IJARETY

International Journal of Advanced Research in Education and TechnologY (IJARETY)

Volume 11, Issue 4, July-August 2024

Impact Factor: 7.394

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

Data Analytics and Visualization using Tableau for Churn in the E-Commerce Sector

Mohammed Hassan Farukh Faizan[1], Mohammed Yousuf Hussain[1], Syed Miran Shah Quadri[1], Dr T Praveen Kumar[2]

UG Student, Department of Artificial Intelligence and Data Science, Methodist College of Engineering and Technology, Hyderabad, India[1]

Associate Professor, Department of Computer Science and Engineering, Methodist College of Engineering and Technology, Hyderabad, India[2]

ABSTRACT: With the increasing popularity and rapid development of e-commerce, competition among companies has intensified[1]. Retaining customers through excellent service and competitive pricing is crucial. To meet high customer demand, companies must attract customers with customized services and targeted strategies to increase loyalty. E-commerce customer churn data shows nonlinear changes and imbalanced samples, where the number of churned customers is significantly different from non-churned ones. This research proposes methods and models for e-commerce companies to proactively reduce customer churn.

I. INTRODUCTION

1.1 Background on Customer Churn in the E-commerce Sector

In E-commerce, it turns out it is much more expensive for a company to attract new customers than to retain current ones[1]. For this specific reason, knowing in advance which customers will leave the company will enable the businesses to create offers or reduce the consumption of its products or services in a relevant way is crucial to increase their retention, build a good Customer Relationship and save acquisition costs. In today's competitive market, there is an immense variety of products and services to choose from. Accordingly, most consumers got used to walking freely from one brand to another, from one supplier to another, looking for the product or service that suits their needs. E-commerce Companies have been suffering from this phenomenon, known as customer “churn,”. Therefore, instead of focusing on retaining their current customers, they often invest effort and allocate huge amounts of money in attracting new customers[2].

1.2 Purpose and Objectives of the Study

The primary purpose of this study is to **analyze customer churn** in the e-commerce sector using data analytics and visualization techniques. The objectives are to:

- Identify key factors influencing customer churn.
- Analyze patterns and trends related to churn.
- Utilize Tableau for effective data visualization to derive actionable insights.

1.3 Overview of Data Visualization

Data visualization is the representation of data or information in a graph, chart, or other visual format. It communicates relationships of the data with images. This is important because it allows trends and patterns to be more easily seen. With the rise of big data upon us, we need to be able to interpret increasingly larger batches of data. But data visualization is not only important for data scientists and data analysts, it is necessary to understand data visualization in any career. Whether you work in finance, marketing, tech, design, or anything else, we need to visualize data. Its main goal is to distill large datasets into visual graphics to allow for easy understanding of complex relationships within the data. It is often used interchangeably with terms such as information graphics, statistical graphics, and information visualization. It is precisely because of this interdisciplinary nature that the visualization field is full of vitality and opportunities[11].

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

II. LITERATURE REVIEW

2.1 Establishment of E-Commerce Customer Churn Prediction Model

Customer churn refers to the fact that the original customers of an enterprise stop to purchase enterprise goods or accept enterprise services, and instead accept the services of competitors[3]. In a non-contractual relationship, even if the termination of this kind of business-customer relationship occurs, it is difficult for the business to detect it in advance [4]. The value of e-commerce customer churn prediction is to merge ecommerce customer data over some time and establish e-commerce customer churn prediction models by analyzing customer purchase behaviors [5]. Then, provide e-commerce customer churn retention measures to reduce customer churn and identify high-value non-churn e-commerce customers and do a respectable job in customer retention. The remaining customers do not need high costs as the new customers want to bring high profit in ecommerce. Comparatively, the customer purchase behavior differs for both existing and new customers; however, it is essential to identify the reasons leading to the customer's loss[4]. In support,[6] stated that in the e-commerce sector, it is extremely important to analyze the loss of customers, predict the customers who might be lost, and then take corresponding measures to retain these customers and avoid their loss. At present, most e-commerce companies have conducted an in-depth analysis of customer basic characteristic information and transaction behavior data, and then use various methods and technologies to establish and study customer churn prediction models, and finally use this to predict customer churn[7].

2.2 Customer Segmentation

Customer segmentation is used for the recognition of the value of customer relationship, a key step prerequisite for more efficient targeted marketing activities [8]. According to the famous Pareto principle, 80% of a company's profits are created by 20% of its customers, and 50% of its profits are lost by the bottom 30% of non-profit customers[9]. Therefore To perform customer segmentation, we must first identify customer value. Focusing on customer value and allocating resources for targeted marketing can enhance a company's core competitiveness. With the rapid development of e-commerce, corporate business activities carried out through the internet are more real-time and interactive, which also transforms the product-centric business type into a customer-centric model for ecommerce business [9]. The churn rate of e-commerce customers is relatively high. If companies want to establish long-term alliances with customers and develop stable and continuous customer relationships[10]. The e-commerce customer base is extensive and complex, and its value varies. The study[10] raised a question about how to accurately identify high-value customers, predict churn and retain them in advance has become a hot spot in the field of e-commerce. Evaluating customer value helps companies identify their most valuable customers and allocate resources effectively to maximize impact.

III. DATA COLLECTION

This project uses a Kaggle dataset on customer consumption behavior on an e-commerce website from June to November 2021. It analyzes consumption patterns and purchase history, along with past platform interactions, to provide insights into customer behavior.

Visualization Tool: Tableau for creating visualizations.

IV. DATA DESCRIPTION

4.1 Dataset Overview

Data is collected from a leading E-commerce platform, it is a historical data containing customer details and experience and its outcome is customer churn flag (churn= 1, no churn = 0). The dataset shows more than 5000 customers and their interaction and preferences in the platform. The effectiveness of this data is that it contains some specific detailed attributes which will help in customer segmentation such as: preferred login device, Satisfaction store and other attributes. These attributes will help us in studying the causes of churn in each customer's segment to identify the triggers leading to customer churn[11].

Attribute	Type
CustomerID	Numeric
Churn	Numeric
Tenure	Numeric
Preferred Login Device	Character
City Tier	Numeric
WarehouseToHome	Numeric
PreferredPaymentMode	Character
Gender	Character

International Journal of Advanced Research in Education and TechnologY(IJARETY)

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

HoursSpentOnApp	Numeric
NumberOfDevicesRegistered	Numeric
PreferedOrderCat	Character
SatisfactionScore	Numeric
MaritalStatus	Character
NumberOfAddress	Numeric
Complain	Numeric
OrderAmountHikeFromLastYear	Numeric
CouponUsed	Numeric
OrderCount	Numeric
DaySinceLastOrder	Numeric
CashbackAmount	Numeric

Table 1: Data Dictionary

4.2 Sample Data

CustomerID	Churn	Tenure	PreferredLoginDevice	CityTier	WarehouseToHome	PreferredPaymentMode	Gender	urSpendOnA	berOfDeviceRegist	referedOrderCa	sfactionSc	aritalStatu	berOfAdd	Complain	untHikeFr	ouponUse	rderCoun	inceLastO	hbackAm
50001	1	4	Phone	3	6	Debit Card	Female	3	3	Laptop & Acces	2	Single	9	1	11	1	1	5	159.93
50002	1	0	Phone	1	8	UPI	Male	3	4	Mobile	3	Single	7	1	15	0	1	0	120.9
50003	1	0	Phone	1	30	Debit Card	Male	2	4	Mobile	3	Single	6	1	14	0	1	3	120.28
50004	1	0	Phone	3	15	Debit Card	Male	2	4	Laptop & Acces	5	Single	8	0	23	0	1	3	134.07
50005	1	0	Phone	1	12	CC	Male	0	3	Mobile	5	Single	3	0	11	1	1	3	129.6
50006	1	0	Computer	1	22	Debit Card	Female	3	5	Mobile Phone	5	Single	2	1	22	4	6	7	139.19
50007	1	0	Phone	3	11	COD	Male	2	3	Laptop & Acces	2	Divorced	4	0	14	0	1	0	120.86
50008	1	0	Phone	1	6	CC	Male	3	3	Mobile	2	Divorced	3	1	16	2	2	0	122.93
50009	1	13	Phone	3	9	E wallet	Male	0	4	Mobile	3	Divorced	2	1	14	0	1	2	126.83
50010	1	0	Phone	1	31	Debit Card	Male	2	5	Mobile	3	Single	2	0	12	1	1	1	122.93
50011	1	4	Phone	1	18	COD	Female	2	3	Others	3	Divorced	2	0	0	9	15	8	295.45
50012	1	11	Phone	1	6	Debit Card	Male	3	4	Fashion	3	Single	10	1	13	0	1	0	153.81
50013	1	0	Phone	1	11	COD	Male	2	3	Mobile	3	Single	2	1	13	2	2	2	134.41
50014	1	0	Phone	1	15	CC	Male	3	4	Mobile	3	Divorced	1	1	17	0	1	0	133.88
50015	1	9	Phone	3	15	CC	Male	3	4	Fashion	2	Single	2	0	16	0	4	7	196.19
50016	1	0	Phone	2	12	UPI	Male	3	3	Mobile	5	Married	5	1	22	1	1	2	120.73
50017	1	0	Computer	1	12	Debit Card	Female	0	4	Mobile	2	Single	2	1	18	1	1	0	129.26
50018	1	0	Phone	3	11	E wallet	Male	2	4	Laptop & Acces	3	Single	2	1	11	1	1	3	157.44
50019	1	0	Computer	1	13	Debit Card	Male	3	5	Laptop & Acces	3	Single	2	1	24	1	1	6	160.74
50020	1	19	Phone	1	20	Debit Card	Female	3	3	Mobile Phone	4	Divorced	10	1	18	1	4	3	149.63
50021	1	0	Phone	3	12	Debit Card	Male	3	5	Laptop & Acces	3	Divorced	5	1	18	6	7	7	161.72
50022	1	20	Phone	1	29	CC	Female	3	3	Fashion	2	Divorced	2	0	12	11	15	6	203.12
50023	1	0	Phone	3	28	E wallet	Male	2	3	Mobile Phone	3	Single	2	1	19	0	1	0	116.75
50024	1	0	Phone	3	26	Debit Card	Female	3	5	Laptop & Acces	3	Divorced	4	1	11	1	2	2	145.59

V. DATA PREPROCESSING

5.1 Handling Missing Values

IFNULL in Tableau is the Logical Function which is used to change a Null value to some other values. It can convert null values to another value and also convert nulls to other values from another field. IFNULL in tableau can convert any data-type nulls values like Null dates, text and numbers and so on. The IFNULL function also accepts aggregate and non-aggregate values.

`IFNULL([Field_name],[Field_output_sameDataType])` can be used on any data type to replace null values to zero or any other values. In our dataset, the Tenure column contains null values(Figure 1). To retain as much data as possible for further analysis and avoid filtering out potentially useful information, we've replaced these null values with zeros using the IFNULL() function. Figure 2 shows the Tenure column after applying this transformation.

CustomerID	Churn	Tenure	PreferredLoginDevice	C
50001	1	4	Mobile Phone	
50002	1		Phone	
50003	1		Phone	
50004	1	0	Phone	
50005	1	0	Phone	
50006	1	0	Computer	
50007	1		Phone	
50008	1		Phone	
50009	1	13	Phone	
50010	1		Phone	

Figure 1: Before filling the null values

CustomerID	Churn	Tenure	PreferredLoginDevice
50001	1	4	Phone
50002	1	0	Phone
50003	1	0	Phone
50004	1	0	Phone
50005	1	0	Phone
50006	1	0	Computer
50007	1	0	Phone
50008	1	0	Phone
50009	1	13	Phone

Figure 2: After filling the null values

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

5.2 Data Transformation

Convert categorical variables into numerical formats if necessary. In our dataset, we have multiple names that refer to the same entity. For example, in the column `PreferredLoginDevice` we have `MobilePhone` and `Phone` that refer to the same entity `Phone` (Figure 3). Similarly, in the `PreferredPaymentMode`, we have `CC` and `COD` as `Credit Card` and `Cash On Delivery`(Figure 4) respectively.

n	Tenure	PreferredLoginDevice	CityTier
1	4	Mobile Phone	3
1		Phone	1
1		Phone	1
1	0	Phone	3
1	0	Phone	1
1	0	Computer	1
1		Phone	3
1		Phone	1
1	13	Phone	3
1		Phone	1
1	4	Mobile Phone	1
1	11	Mobile Phone	1

Figure 3: Duplicate devices names

ne	PreferredPaymentMode	Gender
6	Debit Card	Female
8	UPI	Male
30	Debit Card	Male
15	Debit Card	Male
12	CC	Male
22	Debit Card	Female
11	Cash on Delivery	Male
6	CC	Male
9	E wallet	Male
31	Debit Card	Male
18	Cash on Delivery	Female
6	Debit Card	Male
11	COD	Male
15	CC	Male

Figure 4: Consolidated duplicates

To fix this, we simply replaced the `Mobile Phone` with `Phone` in the `PreferredLoginDevice` and replaced both `Cash On Delivery` and `Credit Card` with `COD` and `CC` respectively.(figure 5)

PreferredLoginDevice	CityTier	WarehouseToHome	PreferredPaymentMode
Phone	3	6	Debit Card
Phone	1	8	UPI
Phone	1	30	Debit Card
Phone	3	15	Debit Card
Phone	1	12	CC
Computer	1	22	Debit Card
Phone	3	11	COD
Phone	1	6	CC
Phone	3	9	E wallet
Phone	1	31	Debit Card
Phone	1	18	COD
Phone	1	6	Debit Card
Phone	1	11	COD
Phone	1	15	CC

Figure 5: Transformed values in both the columns

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

VI. DATA VISUALIZATIONS

6.1 Initial Visualizations to understand the data

In the histogram charts below, we illustrated all the possible numerical attributes in our dataset to study their distribution, further description is shown below.

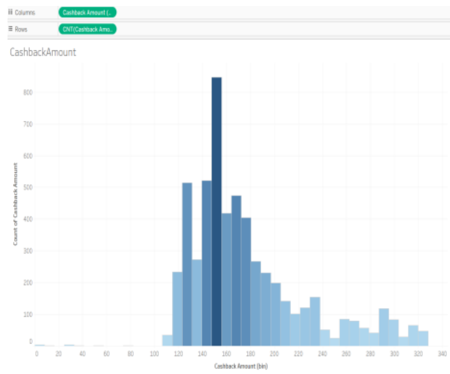

Figure 6: Cashback amount

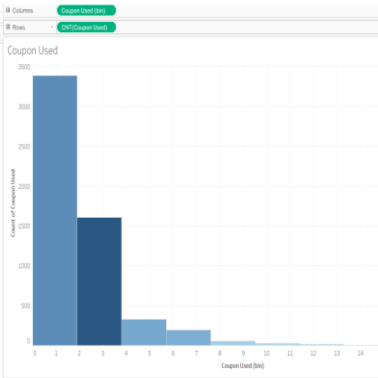

Figure 7: Coupon used

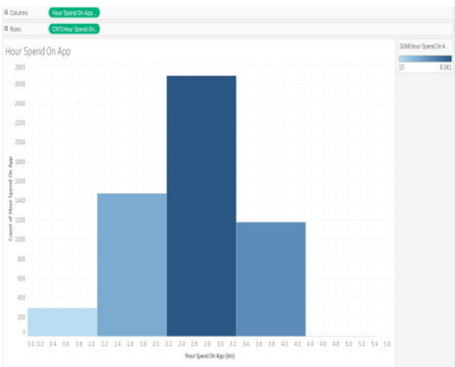

Figure 8: Hour spend on app

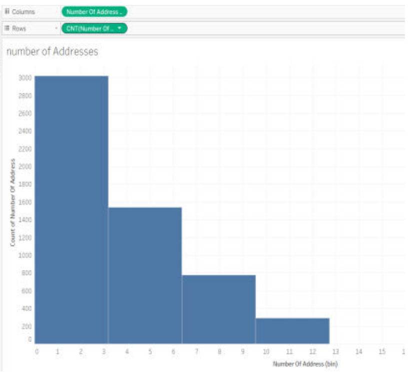

Figure 9: Number of address

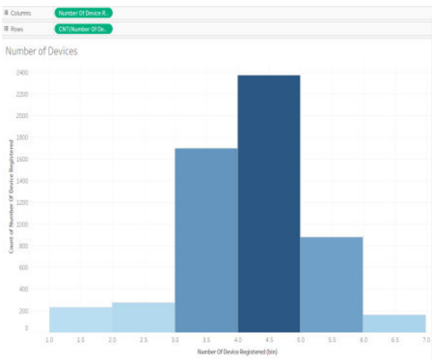

Figure 10: Number of devices

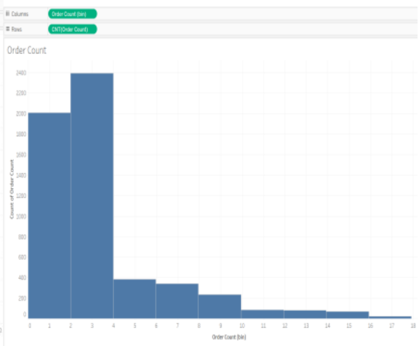

Figure 11: Order count

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

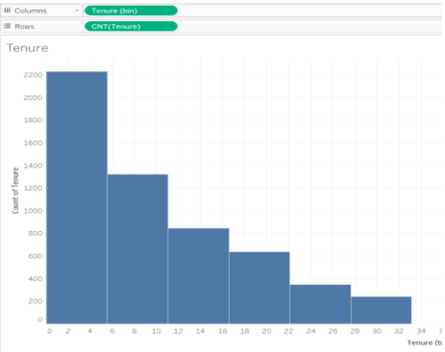

Figure 12: Satisfaction score

Figure 13: Tenure

6.2 Advanced Visualizations and Analysis

6.2.1 Churn by Preferred Login Device:

This bar chart (figure 15) shows the distribution of churned vs. non-churned customers based on their preferred login device (e.g., Phone, Computer, Tablet).

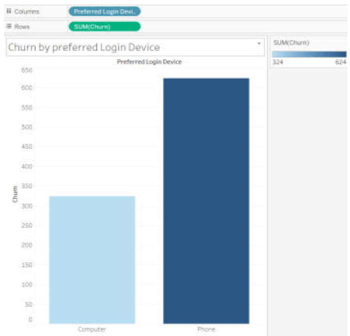

Figure 14: Churn by preferred login device

Insight: If a particular login device shows a higher churn rate, it might indicate usability issues or a lack of engagement on that device.

6.2.2 Churn by Preferred Payment Mode

The bar chart(figure 16) shows counts of churned vs. non-churned for each preferred payment.

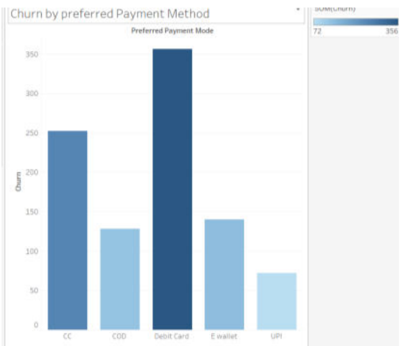

Figure 15: Churn by Preferred Payment Method

Insight: Identifying a payment mode with a higher churn rate could suggest a need to improve the payment process for that method or offer more incentives for its use.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

6.2.3 Churn by Gender

The bar chart (figure 17) compares the count of churned vs. non-churned customers across different genders.

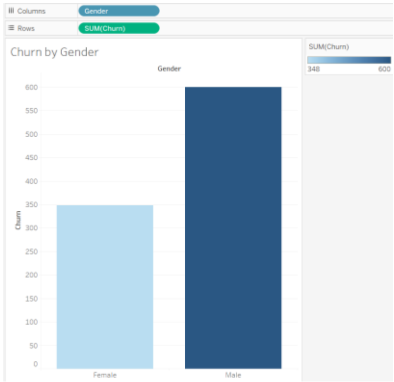

Figure 16: Churn by Gender

Insight: If one gender has a higher churn rate, targeted marketing strategies can be developed to retain that specific group.

VII. KEY FINDINGS AND ACTIONABLE INSIGHTS

7.1 Observation from Categorical Analysis

The analysis revealed that mobile phone users exhibit a higher churn rate compared to computer users, indicating potential issues with the mobile app experience. Additionally, customers who prefer certain payment methods, such as UPI, have a higher churn rate, suggesting payment-related concerns. There is also a noticeable difference in churn rates between male and female customers, highlighting the need for gender-specific strategies.

VIII. RESULTS AND CONCLUSIONS

The churn analysis on the e-commerce dataset revealed several critical insights. Mobile phone users exhibited a higher churn rate compared to computer users, suggesting the need to enhance the mobile app's usability and features to reduce churn among mobile users. Additionally, payment methods such as UPI were associated with higher churn rates, indicating the necessity to streamline and improve these payment processes for better customer satisfaction. Significant differences in churn rates were also observed between male and female customers, highlighting the importance of developing gender-specific marketing and retention strategies. These findings underscore the need for targeted interventions to address user experience issues, simplify payment processes, and cater to the unique preferences of different customer segments. Implementing these actionable insights can improve customer retention, ultimately enhancing the overall performance of e-commerce platforms.

REFERENCES

[1] Saghir, M., Bibi, Z., Bashir, S., & Khan, F. H. (2019, January). Churn prediction using neural network-based individual and ensemble models. In 2019 16th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 634-639). IEEE.

[2] Alshamsi, Abdulrahman, "Customer Churn Prediction in the ECommerce Sector" (2022). Thesis. Rochester Institute of Technology.

[3] Wu, X. J., & Meng, S. S. (2017). Research on e-commerce customer churn prediction based on customer segmentation and Ada-Boost. Industrial Engineering, 20(02), 99 107.

[4] Shao, D. (2016). Analysis and prediction of insurance company's customer loss based on BP neural network. Lanzhou University

[5] Zhang, D. (2015). Establishment and application of customer churn prediction model. Beijing Institute of Technology.

[6] Lu, N., Liu, X. W., & Lee, L. (2018). Research on customer value segmentation of online shops based on RFM. Computer Knowledge and Technology, 14(18), 275-276, 284.

| ISSN: 2394-2975 | www.ijarety.in| | Impact Factor: 7.394| A Bi-Monthly, Double-Blind Peer Reviewed & Referred Journal |

|| Volume 11, Issue 4, July-August 2024 ||

DOI:10.15680/IJARETY.2024.1104017

[7] Huang, J. (2018). A Comparative Study of Social E-Commerce and Traditional E-commerce. Economic and Trade Practice, (23), 188-189.
[8] Feng, X., Wang, C., Liu, Y., Yang, Y., & An, H. G. (2018). Research on customer churn prediction based on comment emotional tendency and neural network. Journal of China Academy of Electronics Science, 13(03), 340-345
[9] Sun, J., Li, H., Fujita, H., Fu, B., & Ai, W. (2020). Class-imbalanced dynamic financial distress prediction based on Adaboost-SVM ensemble combined with SMOTE and time weighting. Information Fusion, 54, 128-144.
[10] Agrawal, S., Das, A., Gaikwad, A., & Dhage, S. (2018, July). Customer churn prediction modeling based on behavioral patterns analysis using deep learning. In 2018 International conference on smart computing and electronic enterprise (ICSCEE) (pp. 1-6). IEEE.
[11] Nikhat A, Nazia T, Dr. Asif P and Dr. Yusuf P. (2020). Data analytics and visualization using Tableau utilitarian for COVID-19 (Coronavirus). Global Journal of Engineering and Technology Advances, 3(2), 28-50.

IJARETY

INNO SPACE
SJIF Scientific Journal Impact Factor

International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975 Impact Factor: 7.394

www.ijarety.in editor.ijarety@gmail.com