

SRUTHI KRISHNA K.S

962221106101

CUSTOMER CHURN PREDICTION:

PHASE 4: PROJECT SUBMISSION

Development part 2

INTRODUCTION:

Churn prediction means detecting which customers are likely to leave a service or to cancel a subscription to a service. It is a critical prediction for many businesses because acquiring new clients often costs more than retaining existing ones. Once you can identify those customers that are at risk of cancelling, you should know exactly what marketing action to take for each individual customer to maximise the chances that the customer will remain.

**WHY IT IS SO IMPORTANT?**

Customer churn is a common problem across businesses in many sectors. If you want to grow as a company, you have to invest in acquiring new clients. Every time a client leaves, it represents a significant investment lost. Both time and effort need to be channelled into replacing them. Being able to predict when a client is likely to leave, and offer them incentives to stay, can offer huge savings to a business.

As a result, understanding what keeps customers engaged is extremely valuable knowledge, as it can help you to develop your retention strategies, and to roll out operational practices aimed at keeping customers from walking out the door.

**WHAT ARE THE MAIN CHALLENGES?**

**1) To succeed at retaining customers who are ready to abandon your business, Marketers & Customer Success experts must be able to predict in advance which customers are going to churn and set up a plan of marketing actions that will have the greatest retention impact on each customer. The key here is to be proactive and engage with these customers. While simple in theory, the realities involved with achieving this “proactive retention” goal are extremely challenging.**

**2)The accuracy of the technique is critical to the success of any proactive retention efforts. If the Marketer is unaware of a customer about to churn, no action will be taken to retain that customer.**

**3)Special retention-focused offers or incentives may be provided to happy, active customers, resulting in reduced revenues for no good reason.**

**4)Your churn prediction model should rely on (almost) real-time data to quantify the risk of churning, not on static data. Although you will be able to identify a certain percentage of at-risk customers with even static data, your predictions will be inaccurate.**

**DATASHEET RELATED TO PYCARET :**

```
# import libraries
```

```
import pandas as pd
```

```
import numpy as np
```

```
# read csv data
```

```
data = pd.read_csv('https://raw.githubusercontent.com/srees1988/predict-churn-py/main/customer_churn_data.csv')
```

	customerID	gender	SeniorCitizen	Partner	Dependents	tenure	PhoneService	MultipleLines	InternetService	OnlineSecurity	...	DeviceProtection	TechSup
0	7590-VHVEG	Female	0	Yes	No	1	No	No phone service	DSL	No	...	No	
1	5575-GNVDE	Male	0	No	No	34	Yes	No	DSL	Yes	...	Yes	
2	3668-QPYBK	Male	0	No	No	2	Yes	No	DSL	Yes	...	No	
3	7795-CFOCW	Male	0	No	No	45	No	No phone service	DSL	Yes	...	Yes	
4	9237-HQITU	Female	0	No	No	2	Yes	No	Fiber optic	No	...	No	

5 rows × 21 columns

## EXPLORATORY DATA ANALYSIS:

customerID	object
gender	object
SeniorCitizen	int64
Partner	object
Dependents	object
tenure	int64
PhoneService	object
MultipleLines	object
InternetService	object
OnlineSecurity	object
OnlineBackup	object
DeviceProtection	object
TechSupport	object
StreamingTV	object
StreamingMovies	object
Contract	object
PaperlessBilling	object
PaymentMethod	object
MonthlyCharges	float64
TotalCharges	object
Churn	object
dtype:	object

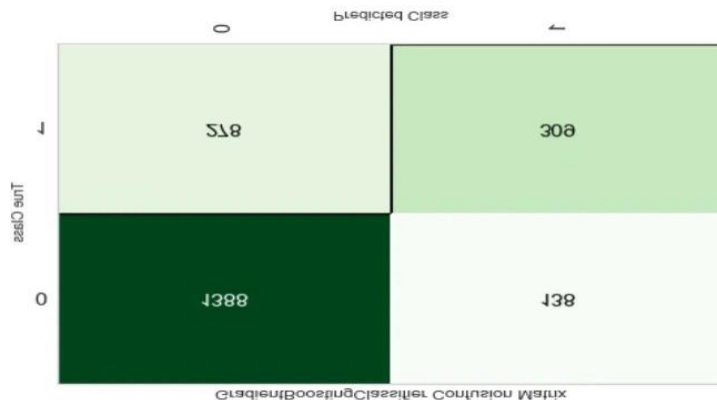
MODEL ANALYSIS:

# AUC Plot

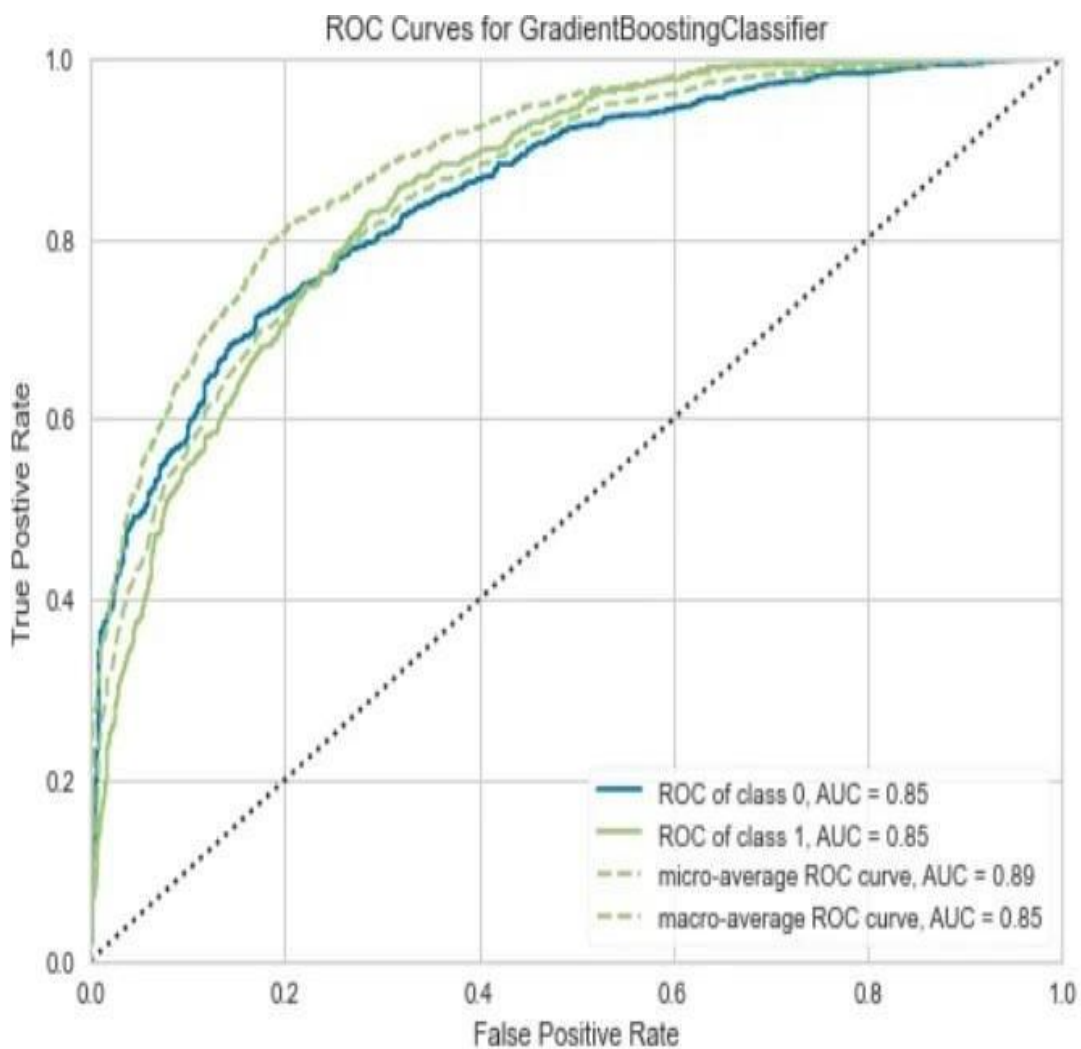
plot\_model(tuned\_best\_model, plot = 'auc')

# Confusion Matrix

plot\_model(tuned\_best\_model, plot =



'confusion\_matrix')



In a churn model, often the reward of true positives is way different than the cost of false positives. Let's use the following assumptions:

\$1,000 voucher will be offered to all the customers identified as churn (True Positive + False Positive);

If we are able to stop the churn, we will gain \$5,000 in customer lifetime value.

Master of Science in Business Analytics

Master of Science in Business Analytics

Customer Churn Prediction Using Machine Learning: Main Approaches and Models

We reach out to experts from HubSpot and ScienceSoft to discuss how SaaS companies handle the problem of customer churn prediction using Machine Learning.

FacebookTwitterLinkedInRedditEmailShare

comments

By Altexsoft.

Customer retention is one of the primary growth pillars for products with a subscription-based business model. Competition is tough in the SaaS market where customers are free to choose from plenty of providers even within one product category. Several bad experiences – or even one – and a customer may quit. And if droves of unsatisfied customers churn at a clip, both material losses and damage to reputation would be enormous.

For this article, we reached out to experts from HubSpot and ScienceSoft to discuss how SaaS companies handle the problem of customer churn with predictive modeling. You will discover approaches and best practices for solving this problem. We'll discuss collecting data about client relationship with a brand, characteristics of customer behavior that correlate the most with churn and explore the logic behind selecting the best-performing machine learning models.

What does the future hold for digital chemistry data?

What does the future hold for digital chemistry data?

What is customer churn?

Customer churn (or customer attrition) is a tendency of customers to abandon a brand and stop being a paying client of a particular business. The percentage of customers that discontinue using a company's products or services during a particular time period is called a customer churn (attrition) rate. One of the ways to calculate a churn rate is to divide the number of customers lost during a given time interval by the number of acquired customers, and then multiply that number by 100 percent. For example, if you got 150 customers and lost three last month, then your monthly churn rate is 2 percent.

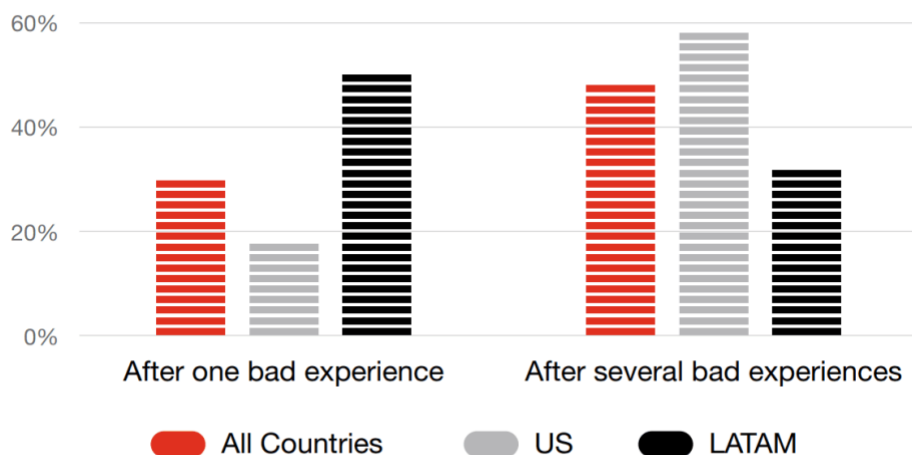
Churn rate is a health indicator for businesses whose customers are subscribers and paying for services on a recurring basis, notes head of data analytics department at ScienceSoft Alex Bekker, "Customers [of subscription-driven businesses] opt for a product or a service for a particular period, which can be rather short – say, a month. Thus, a customer stays open for more interesting or advantageous offers. Plus, each time their current commitment ends, customers have a chance to reconsider and choose not to continue with the company. Of

course, some natural churn is inevitable, and the figure differs from industry to industry. But having a higher churn figure than that is a definite sign that a business is doing something wrong.”

There are many things brands may do wrong, from complicated onboarding when customers aren’t given easy-to-understand information about product usage and its capabilities to poor communication, e.g. the lack of feedback or delayed answers to queries. Another situation: Longtime clients may feel unappreciated because they don’t get as many bonuses as the new ones.

In general, it’s the overall customer experience that defines brand perception and influences how customers recognize value for money of products or services they use.

**Figure 4:** When do consumers stop interacting with a brand they love?



Q: At what point would you stop interacting with a company that you love shopping at or using?  
Source: PwC Future of Customer Experience Survey 2017/18

**DETECTING CUSTOMERS AT RISK OF CHURN :**



## Advertisement

### Customer Churn Prediction Using Machine Learning: Main Approaches and Models

We reach out to experts from HubSpot and ScienceSoft to discuss how SaaS companies handle the problem of customer churn prediction using Machine Learning.

FacebookTwitterLinkedInRedditEmailShare

comments

By Altexsoft.

Customer retention is one of the primary growth pillars for products with a subscription-based business model. Competition is tough in the SaaS market where customers are free to choose from plenty of providers even within one product category. Several bad experiences – or even one – and a customer may quit. And if droves of unsatisfied customers churn at a clip, both material losses and damage to reputation would be enormous.

For this article, we reached out to experts from HubSpot and ScienceSoft to discuss how SaaS companies handle the problem of customer churn with predictive modeling. You will discover approaches and best practices for solving this problem. We'll discuss collecting data about client relationship with a brand, characteristics of customer behavior that correlate the most with churn and explore the logic behind selecting the best-performing machine learning models.

What does the future hold for digital chemistry data?

What does the future hold for digital chemistry data?

What is customer churn?

Customer churn (or customer attrition) is a tendency of customers to abandon a brand and stop being a paying client of a particular business. The percentage of customers that discontinue using a company's products or services during a particular time period is called a customer churn (attrition) rate. One of the ways to calculate a churn rate is to divide the number of customers lost during a given time interval by the number of acquired customers, and then multiply that number by 100 percent. For example, if you got 150 customers and lost three last month, then your monthly churn rate is 2 percent.

Churn rate is a health indicator for businesses whose customers are subscribers and paying for services on a recurring basis, notes head of data analytics department at ScienceSoft Alex Bekker, "Customers [of subscription-driven businesses] opt for a product or a service for a particular period, which can be rather short – say, a month. Thus, a customer stays open for more interesting or advantageous offers. Plus, each time their current commitment ends, customers have a chance to reconsider and choose not to continue with the company. Of course, some natural churn is inevitable, and the figure differs from industry to industry. But having a higher churn figure than that is a definite sign that a business is doing something wrong."

There are many things brands may do wrong, from complicated onboarding when customers aren't given easy-to-understand information about product usage and its capabilities to poor communication, e.g. the lack of feedback or delayed answers to queries. Another situation: Longtime clients may feel unappreciated because they don't get as many bonuses as the new ones.

In general, it's the overall customer experience that defines brand perception and influences how customers recognize value for money of products or services they use.

The reality is that even loyal customers won't tolerate a brand if they've had one or several issues with it. For instance, 59 percent of US respondents to the survey by PricewaterhouseCoopers (PwC) noted that they will say goodbye to a brand after several bad experiences, and 17 percent of them after just one bad experience.

Bad experiences may alienate even loyal customers. Source: PwC

Impact of customer churn on businesses

Well, churn is bad. But how exactly does it affect company performance in the long run?

Don't underestimate the impact of even a tiny percentage of churn, says Michael Redbord, general manager of Service Hub at HubSpot. "In a subscription-based business, even a small rate of monthly/quarterly churn will compound quickly over time. Just 1 percent monthly churn translates to almost 12 percent yearly churn. Given that it's far more expensive to acquire a new customer than to retain an existing one, businesses with high churn rates will quickly find themselves in a financial hole as they have to devote more and more resources to new customer acquisition."

Many surveys focusing on customer acquisition and retention costs are available online. According to this one by Invesp, conversion rate optimization company, getting a new customer may cost up to five times more than retaining an existing customer.

Churn rates do correlate with lost revenue and increased acquisition spend. In addition, they play a more nuanced role in a company's growth potential, continues Michael, "Today's buyers aren't shy about sharing their experiences with vendors through channels like review sites and social media, as well as peer-to-peer networks. HubSpot Research found that 49 percent of buyers reported sharing an experience they had with a company on social media. In a world of eroding trust in businesses, word of mouth plays a more critical role in the buying process than ever before. From the same HubSpot Research study, 55 percent of buyers no longer trust the companies they buy from as much as they used to, 65 percent don't trust company press releases, 69 percent don't trust advertisements, and 71 percent don't trust sponsored ads on social networks."

A glance at the state of customer trust towards businesses. Source: HubSpot Research Trust Survey

The expert concludes that companies with high churn rates aren't only failing to deliver in their relationships with ex-customers but also damage their future acquisition efforts by creating negative word-of-mouth around their products.

CallMiner conversational analytics solutions provider interviewed 1000 adults to learn why and how they interact with companies. The survey revealed that US businesses lose about \$136 billion a year due to customer attrition. What's more,

the company behaviors that caused customers to cut ties with brands could have been corrected.

### Use cases for customer churn prediction

As we mentioned before, churn rate is one of the critical performance indicators for subscription businesses. The subscription business model – pioneered by English book publishers in the 17th century – is very popular among modern service providers. Let's take a quick look at these companies:

Music and video streaming services are probably the most commonly associated with the subscription business model (Netflix, YouTube, Apple Music, Google Play, Spotify, Hulu, Amazon Video, Deezer, etc.).

Media. Digital presence is a must among the press, so news companies offer readers digital subscriptions besides print ones (Bloomberg, The Guardian, Financial Times, The New York Times, Medium etc.).

Telecom companies (cable or wireless). These companies may provide a full range of products and services, including wireless network, internet, TV, cell phone, and home phone services (AT&T, Sprint, Verizon, Cox Communications, etc.). Some specialize in mobile telecommunications (China Mobile, Vodafone, T-Mobile, etc.).

Software as a service providers. The adoption of cloud-hosted software is growing. According to Gartner, the SaaS market remains the largest segment of the cloud market. Its revenue is expected to grow 17.8 percent and reach \$85.1 billion in 2019. The product range of SaaS providers is extensive: graphic and video editing (Adobe Creative Cloud, Canva), accounting (Sage 50cloud, FreshBooks),

eCommerce (BigCommerce, Shopify), email marketing (MailChimp, Zoho Campaigns), and many others.

These company types may use churn rate to measure the effectiveness of cross-department operations and product management.

Identifying at-risk customers with machine learning: problem-solving at a glance

Companies that constantly monitor how people engage with products, encourage clients to share opinions, and solve their issues promptly have greater opportunities to maintain mutually beneficial client relationships.

And now imagine a company that has been gathering customer data for a while, so it can use it to identify behavior patterns of potential churners, segment these at-risk customers, and take appropriate actions to gain back their trust. Those following a proactive approach to customer churn management use predictive analytics. That's one of four analytics types that entails forecasting the probability of future outcomes, events, or values by analyzing current and historical data. Predictive analytics utilizes various statistical techniques, such as data mining (pattern recognition) and machine learning (ML).

"The one weakness of tracking just real churn is that it serves only as a lagging indicator of poor customer experience, which is where a predictive churn model becomes extremely valuable," notes Michael Redbord from HubSpot.

The main trait of machine learning is building systems capable of finding patterns in data, learning from it without explicit programming. In the context of customer

churn prediction, these are online behavior characteristics that indicate decreasing customer satisfaction from using company services/products.

Detecting customers at risk of churn helps take measures in advance

ScienceSoft's Alex Bekker also stresses the importance of machine learning for proactive churn management: "As to identifying potential churners, machine learning algorithms can do a great job here. They reveal some shared behavior patterns of those customers who have already left the company. Then, ML algorithms check the behavior of current customers against such patterns and signal if they discover potential churners."

Subscription-based businesses leverage ML for predictive analytics to find out which current users aren't fully satisfied with their services and address their issues when it's not too late: "Identifying customers at risk of churn as many as 11 months before their renewal enables our customer success team to engage these customers, understand their pain points, and with them, put together a long term plan focused on helping the customer realize value from the service they bought," explains Michael.

Use cases for predictive churn modeling go beyond proactive engagement with prospective churning customers and selecting effective retention actions. According to Redbord, ML-based software allows customer success managers to define which customers they should contact. In other words, employees can be sure they're speaking with the right customers at the right time.

Sales, customer success, and marketing teams can also use the knowledge from the data analysis to align their actions. “For example, if a customer is showing signs of churn risk, that’s probably not a great time for sales to reach out with information about additional services the customer might be interested in. Rather, that engagement should be with the CSM so they can help the customer become re-engaged and see value in the products they currently have. Like sales, marketing can engage with customers differently depending on their current indication of churn risk: For example, non-churn risk customers are better candidates to participate in a case study than a customer who is currently a churn risk,” the expert of HubSpot explains. Generally speaking, the strategy of customer interaction should be based on ethics and sense of timing. And using machine learning for customer data analysis can bring insights to power this strategy.

#### STEPS FOR EFFECTIVE CHURN PREDICTION:

##### 1). Reliable customer segmentation

Churn prediction is entirely based around the use of your company’s historical data on your customer. You’ll need your customer analytics to predict how customer churn is affecting your business accurately. Begin by exporting all historical data types that could potentially affect a customer’s likelihood to churn.

Demographics and behavioral data:

- \*)Is this user a single user or using your product on behalf of their company?
- \*)What is their rate of use of the product in general and the specific features that are available to them?
- \*)How often does this user submit tickets to your support service?

Importing Telco Churn dataset:

```
# Import required libraries
```

```
import numpy as np
```



```

import pandas as pd

# Import the dataset

dataset = pd.read_csv('telcochurndata.csv')

# Glance at the first five records

dataset.head()

# Print all the features of the data

dataset.columns

```

	State	Account length	Area code	International plan	Voice mail plan	Number vmail messages	Total day minutes	Total day calls	Total day charge	Total eve minutes	Total eve calls	Total eve charge	Total night minutes	Total night calls	Total night charge	Total intl minutes	Total intl calls	Total intl charge
0	KS	128	415	No	Yes	25	285.1	110	45.07	197.4	99	16.78	244.7	91	11.01	10.0	3	2.70
1	OH	107	415	No	Yes	26	161.6	123	27.47	195.5	103	16.62	254.4	103	11.45	13.7	3	3.70
2	NJ	137	415	No	No	0	243.4	114	41.38	121.2	110	10.30	162.6	104	7.32	12.2	5	3.29
3	OH	84	408	Yes	No	0	299.4	71	50.90	61.9	88	5.26	196.9	89	8.86	6.6	7	1.78
4	OK	75	415	Yes	No	0	186.7	113	28.34	148.3	122	12.61	186.9	121	8.41	10.1	3	2.73

Exploring Data Visualizations : To understand how variables are distributed.

```

# Import matplotlib and seaborn

import matplotlib.pyplot as plt

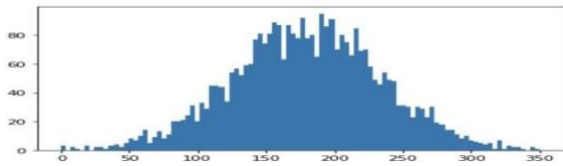
import seaborn as sns

```

```
# Visualize the distribution of 'Total day minutes'
```

```
plt.hist(dataset['Total day minutes'], bins = 100)
```

```
# Display the plot
```



CONCLUSION:

Features and Labels

```
X = dataset.iloc[:, 0:19].values
```

```
y = dataset.iloc[:, 19].values # Churn
```

```
# Encoding categorical data in X
```

```
from sklearn.preprocessing import LabelEncoder
```

```
labelencoder_X_1 = LabelEncoder()
```

```
X[:, 3] = labelencoder_X_1.fit_transform(X[:, 3])
```

```
labelencoder_X_2 = LabelEncoder()
```

```
X[:, 4] = labelencoder_X_2.fit_transform(X[:, 4])
```

```
# Encoding categorical data in y
```

```
labelencoder_y = LabelEncoder()
```

```
y = labelencoder_y.fit_transform(y)
```

Code: Encoding State feature using One hot encoding

```
# Removing extra column to avoid dummy variable trap
```

```
X_State = pd.get_dummies(X[:, 0], drop_first = True)
```

```
# Converting X to a dataframe
```

```
X = pd.DataFrame(X)
```

```
# Dropping the 'State' column
```

```
X = X.drop([0], axis = 1)
```

```
# Merging two dataframes
```

```
frames = [X_State, X]
```

```
result = pd.concat(frames, axis = 1, ignore_index = True)
```

```
# Final dataset with all numeric features
```

```
X = result
```

Code : To Create Training and Test sets

```
# Splitting the dataset into the Training and Test sets
```

```
from sklearn.model_selection import train_test_split
```

```
X_train, X_test, y_train, y_test = train_test_split(X, y,  
                                                    test_size = 0.2,
```

```
                                                    random_state = 0)
```

Code: To scale features of the training and test sets

```
# Feature Scaling
```

```
from sklearn.preprocessing import StandardScaler
```

```
sc = StandardScaler()
```

```
X_train = sc.fit_transform(X_train)
```

```
X_test = sc.transform(X_test)
```

Code: To train a Random Forest classifier model on the training set.

```
# Import RandomForestClassifier
```

```
from sklearn.ensemble import RandomForestClassifier
```

```
# Instantiate the classifier
```

```
clf = RandomForestClassifier()
```

```
# Fit to the training data
```

```
clf.fit(X_train, y_train)
```

Code : Making Predictions

```
# Predict the labels for the test set
```

```
y_pred = clf.predict(X_test)
```

Code: Evaluating Model Performance

```
# Compute accuracy
```

```
from sklearn.metrics import accuracy_score
```

```
accuracy_score(y_test, y_pred)
```

OUTPUT:

Features and Labels

---

```
[[575  4]
 [ 37 51]]
```

CONCLUSION:

Data science customer analytics can provide companies with a competitive edge. Businesses may set themselves apart from their rivals and draw in and keep more customers by getting to know their clients better and being able to provide customized services and retention methods. This can help businesses establish a strong brand identity and maintain a loyal customer base.