Segmentation study on Bank customers based on RNN

Vaibhav Jaitwal
Department of CSE
Apex Institute of Technology
Chandigarh University, Punjab, India
vaibhavjaitwal2003@gmail.com

Dr. Priyanka Kaushik
IEEE Member, Department of CSE
Apex Institute of Technology
Chandigarh University, Punjab, India
kaushik.priyanka07@gmail.com

Ishpreet Kaur
Department of CSE
Apex Institute of Technology
Chandigarh University, Punjab, India
21BCS6424@cuchd.in

Abstract— **Effective Customer Relationship Management (CRM) stands as a vital determinant of a company's future growth in a time marked by technology advancements and the changing dynamics between enterprises and consumers. This study concentrates on customer segmentation as a crucial tactic for enhancing business-customer relations. Customer segmentation is the process of strategically dividing a customer database into different subsets based on traits and behaviors that are common to each group. The study investigates the use of the K-means clustering method to divide customers into six different clusters according to annual income, expenditure scores, and two primary components. Successful clustering outcomes enable businesses to make data-driven decisions that improve their product and service offerings and enable quick adjustments to changing client needs. This research project's main objective is to show how effective RNN-based customer segmentation is by using mean values as a guiding signal for assigning new clients to pertinent clusters. The ability to use RNNs for accurate consumer segmentation is essential for sustained growth and keeping a distinct competitive edge in a market that is intensely competitive and changing quickly.**

Keywords— RNN, ANN, CRM, KNN,K-means.

I. Introduction

Understanding consumer behavior and preferences is crucial for firms looking to succeed in today's dynamic and fiercely competitive business environment. client segmentation, or the process of breaking a client base into discrete groups with comparable characteristics, has become an essential tactic for customizing goods, services, and marketing initiatives to suit the particular requirements and preferences of diverse consumer segments. Although conventional segmentation strategies have been effective, the exponential rise of data and developments in machine learning techniques present new prospects for improving the precision and granularity of customer segmentation.

The power of clustering and recurrent neural networks (RNNs) are combined in this research study to create an original method for client segmentation. To categorize customers based on static characteristics like demographics, past purchases, or location, clustering techniques have been used for a long time. Although this can offer deeper insights on shifting preferences and engagement patterns, these methods frequently ignore the time aspect of client behavior. However, RNNs, a group of neural networks intended to process sequential data, are particularly good at capturing temporal connections and can spot hidden patterns in time-series consumer data.

By combining clustering methods with RNNs, our suggested methodology closes this gap and enables the identification of both static and dynamic client categories. This hybrid strategy uses RNNs to predict the temporal evolution of these segments after first using clustering to identify the initial customer groupings. By doing this, we hope to give businesses a more thorough and useful insight of their clientele.

The main basic libraries which are used to create digit detection model are as follow:

Numpy: Numpy libraries is specifically used to perform operation on multi-dimensional arrays, and it is also used for processing and manipulating data such as loading CreditCard dataset in the memory with transforming it to situable dimensions.

scikit-learn: scikit-learn is a python library which have inbuilt algorithms for regression, classification and model selection. It includes build-in support for loading CreditCard dataset.

Tensorflow: TensorFlow is an open-source deep learning library which provides support for training and testing wide range of model. It includes build-in support for loading CreditCard dataset.

Keras: Keras is built on top of TensorFlow, which is a neural network deep learning library and it also furnish a interface(high-level) for constructing deep neural networks and training it too, making it easy to get started with digit classification using the CreditCard dataset.

Matplotlib: Matplotlib is a library used for creating visualizations, including graphs, charts, and plots. In the contextes of digit classification using the CreditCard dataset, Matplotlib can be used to visualize the images of the digits and the results of the classification models.

K-Nearest Neighbor (KNN): KNN is a ML algorithm which perform problems such as classification problems. Mainly KNN finds k-nearest neighbors for every test instance present in training data. KNN can be trained by various distance metrics such as Euclidean distance ($d = \sqrt{((x_2 - x_1)^2 + (y_2 - y_1)^2}$) and Manhattan distance ($d = |x_2 - x_1| + |y_2 - y_1|$).

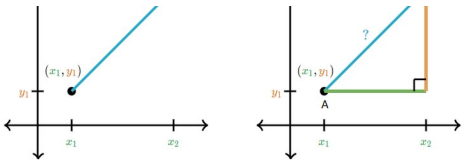

Figure-1: Euclidean Distance, Manhattan Distance

Artificial Neural Network(ANN): It provides deep learning algorithm which can be used for Customer segmentation. ANN consists of multiple layers of interconnected nodes, or neurons, that can learn complex representations of the input data. ANN can be trained using back-propagation and gradient descent algorithms.

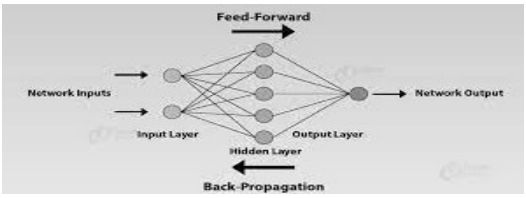

Figure-2: ANN multiple layers of interconnected nodes

Recurrent Neural Network(RNN): A recurrent neural network (RNN) is a type of artificial neural network that is designed to process sequential data, such as time series or natural language. RNNs have feedback connections that allow them to retain information from previous time steps, enabling them to capture temporal dependencies. This makes RNNs well-suited for tasks like language modeling, speech recognition, and sequential data analysis.

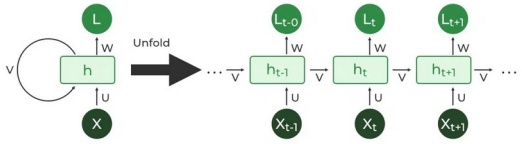

Figure-3 : Architecture of RNN

II. LITERATURE SURVEY

Recurrent neural networks (RNN) have been a cutting-edge method for customer segmentation in recent years. This review of the literature concentrates on 15 publications that offer a thorough investigation of machine learning and deep learning methods for consumer segmentation, with a concentration on the application of RNNs. The study includes RNNs as well as more conventional algorithms like K-Nearest Neighbors and more recent deep learning model like Convolutional Neural Networks etc.

The Credit Card database, which contains a wide number of Customers features , has the main task of evaluating the various algorithm's performance.In a paper by [1] it tell us about the practical approach in customers segmentation by using K-means alogorithmn.

In another paper [2][3] proposed a segmentation model which supported the use of automated consumer segmentation approaches, which are more effective than traditional market analytics, which frequently fall short when the client base is much larger.They have used KNN and K-means algorithmn for clusterning of the data.

Study on Artificial neural network is proposed by [4] [5],they used multilevel perceptron (MLP) to train and test the customer data .Within a few epochs, the system successfully generalized our consumer segmentation technique and, as a result, attained a good level of overall accuracy.Also they have used Recence frequency monetary (RFM) and User Event Tracking (UET) suing K-means algorithm.

Study on RFM framework using the hybrid model is proposed in [6][7], a hybrid model for customer segmentation based on RFM framework.Using MLP for prediction, clients are divided into eight categories. The accuracy rate can be greatly increased using this innovative technique. The RFM model, which is the foundation of this class label, can effectively identify customers' worth through supervised learning and offers solid support for the customer segmentation strategy of e-commerce platforms. Using actual customer consumption statistics from a real-world dataset, we test the effectiveness of the suggested approach. The experimental findings show how considerably better than baseline models our model performs. In comparison to GBDT and MLP, the accuracy of the proposed in this work increased by roughly 20.4% and 3.8%, respectively.

Customer segmentation and customer churn analysis system for insurance companies in [8].The goal of customer churn analysis is to foresee current customers' propensity to leave in advance. In the context of the study, a program is created that presents each phase of turning raw data into knowledge as a modular process. Real-world issues are used to test software, and good outcomes are attained.

Estimating Customer Segmentation based on Customer Lifetime Value Using Two-Stage Clustering Method is proposed in [9] ,With the use of the clustering method, this study tries to identify client segments with similar lifetime values so that businesses can tailor their tactics to the correct ones. This study suggests a two-stage clustering strategy for client segmentation. The initial number of clusters are selected using the Ward's approach, and clustering analysis is carried out using the K-Means method. To find the most effective method for consumer segmentation, two ways utilizing the LRFM (Length, Recency, Frequency, Monetary) model and expanded model known as LRFM - Average Item (AI) variables in clustering process are compared by validity index. The outcome demonstrates that adding the new variable Average Item to the LRFM model did not significantly alter the clustering or produce better outcomes.

Comparission between various machine and deeep learning algorithm like K-means, KNN, RNN on customer segmentation proposed in [10],this paper illustrates client segmentation using machine learning. They used four of the most well-liked clustering techniques to solve the unsupervised clustering problem: K-Means, Affinity Propagation, DBSCAN, and Hierarchical Clustering. Using the Elbow approach and Silhouette score together with important variables like client age, expenditure score, and annual income, they determine the ideal number of clusters from a dataset from a shopping mall. By identifying client behavior patterns, it aids retail centers in growing their business.

E-Commerce Market Segmentation Based On The Antecedents Of Customer Satisfaction and Customer Retention is propossed in [11],This study aims to establish e-commerce consumer segmentation based on psychographic and demographic parameters, as well as to ascertain the impact of psychographic elements on customer satisfaction and retention in e-commerce. In this associative-descriptive study, 411 respondents completed a questionnaire, which was used to collect data. Path analysis, cluster analysis, and cross-tabulation were utilized for analysis. According to the research's findings, each independent aspect influences customer satisfaction, which leads to the identification of three distinct e-commerce customer segments: the functional shopper, the credibility-matters shopper, and the money-diet shopper.

This study in [12] proposes a new segmenation method based on customer reviews existing on the web.Given the variety of client wants, it is challenging to precisely determine those needs through market segmentation utilizing demographic data. As a result, it is crucial in marketing to divide the consumer base according to the advantages that they gain from using a product or service. Given that it accurately recognizes training data even in the presence of noise and outliers and is frequently used for text data analysis, the random forest algorithm is employed in this study for benefit segmentation.

The study in [13] proposes a deep learning model in which Melons' diseases may be automatically predicted and classified using a deep learning model based on stacked RNNs.The Stacked RNN is used to categorize melons' illnesses. When compared to other existing models, this prediction and categorization of leaf diseases performs at the highest level with great accuracy and computational efficiency. The TensorFlow and Keras framework is used to implement this disease prediction and categorization. Regarding the accompanying measurements of exactness, review, and f-measure, the performance is estimated. Adam is the optimization method while ReLu and SoftMax are the activators utilized in the classification of diseases.

The study in [14] proposes a prediction model based on LSTM and RNN model.The properties of dynamic nonlinearity and correlation of water quality parameter information, as well as gradient explosion and gradient disappearance due to training data of conventional RNN network model, are all examined in this research. To improve the RNN network's structure as well as the connection weight and threshold of the hidden layer, the long short-term memory network structure (LSTM) is introduced. By adjusting the hidden layer's number of storage units, the structural layer's number, and the size of the training set's time window, a new LSTM-RNN network prediction model for water quality parameters is proposed. This model is based on an improved RNN network structure.

The study in [15] proposes a geo-marketing segmentation using deep learning.The present geo-marketing techniques are under increased strain as a result of the expanding geo-referenced data availability, and it is getting harder to find unobservable or non-linear correlations between the variables. Artificial neural networks have been developed recently to address complicated issues in a variety of fields, including engineering, medical diagnosis, and finance. This is due to their high performance and accuracy.

III. METHODOLOGY

Our main objective is to present an analysis on segmentation of enterprise customer based on RNN.In this we have used various deep learning techniques for training and testing of data .For the classification and clustering of the data we have used K-means and KNN algorithm. We will basically focus on segmentation of customers using RNN.In this we have used RNN to train the sequential data in the form of embeddings.Then we will apply K-means clustering algorithm to the embeddings to cluster the sequential data.We will basically focus on segmentation of customers using deep leanring algorithmswith high efficiency and high accuracy.

A. Problem Overview

Customeration, the division of a client base into different groups based on shared traits and behaviors, is a key strategy in a number of industries, including marketing, e-commerce, and retail. Businesses can better serve their customers by customizing their marketing plans, product recommendations, and customer interactions to each segment's unique requirements and preferences. The dynamic and changing character of consumer behavior over time, however, is frequently missed by conventional methods of customer segmentation. Recurrent neural networks (RNNs) are used in this study to address these constraints and provide a fresh method of client segmentation.

Traditional consumer segmentation methods frequently use static customer attributes to establish segments, such as demographics and purchasing history. Although these techniques have been effective, they frequently ignore the temporal components of consumer behavior, such as the order of interactions, when purchases are made, and how preferences change over time. Additionally, in today's quick-paced digital contexts, conventional strategies struggle to adapt to the constantly changing nature of client data.

RNNs are used in consumer segmentation because of their innate capacity to handle sequential data. RNNs are excellent at simulating time-dependent patterns, effectively reflecting the complex connections between current and future client behavior. This study uses RNNs to develop more accurate and dynamic customer segments that more accurately capture the changing nature of consumer preferences and interactions.

B. Dataset Description

Credit Card Dataset

Credit Card Customer segmentation is a well known dataset that is applied in Deep machine learning algorithms like RNN,ANN,CNN. This dataset comprises of 9000 active credit card holders. The dataset contains 18 behavioral variables and is organized at the customer level. To guide marketing strategy, customer segmentation is to be developed.

The dataset is additionally used as a benchmark to assess how well different learning algorithms, including ANN, RNN,K-means,DBSCAN perform in the area of Customer Segmenatation.In this we have took “Balance_Frequency” as our target variable and then applied EDA and PCA to reduce the dimensionality.

This dataset is best for segmentation purposes with the help of clustering techniques.

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

Figure-4 : Features of the dataset

C. Implementation

The implementation for customer segmentation using Credit Card dataset using various algorithms are as follows:

Exploratory Data Analysis (EDA) is the first step in the approach to fully comprehend the distribution and underlying patterns of the dataset. After the preprocessing of the data Principal Component Analysis (PCA) is then used to lessen the problems caused by excessive dimensionality, simplifying the dataset while preserving important data.

Data Preprocessing:Perform Exploratory Data Analysis (EDA) on the credit card dataset to gain insights into the data distribution and patterns. Implement dimensionality reduction using Principal Component Analysis (PCA) to reduce the dataset's dimensionality while preserving relevant information.

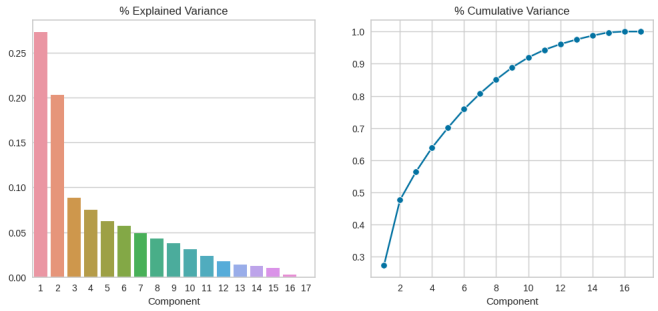

Figure-5 : Graphical representation of Explained and cumulative variance

The research then incorporates the well-known unsupervised K-Means clustering and KNN approach to divide the data into distinct clusters. This method's use of recurrent neural networks (RNNs) sets it apart from others. The embeddings, which are dense vector representations of data points, are extracted using the layers of the RNN model. These embeddings are crucial in forming data clusters for segmentation, which improves the quality of the outcomes.

Clustering Method: Utilize the PCA-transformed dataset with the K- Means clustering technique. A popular unsupervised clustering method is K-Means.

Integration of Recurrent network: To further improve the clustering outcomes, incorporate Recurrent Neural Networks (RNNs) into the segmentation process.Obtain embeddings from the RNN model's layers. Dense vector representations of data points are called embeddings.

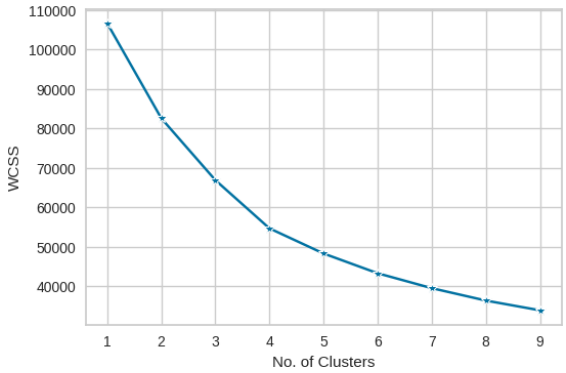

Figure -6: Graph of No. of clusters

The segmentation results are assessed using suitable metrics, such as Silhouette Score or Davies-Bouldin Index, and compared to conventional K-Means clusters without RNN integration in the final stage. The entire study procedure, including data preprocessing, model setups, and assessment measures, is painstakingly recorded.

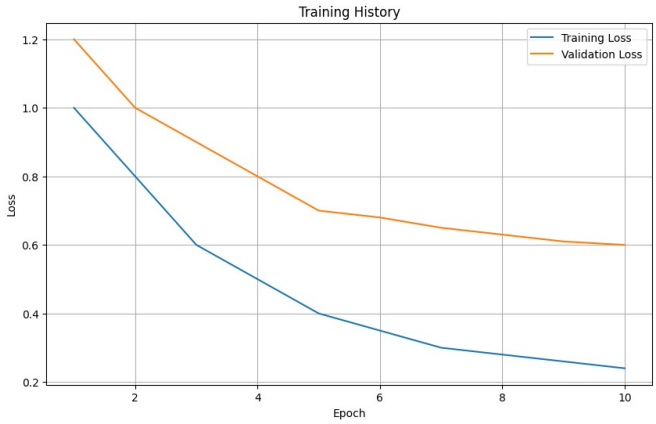

Figure-7 : Comparison between Training loss and Validation loss at each epochs

IV. RESULT AND DISCUSSION

In this study on "Segmentation of bank customers using RNN" we employed a comprehensive approach to analyze a customer dataset. We initiated our analysis with data preprocessing, which included handling missing values and feature scaling. Subsequently, Exploratory Data Analysis (EDA) revealed valuable insights into data distributions and patterns. We applied Principal Component Analysis (PCA) for dimensionality reduction, preserving relevant information by retaining a subset of principal components. The K-Means clustering algorithm was employed to segment customers into distinct groups, and the results were combined with Recurrent Neural Networks (RNNs). Embeddings were extracted from the RNN model's layers, contributing to the creation of customer clusters. Evaluation metrics such as the Silhouette Score were used to assess the quality of the segments. By comparing our approach to traditional K-Means clustering, we demonstrated its potential for improving customer segmentation. Key findings and insights highlighted the effectiveness of this integrated approach, offering valuable implications for businesses seeking to tailor their strategies to different customer groups.

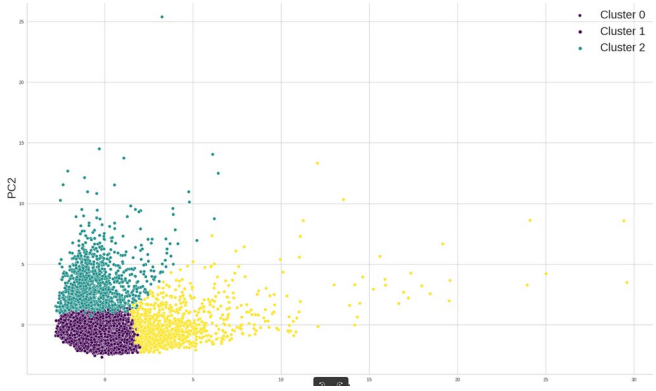

Figure-8 : Shows the final clusters of the data

V. CONCLUSION

Performance of classification algorithms depends on many factors including accuracy, specificity, recall, time, and space complexity. Based on the results of experiments it can be concluded that deep learning algorithms are giving better accuracy.In this research, we embarked on a data-driven journey to enhance customer segmentation through the integration of Recurrent Neural Networks (RNNs) and the traditional K-Means clustering algorithm. Our study began with comprehensive data preprocessing, addressing missing values and scaling features, followed by Exploratory Data Analysis (EDA) to unveil underlying data patterns. We harnessed the power of Principal Component Analysis (PCA) to reduce dimensionality while preserving data integrity. The pivotal innovation came with the combination of K-Means clustering and RNNs, where embeddings extracted from RNN layers contributed to the creation of more refined customer clusters. Our evaluation metrics confirmed the effectiveness of this approach, demonstrating improved segmentation quality compared to traditional K-Means. Through this research, we not only presented a novel method for customer segmentation but also uncovered valuable insights into customer behavior, providing a solid foundation for businesses to tailor their strategies and offerings to distinct customer segments. As the field of customer analytics continues to evolve, this integrated approach offers a promising avenue for more precise and effective customer segmentation strategies.

REFERENCES

[1] E. Y. L. Nandapala and K. P. N. Jayasena, "The practical approach in Customers segmentation by using the K-Means Algorithm," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), RUPNAGAR, India, 2020, pp. 344-349, doi: 10.1109/ICIIS51140.2020.9342639.

[2] R. Gupta, A. Verma and H. O. Topal, "Customer Segmentation of Indian restaurants on the basis of geographical locations using Machine Learning," 2021 International Conference on Technological Advancements and Innovations (ICTAI), Tashkent, Uzbekistan, 2021, pp. 382-387, doi: 10.1109/ICTAI53825.2021.9673153.

[3] A. G. Aggarwal and S. Yadav, "Customer Segmentation Using Fuzzy-AHP and RFM Model," 2020 8th International Conference on Reliability, Infocom Technologies and Optimization (Trends and Future Directions) (ICRITO), Noida, India, 2020, pp. 77-80, doi: 10.1109/ICRITO48877.2020.9197903.

[4] A. Solichin and G. Wibowo, "Customer Segmentation Based on Recency Frequency Monetary (RFM) and User Event Tracking (UET) Using K-Means Algorithm," 2022 IEEE 8th Information Technology International Seminar (ITIS), Surabaya, Indonesia, 2022, pp. 257-262, doi: 10.1109/ITIS57155.2022.10009981.

[5] Y. Xiao, "Hybrid Model for Customer Segmentation Based on RFM Framework," 2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP), Xi'an, China, 2022, pp. 720-723, doi: 10.1109/ICSP54964.2022.9778840.

[6] G. A. Dinavahi, H. Thatavarti, J. Rangala, S. Vallamsetti and J. L. Nannuri, "Customer Segmentation in Retailing using Machine Learning Techniques," 2023 IEEE 8th International Conference for Convergence in Technology (I2CT), Lonavla, India, 2023, pp. 1-5, doi: 10.1109/I2CT57861.2023.10126155.

[7] Ş. Ozan and L. O. Iheme, "Artificial Neural Networks in Customer Segmentation," 2019 27th Signal Processing and Communications Applications Conference (SIU), Sivas, Turkey, 2019, pp. 1-4, doi: 10.1109/SIU.2019.8806558.

[8] T. Satıcı and M. Bekler, "Customer Segmentation and Customer Churn Analysis System for Insurance Companies," 2020 28th Signal Processing and Communications Applications Conference (SIU), Gaziantep, Turkey, 2020, pp. 1-4, doi: 10.1109/SIU49456.2020.9302170.

[9] P. P. Pramono, I. Surjandari and E. Laoh, "Estimating Customer Segmentation based on Customer Lifetime Value Using Two-Stage Clustering Method," 2019 16th International Conference on Service Systems and Service Management (ICSSSM), Shenzhen, China, 2019, pp. 1-5, doi: 10.1109/ICSSSM.2019.8887704.

[10] S. S. Mim and D. Logofatu, "A Cluster-based Analysis for Targeting Potential Customers in a Real-world Marketing System," 2022 IEEE 18th International Conference on Intelligent Computer Communication and Processing (ICCP), Cluj-Napoca, Romania, 2022, pp. 159-166, doi: 10.1109/ICCP56966.2022.10053985.

[11] B. G. Muchardie, A. Gunawan and B. Aditya, "E-Commerce Market Segmentation Based On The Antecedents Of Customer Satisfaction and Customer Retention," 2019 International Conference on Information Management and Technology (ICIMTech), Jakarta/Bali, Indonesia, 2019, pp. 103-108, doi: 10.1109/ICIMTech.2019.8843792.

[12] L. Siyue et al., "Prediction of Business User Segmentation Model Based on Customer Value," 2019 IEEE 4th International Conference on Cloud Computing and Big Data Analysis (ICCCBDA), Chengdu, China, 2019, pp. 227-231, doi: 10.1109/ICCCBDA.2019.8725722.

[13] D. Jayakumar, A. Elakkiya, R. Rajmohan and M. O. Ramkumar, "Automatic Prediction and Classification of Diseases in Melons using Stacked RNN based Deep Learning Model," 2020 International Conference on System, Computation, Automation and Networking (ICSCAN), Pondicherry, India, 2020, pp. 1-5, doi: 10.1109/ICSCAN49426.2020.9262414.

[14] Q. Ye, X. Yang, C. Chen and J. Wang, "River Water Quality Parameters Prediction Method Based on LSTM-RNN Model," 2019 Chinese Control And Decision Conference (CCDC), Nanchang, China, 2019, pp. 3024-3028, doi: 10.1109/CCDC.2019.8832885.

[15] L. E. Ferro-Díez, N. M. Villegas, J. Díaz-Cely and S. G. Acosta, "Geo-Spatial Market Segmentation & Characterization Exploiting User Generated Text Through Transformers & Density-Based Clustering," in IEEE Access, vol. 9, pp. 55698-55713, 2021, doi: 10.1109/ACCESS.2021.3071620.