churn Business Intelligence Model Last Checkpoint: 13 minutes ago (autosaved)
File Edit View Insert Cell Kernel Widgets Help Trusted Python 3 (ipykernel) Logout
In [1]: # Model Creation ----KNN
from sklearn.neighbors import KNeighborsClassifier
knn=KNeighborsClassifier(n_neighbors=5) # model is created on training dataset

In [2]: #KNN Model Training
knn.fit(X1_train, Y1_train)

In [3]: #KNN Model Testing
# Prediction
Y2_pred=knn.predict(X1_test)
Y2_pred

In [4]: print("Classification Report Algorithm 2 is:\n",classification_report(Y1_test, Y2_pred))
Classification Report Algorithm 2 is:
precision    recall    f1-score   support
          0       0.84      0.90      0.87     1294
          1       0.66      0.51      0.58      464

   accuracy                           0.80      1758
  macro avg       0.75      0.71      0.72      1758
weighted avg       0.79      0.80      0.79      1758

In [5]: acc2 = accuracy_score(Y1_test, Y2_pred)*100
acc2
Out[5]: 80.09412060283276

In [6]: from sklearn.metrics import confusion_matrix
cm2=confusion_matrix(Y1_test, Y2_pred)
cm2

Out[6]: array([[1169,  121],
       [ 208,  231]], dtype=int64)

In [7]: plt.figure(figsize=(10,7))
sns.heatmap(cm2, annot=True, fmt=".2f")
plt.title("Heatmap on KNN-Classification Model")
plt.xlabel("Predicted")
plt.ylabel("Truth")
plt.show()

```



Jupyter Customer Churn Business Intelligence Model Last Checkpoint: 12 minutes ago (autosaved)

```
In [554]: # Model Creation ---- SVM(Support Vector Machine)
from sklearn.svm import SVC
model=SVC(kernel='rbf')# model is created
In [555]: # SVM Model Training
model.fit(X1_train, Y1_train)
#A second liblinear-pkg from sklearn/utils/validation.py:993: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example via y.ravel().
y = column_or_1d(y, warn=True)
Out[555]: SVC()

In [556]: # SVM Model Testing
Y2_pred=model.predict(X1_test)
Y2_pred
Out[556]: array([0, 1, 0, ..., 0, 0, 0])

In [557]: from sklearn.metrics import classification_report
print("Classification Report Algorithm 1 is:\n",
      ">",classification_report(Y1_test, Y2_pred))
Classification Report Algorithm 1 is:
precision    recall   f1-score   support
          0       0.84      0.90      0.87     1294
          1       0.66      0.51      0.58      464

   accuracy                           0.80      1758
  macro avg       0.75      0.71      0.72      1758
weighted avg       0.79      0.80      0.79      1758

In [558]: from sklearn.metrics import accuracy_score
acc1=accuracy_score(Y1_test, Y2_pred)*100
acc1
Out[558]: 80.034196283276

In [559]: from sklearn.metrics import confusion_matrix
cm1=confusion_matrix(Y1_test, Y2_pred)
cm1
```

Support Vector Algorithm

Jupyter Customer Churn Business Intelligence Model Last Checkpoint: 13 minutes ago (autosaved)

```
In [564]: # Model Creation ---- Random Forest Model
from sklearn.ensemble import RandomForestClassifier as RFC
random = RFC(n_estimators=40)

In [565]: #Random Forest Model Training
random.fit(X1_train, Y1_train)
#C:\Users\DELL\AppData\Local\Temp\ipykernel_11100\578160665.py:2: DataConversionWarning: A column-vector y was passed when a 1d array was expected. Please change the shape of y to (n_samples,), for example via y.ravel().
array(..., dtype=uint8)[:, newaxis]
random.fit(X1_train, Y1_train)

Out[565]: RandomForestClassifier(n_estimators=40)

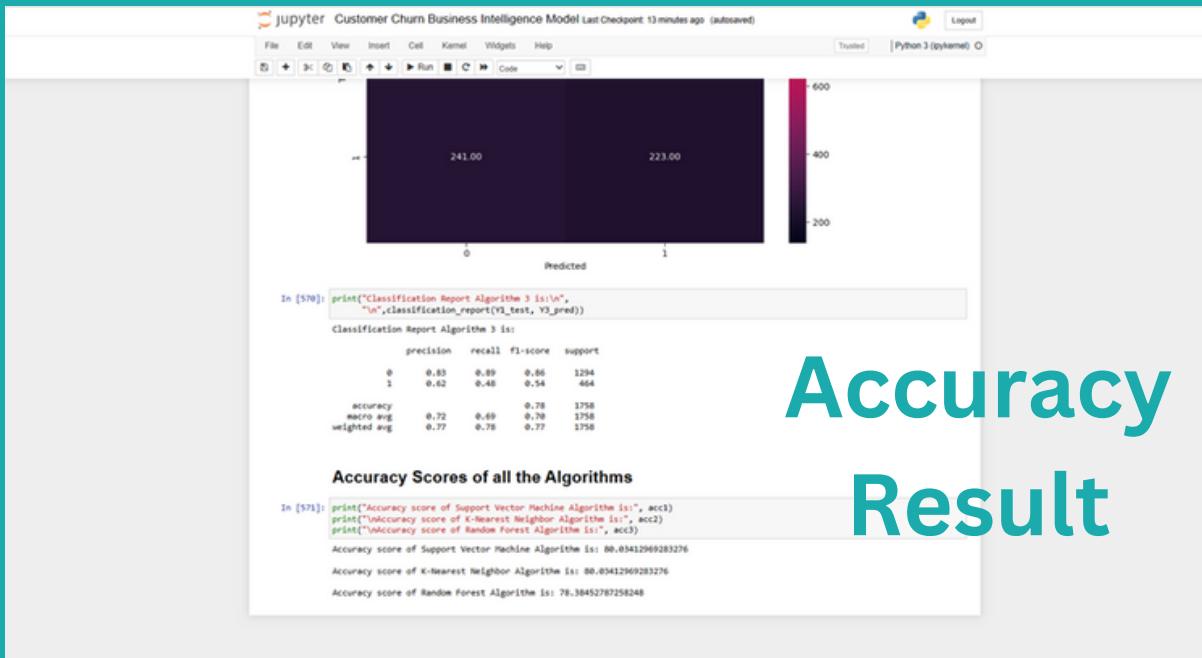
In [566]: #Random Forest Model Testing
Y3_pred=random.predict(X1_test)
Y3_pred
Out[566]: array([0, 1, 0, ..., 0, 0, 0])

In [567]: acc2=accuracy_score(Y1_test, Y3_pred)*100
acc2
Out[567]: 78.36453757258248

In [568]: cm2=confusion_matrix(Y1_test, Y3_pred)
cm2
Out[568]: array([[1155, 139],
               [241, 223]], dtype=int64)

In [569]: plt.figure(figsize=(10,7))
plt.title("Heatmap on Random Forest Model")
plt.xlabel('Predicted')
plt.ylabel('Truth')
plt.show()
```

Heatmap on Random Forest Model



Accuracy Result

• • • •

GitHub Repository: 10

