Customer Churn Prediction using Machine Learning

June 17, 2023

-By Utkarsh Sinha

Step 1: Prototype Selection

Abstract

The objective of this report is to outline the process of developing a customer churn prediction system and exploring potential monetization strategies for transforming the idea into a profitable business. Customer churn prediction is a critical analysis for businesses aiming to retain existing customers, and it involves leveraging historical data and machine learning techniques to identify customers likely to discontinue their relationship with a company.

The report will delve into the various steps involved in building a customer churn prediction model, including data collection, preprocessing, feature engineering, model selection, training, validation, and performance evaluation. By following these steps, businesses can develop an accurate churn prediction system that assists in understanding and predicting customer attrition.

Furthermore, this report explores monetization strategies to transform the churn prediction idea into a viable business. It discusses potential avenues such as offering subscription-based services, providing consulting and advisory services, integrating with existing customer relationship management systems, delivering data analytics and insights, developing churn prevention solutions, and establishing partnerships and collaborations.

The success of the proposed business relies on validating the demand for churn prediction services, conducting market research, and refining the business model based on customer feedback and industry trends. Providing exceptional customer support,

showcasing the value of churn prediction solutions, and continuously innovating will be pivotal in building a profitable business in this domain.

Ultimately, this report aims to provide an informative overview of customer churn prediction, outlining the steps involved in developing a churn prediction model, and offering insights into potential monetization strategies for translating the idea into a successful business venture.

1. Problem Statement

"Businesses lack an efficient and accurate customer churn prediction system, hindering their ability to proactively identify and address customer attrition. This leads to financial losses, reduced market share, and increased customer acquisition costs."

Addressing this problem requires the development of robust churn prediction models that leverage historical data and advanced analytics techniques. Additionally, businesses need monetization strategies to transform the churn prediction idea into a profitable venture, enabling them to offer valuable solutions to organizations seeking to mitigate customer churn and enhance customer retention strategies.

2. Market/Customer/Business Need Assessment

There has been a huge downfall for Retail Shops and Vendors because of the unfortunate pandemic which has led many to buy things online instead of shops. So, it is extremely crucial for them to optimize their selling techniques by managing their inventory correctly by purchasing items which are in more demand and providing schemes on items which are generally grouped together to maximize their profits.

In addition, the epidemic has substantially altered client purchasing preferences. As a result of this technique, we hope to supply small businesses with beneficial data insights and revenue-generating opportunities.

3. Target Specifications and Characterization

1. Prediction Accuracy: The churn prediction system should aim for high accuracy in identifying customers who are likely to churn. The model should be trained and fine-tuned to achieve a balance between precision and recall, ensuring that both false positives and false negatives are minimized.
2. Real-time Analysis: The system should be designed to process and analyze customer data in real-time or near-real-time, allowing businesses to take proactive measures to prevent churn as soon as potential risks are identified.
3. Scalability and Efficiency: The system should be scalable to handle large volumes of customer data efficiently. It should be able to process and analyze data from diverse sources, such as transactional data, demographic information, customer interactions, and more.
4. Actionable Insights: The churn prediction system should provide actionable insights and recommendations to businesses, enabling them to develop effective retention strategies. These insights should be presented in a user-friendly and easily interpretable manner, empowering businesses to make informed decisions.
5. Monetization Potential: The chosen monetization strategies should be aligned with the needs and preferences of target customers. The system should offer flexible pricing models, such as subscription-based services, licensing, or revenue-sharing agreements, to cater to different business sizes and budgets.
6. Customization and Integration: The system should allow for customization to accommodate the unique requirements of different industries or business domains. It should also facilitate seamless integration

with existing customer relationship management (CRM) systems, maximizing the utility for businesses.

7. Data Security and Privacy: The system should prioritize data security and comply with relevant data protection regulations. It should implement robust security measures to safeguard customer data and ensure privacy throughout the data collection, storage, and analysis processes.
8. Continuous Improvement: The churn prediction system should have mechanisms in place to continuously update and improve the models based on feedback, evolving customer behavior, and industry trends. Regular evaluations and refinements should be performed to maintain and enhance prediction accuracy.

9.External Search(Information and Data Analysis)

These are some of the sources I visited for more information and need for shopping pattern analysis of customers.

1. **Why Market Basket Analysis is Crucial to Gain a Winning Edge in the Retail Sector**
2. **Advantages of Market Basket Analysis in B2B Marketing**
3. **What is Market Basket analysis?**

I am going to use this Dataset for my code implementation for this report.

Dataset Description:

The dataset used in these models contains customers bought items. Each row corresponds to the item bought by one customer in one invoice. We have to find which items to be added in a buy one gets one deal.

First import the basic libraries for data preprocessing:

In [1]:
```
# Importing the essential Libraries
import pandas as pd
import numpy as np
```

In [2]:
```
# Reading the Dataset
df = pd.read_csv('customer_churn.csv')
df.head()
```

Out[2]:

	RowNumber	CustomerId	Surname	CreditScore	Geography	Gender	Age	Tenure	Balance	NumOfProducts	HasCrCard	IsActiveMember	EstimatedSalary
0	1	15634602	Hargrave	619	France	Female	42	2	0.00	1	1	1	101348.88
1	2	15647311	Hill	608	Spain	Female	41	1	83807.86	1	0	1	112542.58
2	3	15619304	Onio	502	France	Female	42	8	159660.80	3	1	0	113931.57
3	4	15701354	Boni	699	France	Female	39	1	0.00	2	0	0	93826.63
4	5	15737888	Mitchell	850	Spain	Female	43	2	125510.82	1	1	1	79084.10

In [4]:
```
df.isnull().sum()
```

Let's now see more info on our dataset:

DATA PREPROCESSING

In [4]:
```
df.isnull().sum()
```

Out[4]:
```
RowNumber          0
CustomerId         0
Surname            0
CreditScore        0
Geography          0
Gender             0
Age                0
Tenure             0
Balance            0
NumOfProducts      0
HasCrCard          0
IsActiveMember     0
EstimatedSalary    0
Exited             0
dtype: int64
```

In [5]:
```
df.shape
```

Out[5]:
```
(10000, 14)
```

In [6]:
```
df.info()
```

In [6]:
```
df.info()
```

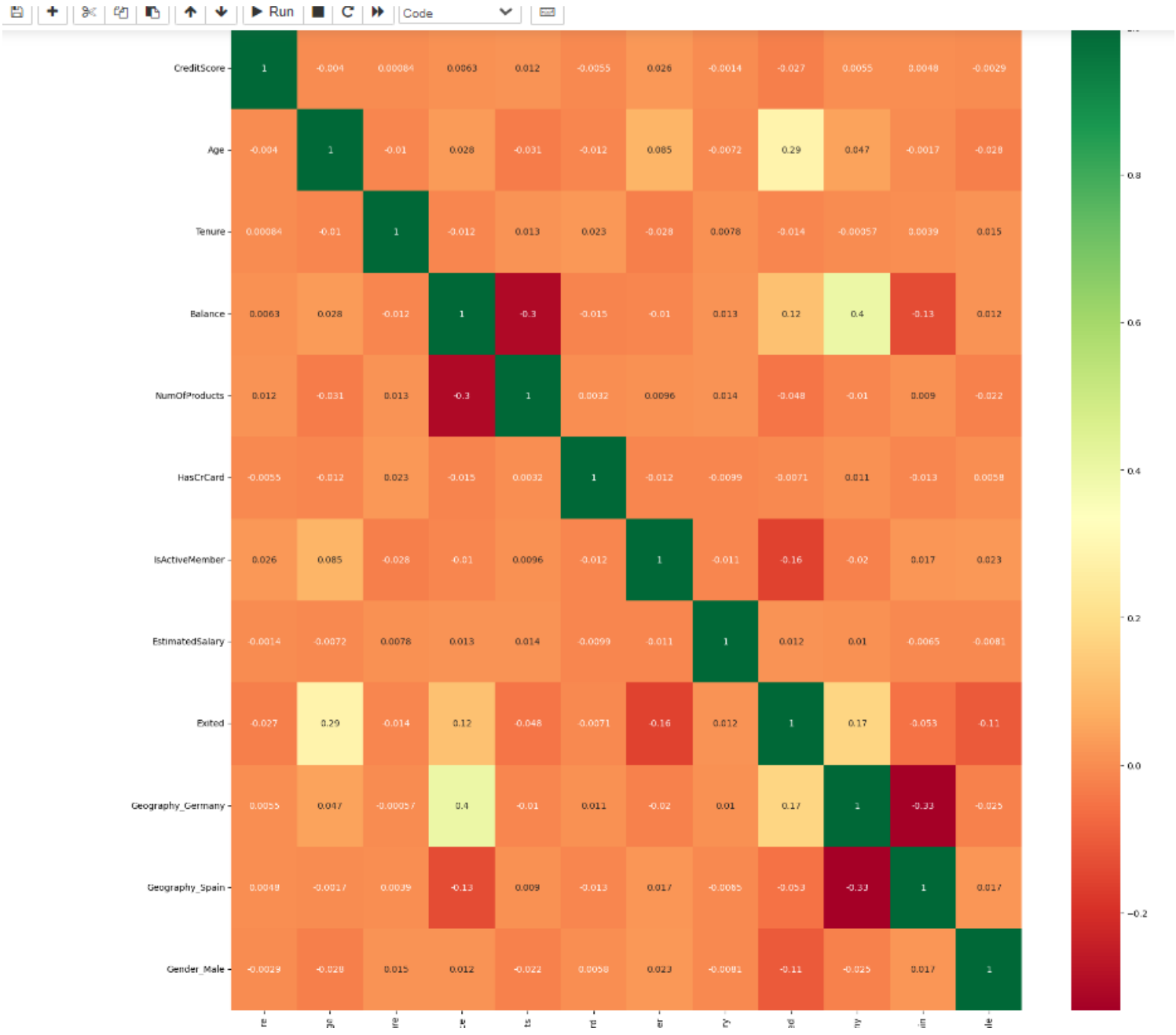
```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 14 columns):
 #   Column           Non-Null Count  Dtype  
---  ------           --------------  -----  
 0   RowNumber        10000 non-null  int64  
 1   CustomerId       10000 non-null  int64  
 2   Surname          10000 non-null  object 
 3   CreditScore      10000 non-null  int64  
 4   Geography        10000 non-null  object 
 5   Gender           10000 non-null  object 
 6   Age              10000 non-null  int64  
 7   Tenure           10000 non-null  int64  
 8   Balance          10000 non-null  float64
 9   NumOfProducts    10000 non-null  int64  
 10  HasCrCard        10000 non-null  int64  
 11  IsActiveMember   10000 non-null  int64  
 12  EstimatedSalary  10000 non-null  float64
 13  Exited           10000 non-null  int64  
dtypes: float64(2), int64(9), object(3)
memory usage: 1.1+ MB
```

In [7]:
```
df.columns
```

Out[7]:
```
Index(['RowNumber', 'CustomerId', 'Surname', 'CreditScore', 'Geography',
       'Gender', 'Age', 'Tenure', 'Balance', 'NumOfProducts', 'HasCrCard',
       'IsActiveMember', 'EstimatedSalary', 'Exited'],
      dtype='object')
```

In [10]:
```
print(df['Gender'].unique())
print(df['Geography'].unique())
```

Description and List of Transactions created for our Association Rule Mining Algorithm:



Splitting the dataset into Training and Testing dataset

In [24]:
```
# Splitting the dataset into Training and Testing Data
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 = 42)
```

In [25]:
```
# Standardizing the Dataset
from sklearn.preprocessing import StandardScaler
sc = StandardScaler()
X_train = sc.fit_transform(X_train)
X_test = sc.transform(X_test)
```

In [26]:
```
print(X_train)
```

```
[[ 0.35649971 -0.6557859   0.34567966 ... -0.57946723 -0.57638802
   0.91324755]
 [-0.20389777  0.29493847 -0.3483691  ...  1.72572313 -0.57638802
   0.91324755]
 [-0.96147213 -1.41636539 -0.69539349 ... -0.57946723  1.73494238
   0.91324755]
 ...
 [ 0.86500853 -0.08535128 -1.38944225 ... -0.57946723 -0.57638802
  -1.09499335]
 [ 0.15932282  0.3900109   1.03972843 ... -0.57946723 -0.57638802
   0.91324755]
 [ 0.47065475  1.15059039 -1.38944225 ...  1.72572313 -0.57638802
   0.91324755]]
```

In [27]:
```
print(y_train)
```

```
[0 0 1 ... 1 1 0]
```

In [28]:
```
## Feature Importance
```

10. Business Opportunity

By leveraging churn prediction, businesses can:

- Retain Customers: Customer retention is a critical aspect of any business's success. By accurately predicting customer churn, businesses can identify at-risk customers and proactively implement targeted retention strategies. This leads to increased customer loyalty, repeat business, and long-term revenue growth.

- Reduce Customer Acquisition Costs: Acquiring new customers is typically more expensive than retaining existing ones. By predicting customer churn, businesses can focus their resources on retaining customers who are at high risk of leaving, thereby optimizing their marketing and customer acquisition efforts. This results in cost savings and improved return on investment (ROI).

- Enhance Customer Satisfaction: Churn prediction allows businesses to gain a deeper understanding of customer behavior, preferences, and pain points. This knowledge enables companies to improve their products, services, and customer experiences, ultimately leading to higher levels of customer satisfaction and loyalty.

- Personalize Marketing and Communication: Churn prediction models can help businesses segment their customer base based on churn risk levels. This segmentation enables targeted marketing campaigns and personalized communication strategies. By delivering tailored offers, incentives, and engagement efforts, businesses can increase customer engagement and satisfaction, reducing the likelihood of churn.

- Optimize Resource Allocation: Churn prediction helps businesses allocate their resources more efficiently. By identifying customers who are unlikely to churn, companies can focus their marketing efforts on upselling, cross-selling, or nurturing relationships with high-value customers. This optimization of resources leads to increased revenue and improved operational efficiency.

- Gain Competitive Advantage: Implementing churn prediction as part of a business strategy gives companies a competitive edge. By being proactive and effectively managing customer churn, businesses can differentiate themselves in the market. This positions them as customer-centric organizations that prioritize customer satisfaction, leading to a stronger market position and improved brand reputation.

- Monetization Opportunities: Businesses that develop churn prediction capabilities can monetize their expertise in several ways. They can offer subscription-based services, consulting and advisory services, or integrate churn prediction into existing customer relationship management (CRM) systems. Additionally, there are opportunities to generate revenue through data analytics and insights, churn prevention solutions, or partnerships and collaborations.

Conclusion

Market basket analysis is being used by an increasing number of companies to acquire beneficial insights about associations and hidden relationships. However, for small businesses, this extension is a fantastic opportunity to boost sales and help them develop and grow their business.