International Journal of Science and Research (IJSR)
ISSN: 2319-7064
SJIF (2022): 7.942

Analyzing Credit Card Consumer Behavior using Unsupervised Machine Learning Techniques

Ritambhara Jha

Email: *jha.ritambhara[at]gmail.com*

Abstract: ***Credit cards are extensively utilized financial products that provide users with ease and flexibility. However, understanding and predicting credit card consumer behavior remains a complex challenge. Machine learning has evolved as an invaluable technique for analyzing massive information and extracting important insights, allowing businesses to better understand their consumers and design effective strategies. This paper analyzes the effective application of data science with ML models in the credit card consumer behavior. It goes over different data sources, machine learning algorithms, and the advantages of using data science.***

Keywords: Customer segmentation, Credit card, Unsupervised ML models

1. Introduction

Credit card is a type of electronic payment that is commonly used to pay for goods or services in installments. Transactions, client profiles, and other sources create vast volumes of data for banks or credit card firms. This information has enormous promise for better understanding customer behavior, recognizing purchasing trends, and forecasting future financial decisions. Data science provides a comprehensive framework for evaluating and deriving important insights from this data.

Segmentation is a vital aspect of developing marketing goals and strategies, and establishing those objectives will often involve a combination of:

a) A study of how products should be offered or created in light of present client segments [2]
b) Identifying new segments as targets for current items or developing new products [3].

As a company's resources are limited and it must focus on how to effectively identify and service its consumers. Strategic segmentation enables a firm to choose the consumer groups it should be targeted to serve and how to present its products and services for each group [3].

2. Related Work

Several research on consumer credit card usage and behavior have been conducted. In the United States of America, payment cards were created in 1930 to allow for payments at a merchant's own outlet, particularly in the "travel and entertainment" sector. Later, in the 1950s, Diners Club introduced the first general-purpose credit card. In 1958, American Express issued its first card, and the following year, Bank of America issued the BankAmericard in response to the bank's discovery that it was losing this market, with the cardholder's innovation being the ability to settle its debt until the deadline. The success was rapid, and the card quickly became the most popular among Americans. Other banks quickly joined the BankAmericard system, which acquired international traction. BankAmericard was renamed Visa in 1977[1].

Research work proposed in [4] utilizes the K-means algorithm to divide customers into groups based on their attributes. The authors of [5] propose a method for segmenting customers based on their CPB profile and multiple instance clustering. Customer segmentation is one of the many applications of unsupervised learning. Using clustering algorithms (K-means, Agglomerative, and Mean Shift), identify customer segements to focus on the potential user base. As a consequence, they divide clients into groups based on comparable characteristics such as gender, age, and hobbies and spending patterns.

3. Implementation

1) **Dataset -** It comprises the behavior of about 9000 active credit card holders during the last 6 months. The file is at a customer level with 18 behavioral variables. Data dictionary helps us understand what each data characteristic signifies. Pre-processing of data involves importing the necessary packages and datasets, reviewing the data summary, dealing with missing values, confirming data types, and selecting the features.

	CUST_ID	BALANCE	BALANCE_FREQUENCY	PURCHASES	ONEOFF_PURCHASES	INSTALLMENTS_PURCHASES	CASH_ADVANCE	PURCHASES_FREQUE
0	C10001	40.900749	0.818182	95.40	0.00	95.4	0.000000	0.16
1	C10002	3202.467416	0.909091	0.00	0.00	0.0	6442.945483	0.00
2	C10003	2495.148862	1.000000	773.17	773.17	0.0	0.000000	1.00
3	C10004	1666.670542	0.636364	1499.00	1499.00	0.0	205.788017	0.08
4	C10005	817.714335	1.000000	16.00	16.00	0.0	0.000000	0.08

Figure 1: Overview of the dataset

Observing the max values, it can be inferred that there are outliers. Dropping the outliers at this point can lead to loss of vital information. So, in this scenario outliers are treated as extreme values.

2) **Normalization** - It is a technique that is frequently used in data preparation for machine learning. The purpose of normalization is to convert the values of numeric columns in a dataset to a common scale while preserving disparities in value ranges. Every dataset does not require normalization for machine learning. It is only necessary when the ranges of characteristics differ.

3) **Principal Component Analysis** - It is impossible to showcase the data in 17 dimensions, PCA has been applied to convert it into two dimensions for visualization. PCA reduces a big number of variables to a smaller set that retains the majority of the information in the larger set, reducing the amount of data variables.

```
pca = PCA(n_components = 2)
X_principal = pca.fit_transform(normalized_df)
X_principal = pd.DataFrame(X_principal)
X_principal.columns = ['P1', 'P2']
X_principal.head(2)
```

	P1	P2
0	-0.489825	-0.679678
1	-0.518791	0.545011

Figure 2: Application of PCA

4) **Clustering:** It is a prominent exploratory data analysis tool for gaining an understanding of the layout of the data. It entails the method to find subgroups in data so that data-points in the same subgroup (cluster) are extremely similar while data points in other clusters are considerably dissimilar.

K-means Clustering algorithm is an iterative technique that attempts to partition a dataset into K unique non-overlapping subgroups (clusters), with each data point belonging to just one. It is done by specifying the number of clusters and the K value. The elbow approach is used to determine the optimal K. It is a relatively easy approach that produces plots in the shape of an elbow. We simply deduce the best number of K from the plot. The silhouette technique can compute the silhouette coefficient and quickly determine the exact value of K.

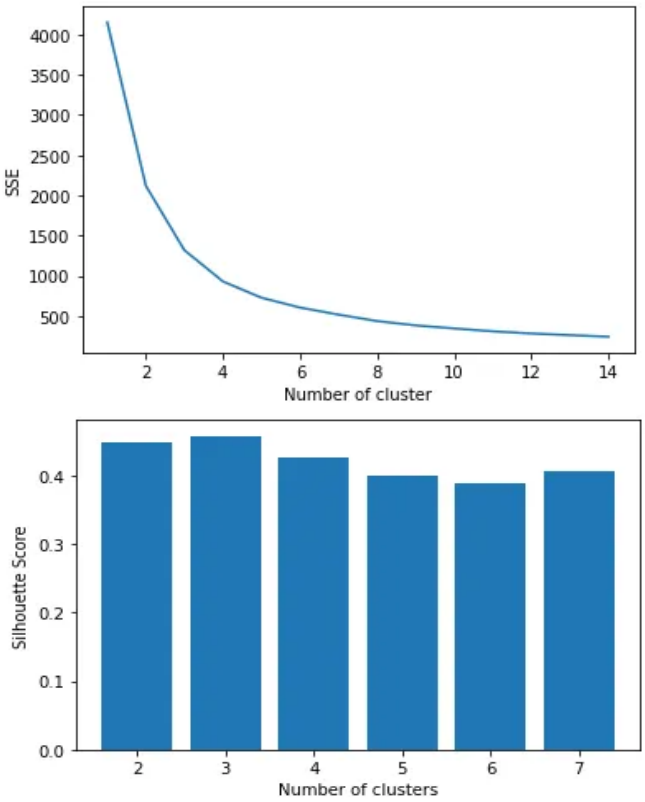

Figure 3: Elbow and Silhouette Approach

The silhouette coefficient has a value between -1 and 1. A score of 1 indicates that the data point is highly compact inside the cluster to which it belongs and is located far away from the other clusters. The poorest possible value is -1. Near-zero values indicate overlapping clusters.

k = 3 has the highest silhouette score. In this scenario, the ideal number of clusters is three.

4. Result

```
Cluster 0

Balance              : low
Balance Frequency    : high (updated frequently)
Purchase             : low
Purchase Frequency   : low
Cash Advance         : low
Minimum Payment      : low
Credit Limit         : low
```

Cluster 1

```
Balance              : medium
Balance Frequency    : high (updated frequently)
Purchase             : high
Purchase Frequency   : high (updated frequently)
Cash Advance         : low
Minimum Payment      : high
Credit Limit         : high
```

Cluster 2

```
Balance              : high
Balance Frequency    : high (updated frequently)
Purchase             : low
Purchase Frequency   : very low (does not updated frequently)
Cash Advance         : high
Minimum Payment      : high
Credit Limit         : high
```

Figure 4: Cluster Details

Cluster 0: This customer group represents a small set of consumers with low balances, low spenders (few purchases), and the lowest credit limit.

Cluster 1: This customer group represents a significant group of consumers with medium balances, heavy spenders (purchases), and the maximum credit limit.

Cluster 2: This customer category represents a limited number of consumers with large balances and cash advances, low purchasing frequency, and a high credit limit. This client sector, we can presume, utilizes their credit cards as a loan.

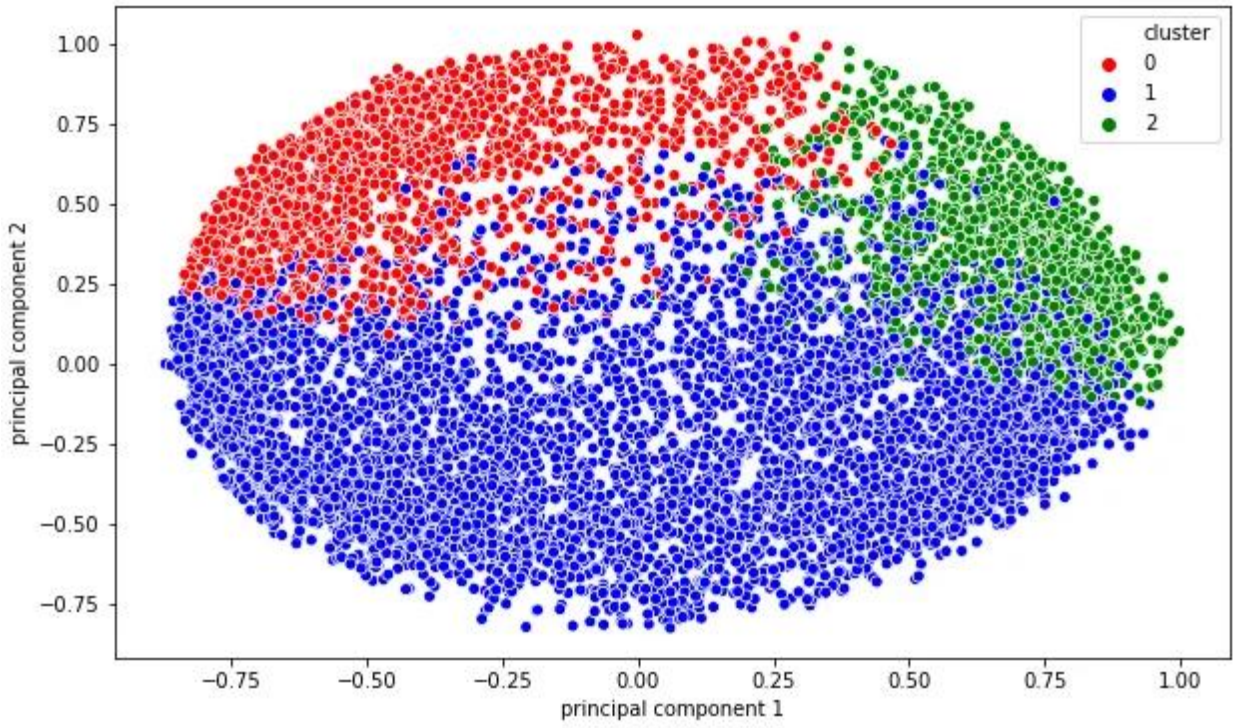

Figure 5: Clusters

5. Conclusion

Unsupervised machine learning (ML) modeling has developed as a valuable tool for credit card businesses to assess user behavior, forecast future events, and make sound judgments. Businesses may enhance consumer segmentation, detect fraud, anticipate credit risk, and establish more successful marketing campaigns by employing data science tools. However, it is crucial to address the challenges of data quality, model interpretability, privacy concerns, and algorithmic bias to ensure responsible and ethical use of data science in credit card consumer behavior analysis.

6. Future Work

Better Explainable AI (XAI) approaches are being developed to increase model interpretability and transparency. Exploring new data sources to acquire deeper insights into client behavior, such as alternative data from social media, wearable devices, and internet of things (IoT) sensors. Deep learning models' potential for increasingly complicated credit card customer behavior research tasks is being investigated. Considering the ethical implications of utilizing data science to predict credit card user behavior, such as fairness, transparency, and responsibility.

By tackling these research topics, data science can have an even larger role in influencing the future of credit card consumer behavior analysis and driving financial sector innovation.

References

[1] R Pereira, Sara Barradas. *Modelling credit card customer behaviour*. Diss. 2019
[2] Ansoff, H.I. (1957). Strategies for diversification. Harvard Business Review, Sept.-Oct.: 113–124
[3] McDonald, M. & Dunbar, I. (2004). Market segmentation: how to do it, how to profit from it. London: Elsevier
[4] Hemashree Kilari, Sailesh Edara, Guna Ratna Sai Yarra and Dileep Varma Gadhiraju, "Customer Segmentation using K-Means Clustering", *International Journal of Engineering Research & Technology (IJERT)*, vol. 11, no. 03, March 2022
[5] vett Fuentes et al., "Customer segmentation using multiple instance clustering and purchasing behaviors", *Progress in Artificial Intelligence and Pattern Recognition: 6th International Workshop IWAIPR 2018 Havana Cuba September 24–26 2018 Proceedings 6*, 2018
[6] S. Raj, S. Roy, S. Jana, S. Roy, T. Goto and S. Sen, "Customer Segmentation Using Credit Card Data Analysis," *2023 IEEE/ACIS 21st International Conference on Software Engineering Research, Management and Applications (SERA)*, Orlando, FL, USA, 2023, pp. 383-388, doi: 10.1109/SERA57763.2023.10197704.