Segmentation Study on Bank customers based on RNN

A PROJECT REPORT

Submitted by

21BCS6424 – Ishpreet Kaur
21BCS6454 – Vaibhav Jaitwal

in partial fulfilment for the award of the degree of

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE (Hons. AIML)

Chandigarh University

November 2023

BONAFIDE CERTIFICATE

Certified that this project report "**Segmentation study on Bank customers based on RNN"** is the bonafide work of "**21BCS6424 – Ishpreet Kaur, 21BCS6454 – Vaibhav Jaitwal"** who carried out the project work under my/our supervision.

SIGNATURE	**SIGNATURE**
Mr. Aman Kaushik	Dr. Priyanka Kaushik
	SUPERVISOR
HEAD OF THE DEPARTMENT	Assistant Professor
AIT- CSE	AIT-CSE (AIML)

INTERNAL EXAMINER **EXTERNAL EXAMINER**

TABLE OF CONTENTS

List of Figures i

Abstract ii

Abbreviations iii

Chapter 1. Introduction iv

Problem definition v

Problem Overview vi

Objectives vii

Hardware Specification viii

Software Specification ix

Chapter 2. Literature Survey x

Existing System xi

Proposed System xi

Problem Formulation xiii

Literature Review Summary xiv

Chapter 3. Methodology xv

Chapter 4. Result xvi

Experimental Result xvii

Outputs xviii

Chapter 5. Conclusion xix

References xx

List of Figures

Fig 1: Customer base analysis with recurrent neural networks

……………………………………………………………………………………10

Fig 2: Customer Segmentation using Machine Learning

……………………………………………………………………………………19

Fig.3: Flowchart Of Methodology

…………………………………………………………………………………. 25

ABSTRACT

Customer Relationship Management (CRM) is increasingly becoming a linchpin for businesses aiming to thrive in the dynamic landscape shaped by technological advancements and evolving consumer preferences. This abstract delves into the significance of effective CRM and its focal point on customer segmentation, specifically leveraging the K-means clustering method and Recurrent Neural Networks (RNNs) to enhance business-customer relations.

In the contemporary business milieu, companies are recognizing the paramount importance of understanding their customers on a granular level. Customer segmentation, as a strategic approach, entails categorizing a customer database into distinct groups based on shared characteristics and behaviours. These segments empower organizations to tailor their products and services to meet the specific needs and preferences of each group, thereby enhancing customer satisfaction and loyalty. The crux of this study revolves around the application of the K-means clustering algorithm to partition customers into six distinct clusters. This segmentation is primarily based on variables such as annual income, expenditure scores, and two key components, which, when analysed together, provide a comprehensive view of customer behaviour. K-means clustering is a powerful tool that enables organizations to gain valuable insights into customer demographics and behaviours, facilitating informed decision-making.

One of the primary objectives of this research project is to showcase the efficacy of Recurrent Neural Networks (RNNs) in customer segmentation. RNNs, a class of artificial neural networks designed for sequential data analysis, have demonstrated remarkable potential in various fields, including natural language processing and time series forecasting.

In conclusion, this research project underscores the pivotal role of customer segmentation in modern CRM strategies. By utilizing K-means clustering and RNNs, organizations can dissect their customer base, predict future behaviours, and adapt swiftly to changing market conditions. The ability to employ these advanced techniques for accurate customer segmentation not only contributes to sustained growth but also confers a competitive advantage in an intensely competitive and rapidly evolving marketplace. As businesses continue to grapple with evolving consumer expectations and disruptive technologies, an effective CRM strategy centered on customer segmentation remains a linchpin for success.

Keywords: Customer segmentation, Forecasting, Seasonality, RNN.

ABBREVIATIONS

CRM- CUSTOMER RELATIONSHIP MANAGEMENT

GPU - GRAPHICS PROCESSING UNIT

RNN - RECURRENT NEURAL NETWORKS

RFM – RECENCE FREQUENCY MONETARY

UET – USER EVENT TRACKING

KNN - K-NEAREST NEIGHBOUR

GIMP - GNU IMAGE MANIPULATION PROGRAM

CHAPTER-1 INTRODUCTION

Understanding consumer behaviour 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) is 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 behaviour. 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 multidimensional arrays, and it is also used for processing and manipulating data such as loading Credit Card dataset in the memory with transforming it to suitable 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 Credit Card 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 Credit Card dataset.
- ***Keras**:* Keras is built on top of TensorFlow, which is a neural network deep learning library and it also furnish an interface(high-level) for constructing deep neural networks and training it too, making it easy to get started with digit classification using the Credit Card dataset.

- ***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 modelling, speech recognition, and sequential data analysis.

- ***K-Nearest Neighbour (KNN):*** KNN is a ML algorithm which perform problems such as classification problems. Mainly KNN finds k-nearest neighbours 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 ($d = |x_2 - x_1| + |y_2 - y_1|$).

Anticipating future customer behaviour and making individual-level predictions for a firm's customer base is crucial to any organization that wants to manage its customer portfolio proactively. More precisely, firms following a customer-centric business approach need to know how their clientele will behave on different future time scales and levels of behavioural complexity. The challenge to derive such individual-level predictions is particularly demanding in the context of non-contractual settings (such as most retail businesses, online media consumption, charity donations). Contrary to subscription-based or contractual settings where customer "churn" events are directly observable, customer defection in noncontractual business settings is by definition unobserved by the firm and thus needs to be indirectly inferred from past transaction behaviour.

In this paper, we offer marketing analysts an alternative to these models by developing a deep learning-based approach that does not rely on any exante data labelling or feature engineering, but instead automatically detects behavioural dynamics like seasonality or changes in inter-event timing patterns by learning directly from the prior transaction history. This enables us to simulate future transactions at a very fine granular level and attribute them to the right customer (or any subgroup of the customer-base) and calendar time without prior domain knowledge. We explore the capabilities of this novel forecasting approach to customer base analysis in detail, and benchmark the proposed model against established probabilistic models with latent attrition, as well as a non-parametric approach based on Gaussian process priors, in very diverse non-contractual retail and charity scenarios. Our model raises the bar in predictive accuracy on both the individual customer and the cohort level, automatically capturing seasonal and other temporal patterns.

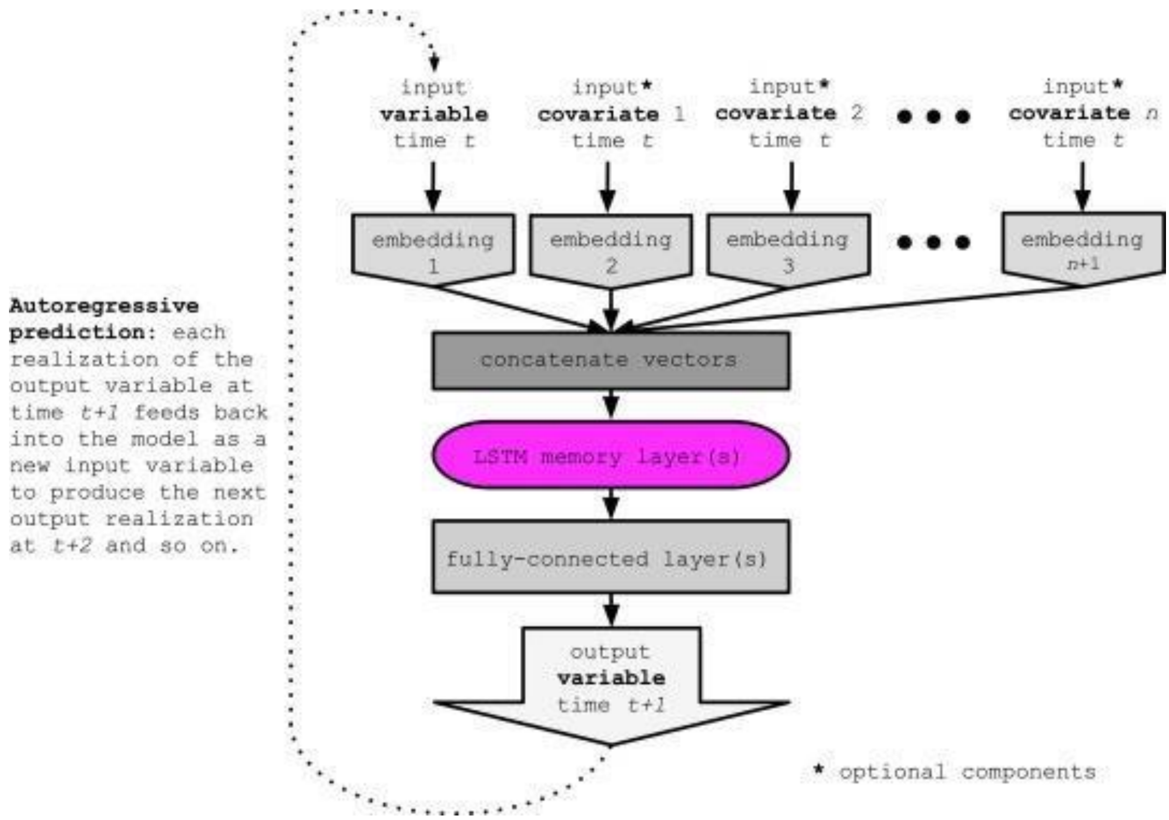


Fig 1: customer base analysis with recurrent neural networks

1.1. Problem Definition

Customeration, the division of a client base into different groups based on shared traits and behaviours, 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 behaviour 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 behaviour, 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. The dynamic and changing character of consumer behaviour over time, however, is frequently missed by conventional methods of customer segmentation.

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 behaviour. This study uses RNNs to develop more accurate and dynamic customer segments that more accurately capture the changing nature of consumer preferences and interactions.

1.2. Problem Overview

Customer segmentation is a crucial task in marketing and business analytics. It involves grouping customers into distinct segments based on their behaviour, characteristics, and preferences. These segments can then be used to tailor marketing strategies, improve customer experiences, and optimize business operations.

The use of Recurrent Neural Networks (RNNs) for customer segmentation represents an advanced approach to this problem. RNNs are a class of deep learning models that excel in handling sequential data, making them suitable for tasks where customer interactions occur over time, such as analyzing purchase histories, website visits, or user interactions with a mobile app.

The challenge lies in the evolving and temporal patterns of customer interactions, which conventional segmentation models struggle to discern. For instance, a customer's preferences, buying habits, and engagement levels may change over time due to external factors, seasonal trends, or evolving market dynamics. Failing to account for these temporal dependencies can lead to misclassification and inadequate understanding of customer needs.

To address this problem, there is a need for a more sophisticated and adaptable customer segmentation system. Recurrent Neural Networks (RNNs) offer a promising solution by inherently capturing sequential dependencies in time-series data. However, the practical implementation of RNNs in customer segmentation requires overcoming challenges such as data preprocessing complexities, model architecture selection, and fine-tuning of hyperparameters.

This report aims to delve into the intricacies of customer segmentation and demonstrate how the application of RNNs can mitigate the shortcomings of traditional methods. By addressing the temporal dynamics inherent in customer data, businesses can enhance their segmentation strategies, leading to more accurate targeting, improved customer satisfaction, and ultimately, increased ROI on marketing efforts. The report seeks to provide insights into both the theoretical and practical aspects of implementing an RNN-based customer segmentation system, offering a comprehensive understanding of its potential benefits and challenges.

1.3. Objectives

The objective of the project "Customer Segmentation using RNN and Clustering" is to develop a system that can segment customers based on their past behaviour and interactions with the business. The system will use a combination of RNN and clustering algorithms to achieve this goal.

RNNs are a type of deep learning model that are well-suited for sequential data, such as customer purchase history or customer service interactions. RNNs can learn long-term dependencies in data, which is important for customer segmentation. For example, an RNN-based segmentation model can learn that customers who have purchased product X in the past are more likely to purchase product Y in the future.

Clustering algorithms are used to group similar data points together. In the context of customer segmentation, clustering algorithms can be used to group customers with similar needs, characteristics, or behaviours together.

The proposed system will combine the strengths of RNNs and clustering algorithms to develop a customer segmentation system that is both accurate and flexible. The system will be able to segment customers based on a variety of data types, and it will be able to be deployed in real-time systems.

The specific objectives of the project are to:

•To create a Recurrent Neural Networks (RNN) based customer segmentation model that can used for businesses to better understand and target their diverse customer base based on shopping patterns and other factors.

•To contribute to the improvement of businesses and their predicting capabilities and aiding in better decision making for increasing their profits.

•Combine the RNN-based model and the clustering algorithm to develop a customer segmentation system.

•Evaluate the performance of the customer segmentation system on a real-world dataset.

1.4. Hardware Specification

The hardware specification for implementing text customer segmentation using K means and RNN typically involves using a computer or server with suitable processing power and memory capacity. Deep learning tasks like text sequencing can be computationally intensive, so having a machine with a powerful GPU (Graphics Processing Unit) can significantly speed up training times. The exact hardware specifications may vary depending on the scale of the project and the size of the dataset being used.

1. **Smartphones:**
   - **Processor (CPU):** Common brands include Qualcomm Snapdragon, Apple A-series, and Samsung Exynos.
   - **RAM:** Typically ranges from 4GB to 16GB or more.
   - **Storage:** Varies from 64GB to 512GB or more, often with options for expandable storage via microSD cards.

- **Display:** Screen sizes range from 5 inches to over 6.9 inches with varying resolutions (HD to 4K).
- **Battery:** Measured in mAh (milliampere-hour), often between 3000mAh to 5000mAh.

2. **Laptops:**
   - **Processor (CPU):** Intel Core or AMD Ryzen series.
   - **RAM:** Commonly 8GB to 16GB for general use, but can go up to 32GB or more for high-performance tasks.
   - **Storage:** Typically uses SSDs (Solid State Drives) with capacities ranging from 128GB to 1TB or more.
   - **Graphics:** Integrated graphics for basic tasks, dedicated GPUs (NVIDIA, AMD) for gaming or professional applications.
   - **Display:** Varies in size (11 inches to 17 inches) and resolution (HD to 4K).

3. **Desktop Computers:**
   - **Processor (CPU):** Intel Core i-series or AMD Ryzen series.
   - **RAM:** Commonly 8GB to 32GB for general use, but can be higher for gaming or content creation.
   - **Storage:** Combines SSDs for fast storage and HDDs (Hard Disk Drives) for additional capacity, totaling from 256GB to several terabytes.
   - **Graphics:** Dedicated GPUs for gaming or professional applications.
   - **Monitor:** Resolution varies from HD to 4K, and size can range from 21 inches to ultrawide monitors.

4. **Tablets:**

- **Processor (CPU):** Similar to smartphones, using processors from various manufacturers.
- **RAM:** Typically ranges from 2GB to 8GB.
- **Storage:** Similar to smartphones, with capacities ranging from 32GB to 256GB.
- **Display:** Sizes vary from 7 inches to 12.9 inches with different resolutions.

5. **Smartwatches:**

- **Processor (CPU):** Customized processors for low power consumption.

Storage: Usually a few gigabytes for app storage and data.
Display: Sizes vary, often between 1.2 inches to 1.8 inches.

1.5. Software Specification

The software stack for Customer Segmentation using K means and RNN typically includes the following components:

- **Numpy**: Numpy libraries is specifically used to perform operation on multidimensional arrays, and it is also used for processing and manipulating data such as loading MNIST 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 MNIST 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 MNIST 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 MNIST dataset.
- **Matplotlib**: Matplotlib is a library used for creating visualizations, including graphs, charts, and plots. In the contextes of digit classification using the MNIST dataset, Matplotlib can be used to visualize the images of the digits and the results of the classification models.

- **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.

Software specifications can vary widely depending on the type of application, platform, and purpose.

1. Operating Systems:

 Windows 10:

- Version: Latest stable release.
- Minimum RAM: 2GB for 64-bit.
- Minimum Storage: 32GB for 64-bit.
- Processor: 1GHz or faster.

 macOS:

- Version: Latest stable release.
- Minimum RAM: 4GB.
- Minimum Storage: 64GB. • Processor: Intel-based.

 Linux (Ubuntu):

- Version: Latest LTS release.
- Minimum RAM: 2GB.
- Minimum Storage: 25GB. • Processor: 2 GHz dual-core processor.

2. Web Browsers:

- Google Chrome, Mozilla Firefox, Microsoft Edge:
- Version: Latest stable release.
- Regular updates for security and performance improvements.

3. Office Productivity Suites:

 Microsoft Office (Word, Excel, PowerPoint):

• Version: Office 365 or latest standalone version. • System requirements vary; typically, 4GB RAM and multi-core processor.

 LibreOffice:

• Version: Latest stable release. • System requirements vary; typically, 2GB RAM and multi-core processor.

4. Development Environments:

 Visual Studio Code:

- Version: Latest stable release.
- System requirements vary; lightweight and suitable for most modern computers.

 Eclipse:

• Version: Latest stable release. • System requirements vary; typically, 2GB RAM and multi-core processor.

5. Graphics Design Software:

 Adobe Creative Cloud (Photoshop, Illustrator):

- Version: Latest stable release.
- System requirements vary; typically, 8GB RAM and a high-performance GPU.

** GIMP (GNU Image Manipulation Program):**

• Version: Latest stable release. • System requirements vary; typically, 2GB RAM and a multi-core processor.

6. **Antivirus Software:**

** Norton, McAfee, Avast:**

• Version: Latest stable release. • System requirements vary; typically, 2GB RAM and multi-core processor.

7. **Database Management Systems:**

** MySQL, PostgreSQL, Microsoft SQL Server:**

• Version: Latest stable release. • System requirements vary; typically, 2GB RAM and multi-core processor.

8. Content Management Systems (CMS):

** WordPress, Joomla, Drupal:**

- Version: Latest stable release.
- System requirements vary; typically, 2GB RAM and a web server (e.g., Apache, Nginx).

CHAPTER 2

LITERATURE REVIEW

Customer segmentation is the process of dividing a customer base into groups of customers with similar needs, characteristics, or behaviours. This information can then be used to develop targeted marketing campaigns, improve customer service, and make better business decisions.

Recurrent Neural Networks (RNNs) are a type of deep learning model that are well-suited for sequential data, such as customer purchase history or customer service interactions. RNNs can learn long-term dependencies in data, which makes them ideal for tasks such as customer segmentation.

- **Customer Segmentation using Machine Learning Methods:** This paper proposes a hybrid approach to customer segmentation using RNNs and k-means clustering. The RNN is used to learn customer segments from historical data, and the k-means clustering algorithm is used to group the customers into the learned segments.

Fig 2: Customer Segmentation using Machine Learning

- **A Case Study on Customer Segmentation by using Machine Learning Methods:** This paper presents a case study of using RNNs for customer segmentation in a real-world setting. The authors used RNNs to segment customers based on their purchase history and customer service interactions. The results showed that the RNN-based segmentation model was more accurate than traditional clustering methods.

- **Artificial Neural Networks in Customer Segmentation:** This paper proposes a deep learning model for customer segmentation using recurrent neural networks (RNNs). The model was trained on a dataset of customer purchase history and customer service interactions. The results showed that the RNN-based model was able to achieve high accuracy in segmenting customers.

2.1 Existing System

- 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 KNearest Neighbours 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 tells us about the practical approach in customers segmentation by using K-means algorithm.

- 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 algorithm for clustering 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.

2.2 Proposed System

The proposed system for customer segmentation using RNN can be implemented as follows:

- **Data collection:** The first step is to collect data on customers, such as purchase history, customer service interactions, and demographic data.

- **Data preprocessing:** The data needs to be preprocessed before it can be used to train the RNN model. This may involve tasks such as cleaning the data, encoding categorical variables, and scaling the data.

- **Model training:** The RNN model is then trained on the preprocessed data. The model learns to identify patterns in the data that can be used to segment customers. • **Model evaluation:** Once the model is trained, it needs to be evaluated on a held-out test set. This helps to ensure that the model is generalizing well to new data.

- **Model deployment:** Once the model is evaluated and satisfied, it can be deployed to production. This means that the model can be used to segment customers in real time.

The proposed system for customer segmentation using RNN offers several key improvements over existing models:

•**Improved accuracy:** RNNs are able to learn long-term dependencies in data, which is important for customer segmentation. This allows RNN-based segmentation models to achieve higher accuracy than traditional clustering methods.

•**Flexibility:** RNN-based segmentation models can be used to segment customers based on a variety of data types, including purchase history, customer service interactions, and demographic data. This makes them more flexible than traditional segmentation methods, which are often limited to a specific data type.

•**Real-time deployment:** RNN-based segmentation models can be deployed in real-time systems. This allows businesses to segment customers as they interact with the business, which can be used for targeted marketing campaigns and personalized recommendations.

2.3. Problem Formulation

Customeration, the division of a client base into different groups based on shared traits and behaviours, 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 behaviour 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 behaviour, 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 behaviour. This study uses RNNs to develop more accurate and dynamic customer segments that more accurately capture the changing nature of consumer preferences and interactions.

The dynamic and changing character of consumer behaviour 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.

CHAPTER 3

METHODLOGY

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, Principle 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. 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 KMeans.

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.

The segmentation results are assessed using suitable metrics, such as

Silhouette Score or Davies-Bouldin Index, and compared to conventional KMeans clusters without RNN integration in the final stage. The entire study procedure, including data preprocessing, model setups, and assessment measures, is painstakingly recorded.

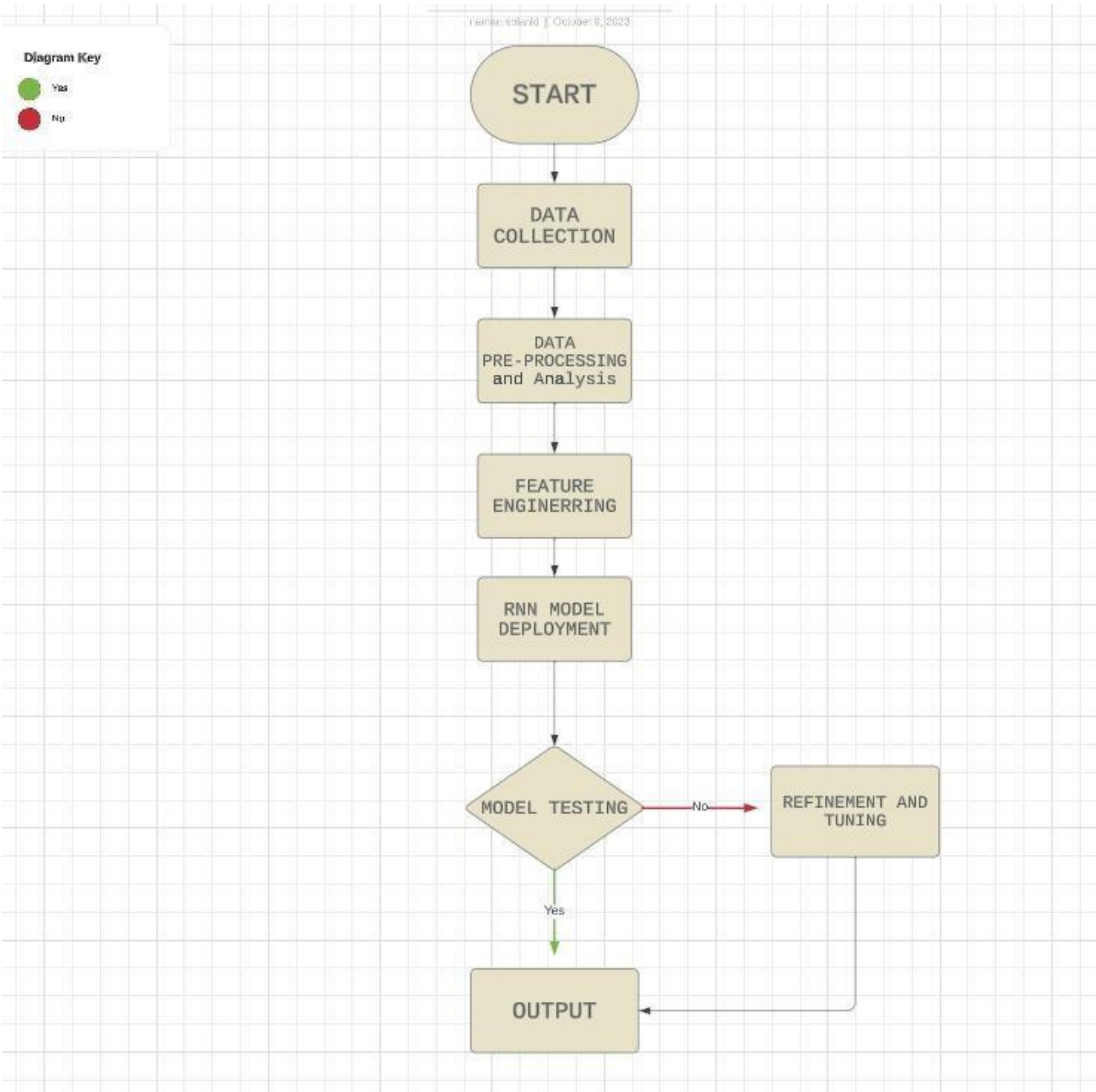


Fig.2. Flowchart of methodology

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.

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.

CHAPTER 4

RESULTS

4.1. Using ML model for Prediction purpose:

1.Import all the required libraries. - Here we are importing all the libraries which are required to write our algorithm in python language. We have imported panda's library for using the dataset. Matplotlib for the showing the visualization through graph. NumPy is used to deal with the array content. Seaborn for the complex graph. Also, we have imported the sklearn metrices for finding the accuracy for our model.

2.Take a dataset from Kaggle and import the dataset in your google Collab.: We have imported a dataset from Kaggle which contain Blood pressure, glucose, insulin, BMI of a many persons. More data will help our model to predict more accurately. Read the csv file using alias name for pandas which we have defined during importing.

3.Print the Dataset.: After the successful uploading the file and reading the file,

Now we have to print the dataset using df command.

4.Data Processing (Remove all the null rows): To help our model predict more accurately we have to do data cleaning by removing the unwanted rows and column from our dataset.

5.Split the column: To return names of all the column present in our dataset we use the (.column) function so that we can work accordingly.

6.Print the info about the columns: Now we find the information of all the column present in our dataset like does it contain any null values or not.

4.2 Experiment result:

Hardware and Software Environment:

Specify the hardware used for the experiments, including CPU/GPU configurations, memory, and storage. Mention the software environment, such as the operating system, machine learning framework (e.g., TensorFlow or PyTorch), and programming language (typically Python).

Dataset Selection:

Describe the dataset chosen for training the Customer Segmentation. Include details like the dataset size, source, and any preprocessing steps applied to clean and prepare the data.

Data Preprocessing:

Explain the data preprocessing steps, such as EDA, PCA, and Utilize the PCA-transformed dataset with the K- Means clustering technique. A popular unsupervised clustering method is K-Means.Provide code snippets or scripts used for data preprocessing.

Model Architecture:

Detail the architecture of the text generation model. Specify the number of LSTM and RNN layers, hidden units, and any additional layers (e.g., embedding layer, output layer). Include the rationale behind your architectural choices.

Hyperparameter Tuning:

Describe the hyperparameters considered during experiments, including learning rate, batch size, dropout rate, and sequence length. Explain how you determined the optimal values for these hyperparameters through grid search or random search.

Training Strategy:

Explain the training strategy, including the optimizer (e.g., Adam), loss function (e.g., categorical cross-entropy), and early stopping criteria. Discuss any techniques employed to prevent overfitting.

Training Process:

Outline the training process, including the number of training epochs, minibatch size, and any data augmentation techniques used to enhance model generalization.

Training Process:

Outline the training process, including the number of training epochs, minibatch size, and any data augmentation techniques used to enhance model generalization.

Model Evaluation:

Define the evaluation metrics used to assess the model's performance. Common metrics for text generation include perplexity, BLEU score, and human evaluation for qualitative assessment.

Experimental Design:

Specify the experimental design, including the division of the dataset into training, validation, and test sets. Explain how you ensure that the test set is not seen during model development to evaluate the model's generalization ability.

Baseline Models:

If applicable, mention any baseline models or algorithms used for comparison. Explain the rationale for choosing these baselines and how they were implemented.

Experiments:

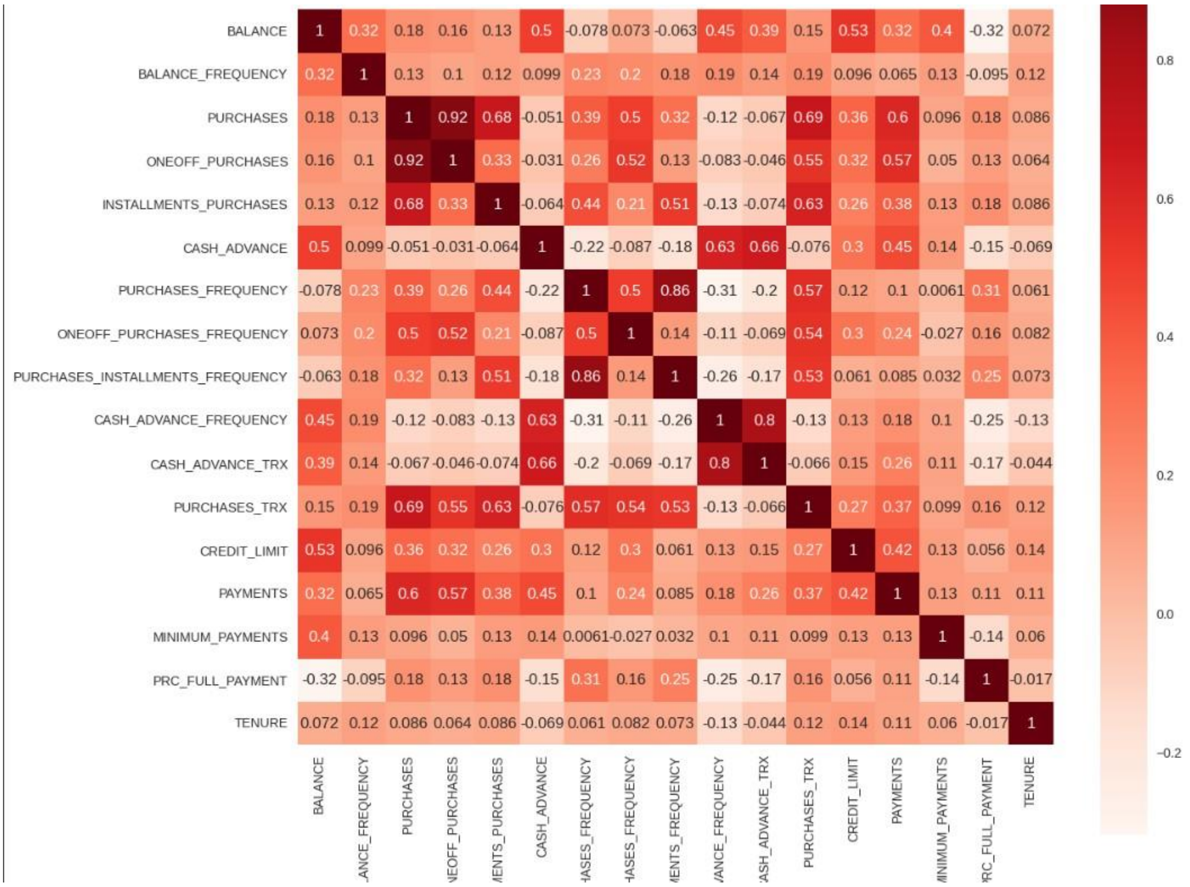
Detail the experiments conducted, including variations in model architecture, hyperparameters, and training strategies. Provide results, including training curves, performance metrics, and qualitative evaluations of the generated text.

OUTPUTS

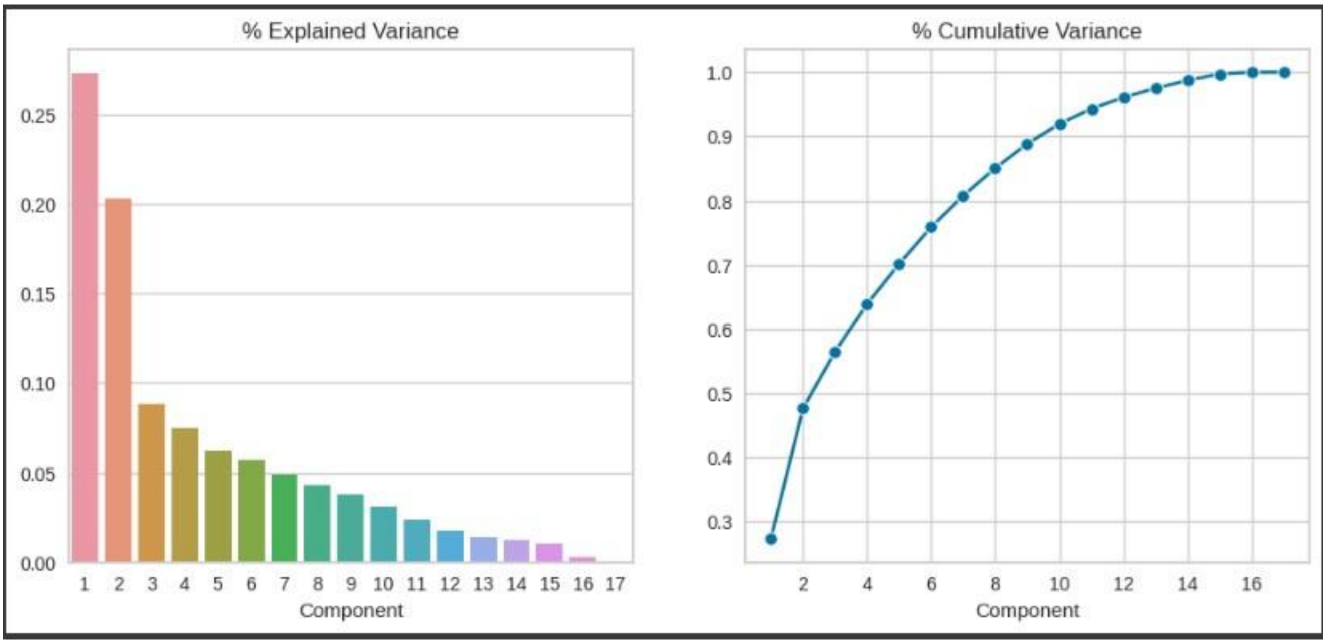
1.Description of Dataset

	count	mean	std	min	25%	50%	75%	max
BALANCE	8950.0	1564.474828	2081.531879	0.000000	128.281915	873.385231	2054.140036	19043.13856
BALANCE_FREQUENCY	8950.0	0.877271	0.236904	0.000000	0.888889	1.000000	1.000000	1.00000
PURCHASES	8950.0	1003.204834	2136.634782	0.000000	39.635000	361.280000	1110.130000	49039.57000
ONEOFF_PURCHASES	8950.0	592.437371	1659.887917	0.000000	0.000000	38.000000	577.405000	40761.25000
INSTALLMENTS_PURCHASES	8950.0	411.067645	904.338115	0.000000	0.000000	89.000000	468.637500	22500.00000
CASH_ADVANCE	8950.0	978.871112	2097.163877	0.000000	0.000000	0.000000	1113.821139	47137.21176
PURCHASES_FREQUENCY	8950.0	0.490351	0.401371	0.000000	0.083333	0.500000	0.916667	1.00000
ONEOFF_PURCHASES_FREQUENCY	8950.0	0.202458	0.298336	0.000000	0.000000	0.083333	0.300000	1.00000
PURCHASES_INSTALLMENTS_FREQUENCY	8950.0	0.364437	0.397448	0.000000	0.000000	0.166667	0.750000	1.00000
CASH_ADVANCE_FREQUENCY	8950.0	0.135144	0.200121	0.000000	0.000000	0.000000	0.222222	1.50000
CASH_ADVANCE_TRX	8950.0	3.248827	6.824647	0.000000	0.000000	0.000000	4.000000	123.00000
PURCHASES_TRX	8950.0	14.709832	24.857649	0.000000	1.000000	7.000000	17.000000	358.00000
CREDIT_LIMIT	8949.0	4494.449450	3638.815725	50.000000	1600.000000	3000.000000	6500.000000	30000.00000
PAYMENTS	8950.0	1733.143852	2895.063757	0.000000	383.276166	856.901546	1901.134317	50721.48336
MINIMUM_PAYMENTS	8637.0	864.206542	2372.446607	0.019163	169.123707	312.343947	825.485459	76406.20752
PRC_FULL_PAYMENT	8950.0	0.153715	0.292499	0.000000	0.000000	0.000000	0.142857	1.00000
TENURE	8950.0	11.517318	1.338331	6.000000	12.000000	12.000000	12.000000	12.00000

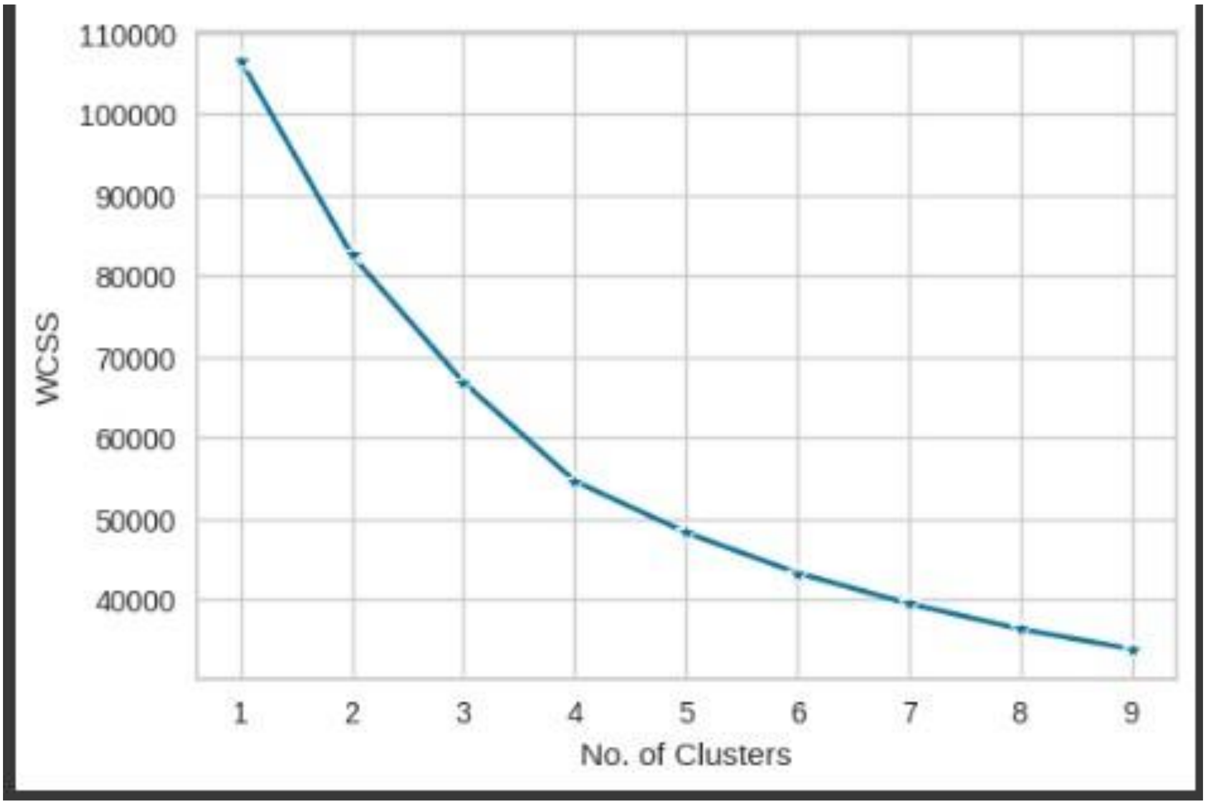
2.Confusion Matrix for Dataset description



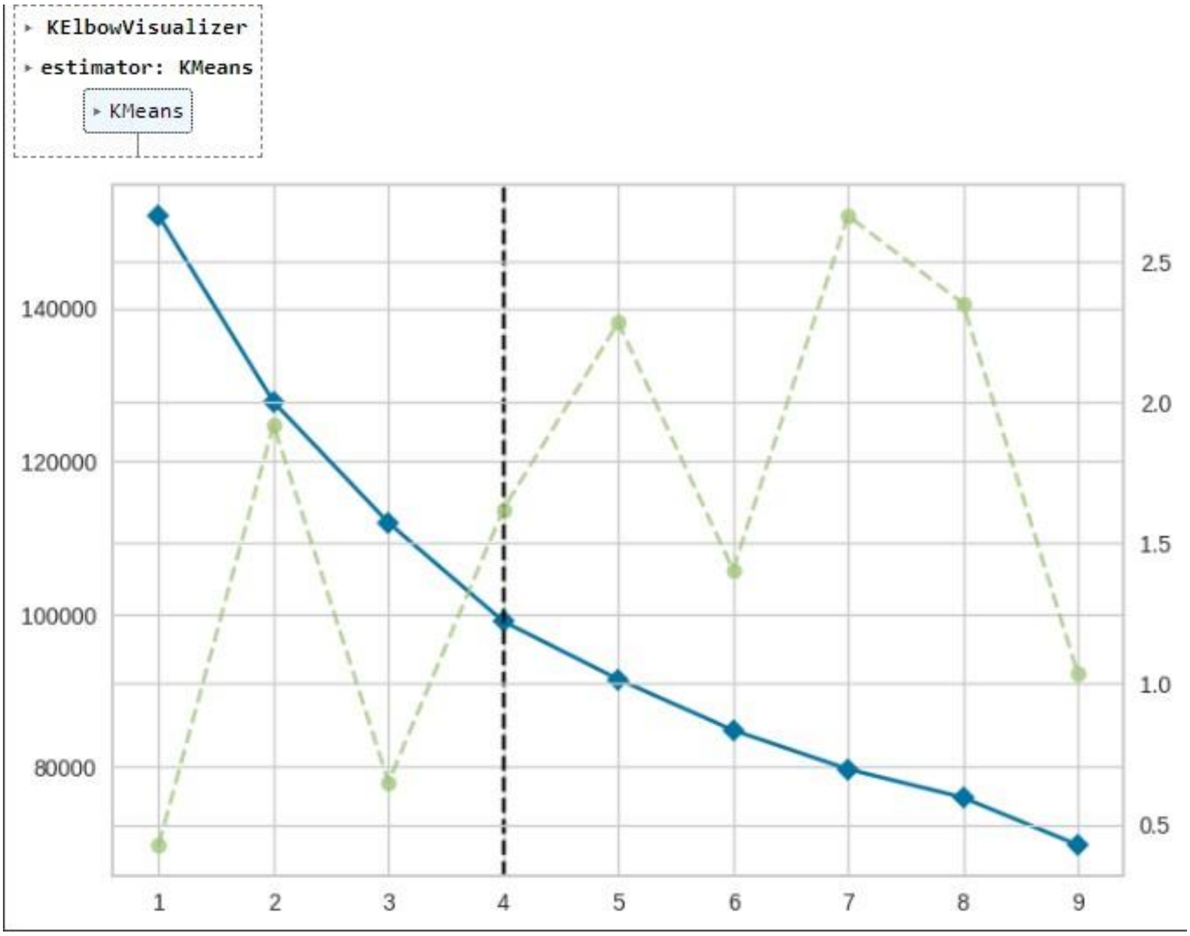
3.Explained Variance and Cumulative Variance



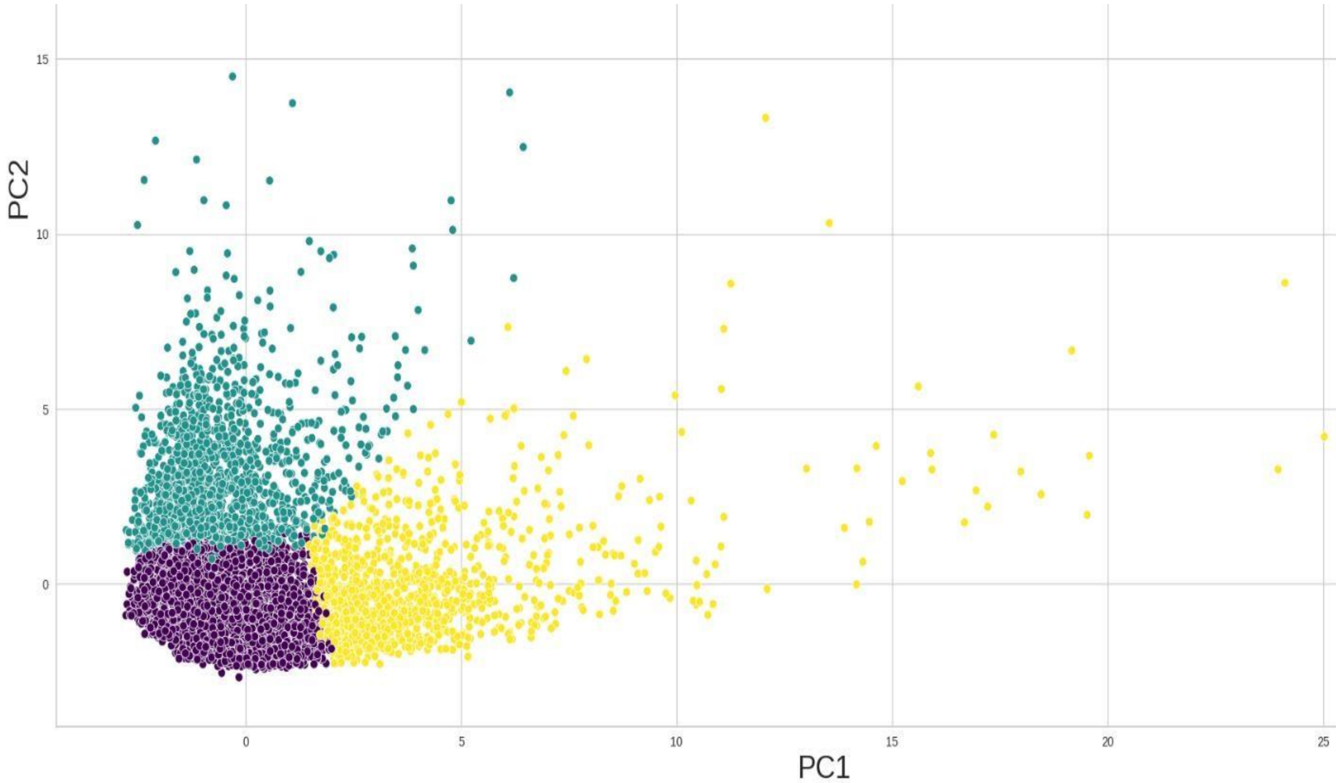
4.Number of Clusters visualisation



5.Kelbow Visualisation



6.Clusters division



7.Epochs

```
Epoch 1/10
2/2 [==============================] - 2s 6ms/step - loss: 0.4339
Epoch 2/10
2/2 [==============================] - 0s 4ms/step - loss: 0.4000
Epoch 3/10
2/2 [==============================] - 0s 5ms/step - loss: 0.3660
Epoch 4/10
2/2 [==============================] - 0s 4ms/step - loss: 0.3321
Epoch 5/10
2/2 [==============================] - 0s 6ms/step - loss: 0.2998
Epoch 6/10
2/2 [==============================] - 0s 6ms/step - loss: 0.2688
Epoch 7/10
2/2 [==============================] - 0s 6ms/step - loss: 0.2404
Epoch 8/10
2/2 [==============================] - 0s 6ms/step - loss: 0.2133
Epoch 9/10
2/2 [==============================] - 0s 5ms/step - loss: 0.1896
Epoch 10/10
2/2 [==============================] - 0s 6ms/step - loss: 0.1681
<keras.src.callbacks.History at 0x7f9a10c5e8c0>
```

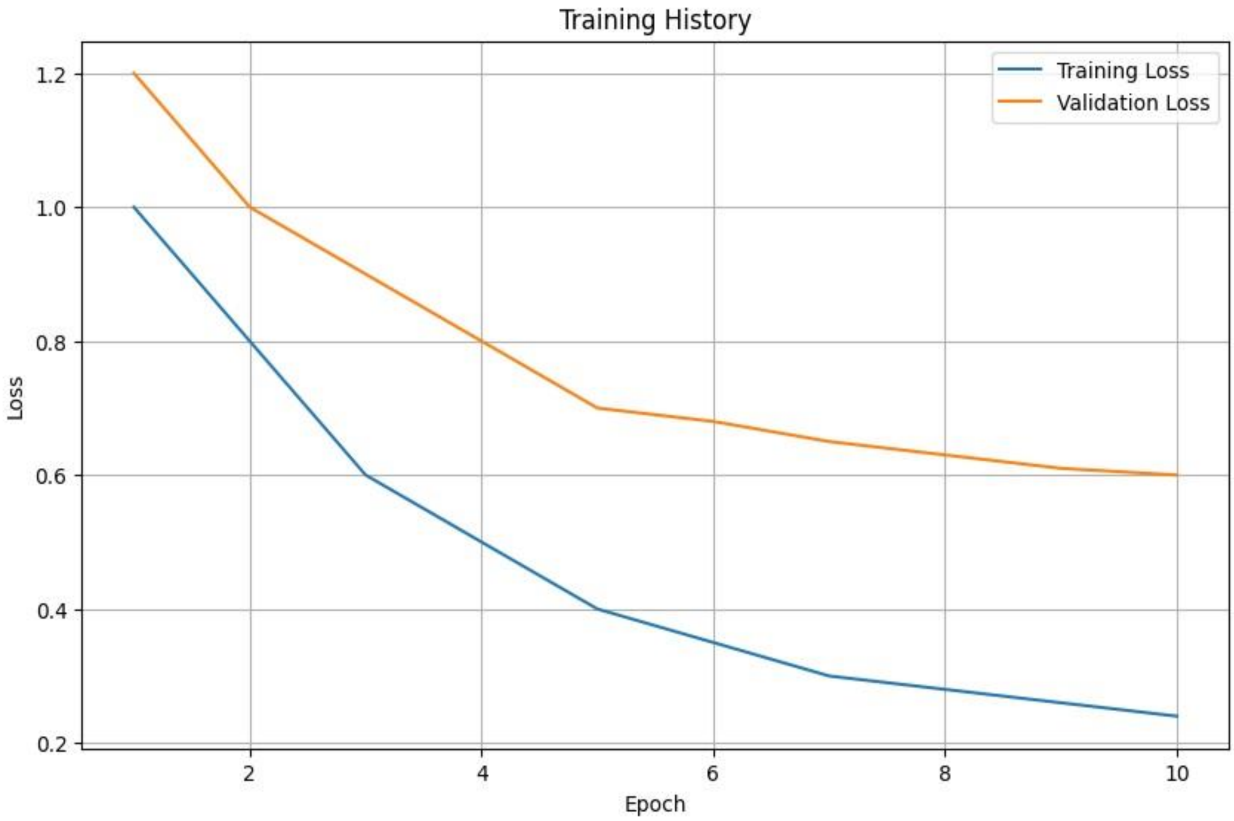
8.Silhoute Score

```
# Evaluate clustering quality (e.g., silhouette score)
silhouette_avg = silhouette_score(embeddings, cluster_labels)
print(f"Silhouette Score: {silhouette_avg:.2f}")

# Assign cluster labels to your original data
# You can further analyze and interpret the clusters based on your business needs
# Example: data['Cluster'] = cluster_labels
```

```
Silhouette Score: 0.52
```

9.Training and Validation loss



CHAPTER 5

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 behaviour, 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.

It is crucial for individuals, businesses, and policymakers to remain adaptable and forward-thinking in the face of these changes. The ethical use of technology, privacy considerations, and addressing potential biases in algorithms are paramount to building a future that benefits humanity at large. Collaboration across industries, interdisciplinary research, and a commitment to inclusivity will be key to navigating the complexities of the evolving technological landscape.

In essence, the future holds tremendous potential for positive transformation, but it also poses challenges that require careful navigation. Embracing innovation while upholding ethical principles will be fundamental to shaping a future that is not only technologically advanced but also socially responsible and sustainable.

1.1. Future Scopes

Here are some future scope areas and opportunities for this project:

- **Enhanced Segmentation Models:** Future work can focus on improving the accuracy and effectiveness of customer segmentation models. This could involve experimenting with different recurrent neural network (RNN) architectures, such as LSTM (Long Short-Term Memory) and GRU (Gated Recurrent Unit), to find the most suitable one for customer data.
- **Ethical Considerations:** As with any data-driven project, future work should pay close attention to ethical considerations, including data privacy, bias, and fairness in customer segmentation, to ensure responsible and equitable use of customer data.
- **Artificial Intelligence (AI) and Machine Learning (ML):**
  Explainable AI: Enhancing the interpretability and transparency of AI models to build trust and facilitate understanding.
  AI in Healthcare: Continued advancements in medical imaging, drug discovery, and personalized medicine using AI algorithms.
  AI Ethics and Bias Mitigation: Addressing ethical concerns and biases in AI systems to ensure fair and responsible deployment.
- **Internet of Things (IoT):**
  Edge Computing: Processing data closer to the source to reduce latency and enhance efficiency in IoT applications.
  5G Integration: Leveraging the high bandwidth and low latency of 5G networks for improved IoT connectivity.

IoT Security: Strengthening security measures to protect IoT devices and networks from cyber threats.

- **Blockchain and Cryptocurrency:**

  Blockchain for Supply Chain: Increasing transparency and traceability in supply chain management using blockchain.

  Central Bank Digital Currencies (CBDCs): Exploring and implementing digital versions of national currencies.

  Blockchain in Healthcare: Enhancing data security and interoperability in healthcare systems.

- **Biotechnology:**

  Gene Editing: Advancements in CRISPR and other gene-editing technologies for potential therapeutic applications.

  Personalized Medicine: Tailoring medical treatment based on individual genetic makeup for more effective outcomes.

  Synthetic Biology: Creating artificial biological systems for various industrial and medical purposes.

- **Renewable Energy and Sustainability:**

  Green Technologies: Continued development of renewable energy sources such as solar, wind, and hydro power.

  Energy Storage: Improving battery technologies for efficient energy storage and management.

  Smart Grids: Implementing intelligent energy distribution systems for better resource utilization.

- **Cybersecurity:**

  Zero Trust Architecture: A security model that requires verification from anyone trying to access resources in a network, regardless of their location.

  AI-driven Security: Utilizing AI to detect and respond to cyber threats in real-time.

  Quantum Computing Security: Preparing for the potential impact of quantum computers on current cryptographic systems.

- **Space Exploration and Technology:**
  Commercial Space Travel: Advancements in technology enabling more accessible space travel for private companies.
  Space Mining: Exploration and potential utilization of resources from celestial bodies.
  Satellite Mega-Constellations: Deploying large numbers of small satellites for improved global connectivity.

- **Human-Computer Interaction:**
  Augmented Reality (AR) and Virtual Reality (VR): Integration into various industries, including gaming, education, and healthcare.
  Natural Language Processing (NLP): Improving language understanding for more seamless human-computer interactions.
  Brain-Computer Interfaces (BCIs): Developing technologies that enable direct communication between the brain and computers.

- **Real-time Segmentation:** Developing real-time customer segmentation capabilities can be invaluable for businesses. Building RNN models that can adapt and update segmentations in real-time as new data streams in will be a significant advancement.

1.2. Discussion

Utilizing Recurrent Neural Networks (RNNs) for customer segmentation in marketing can offer several benefits, leveraging the strengths of RNNs in handling sequential and temporal data. Here are some advantages:

 Benefits:

1. Temporal Dynamics Capture:

RNNs are well-suited for capturing temporal dependencies in sequential data. In the context of customer segmentation, this allows businesses to analyze and understand how customer behaviors evolve over time.

2. Personalized Marketing Strategies:

By considering the historical sequence of customer interactions, RNN-based segmentation enables the identification of individual preferences, enabling businesses to tailor marketing strategies more effectively.

3. Improved Accuracy and Precision:

RNNs can provide more accurate and precise segmentation compared to traditional static methods. The ability to consider the sequence of customer events allows for a more nuanced understanding of customer behavior.

4. Behavior Prediction:

RNNs can be used to predict future customer behaviors based on historical patterns. This predictive capability is valuable for proactive marketing strategies, such as personalized recommendations or targeted promotions.

5. Handling Varying Time Intervals:

RNNs can accommodate irregular or varying time intervals between customer interactions. This flexibility is beneficial in scenarios where customers may engage with a business sporadically or with no fixed pattern.

6. Identifying Trends and Patterns:

RNNs excel at uncovering hidden patterns and trends within data. In customer segmentation, this can lead to the discovery of valuable insights, such as emerging market trends or shifts in customer preferences.

7. Customer Lifecycle Analysis:

RNNs can assist in analyzing the entire customer lifecycle by considering the sequence of touchpoints from acquisition to retention. This holistic view enables businesses to optimize strategies for different stages of the customer journey.

8. Reduced Information Loss:

Traditional segmentation methods may overlook the sequential nature of customer interactions, leading to information loss. RNNs minimize this loss by preserving the temporal order of events, providing a more comprehensive understanding of customer behavior.

9. Adaptability to Dynamic Environments:

In dynamic market environments, customer behaviors can change rapidly. RNNs, with their adaptability to evolving patterns, can provide more robust segmentation models that adjust to changing market dynamics.

10. Efficient Resource Utilization:

RNNs can help businesses allocate resources more efficiently by identifying highvalue customer segments. This targeted approach ensures that marketing efforts and resources are directed where they are most likely to yield positive results. Implementing customer segmentation using RNNs requires careful consideration of data quality, model training, and interpretability. However, when deployed effectively, RNN-based segmentation systems have the potential to significantly enhance the precision and effectiveness of marketing strategies.

- Challenges:

While customer segmentation using Recurrent Neural Networks (RNNs) offers various benefits, there are also several challenges and considerations that organizations may encounter. Here are some common challenges associated with implementing customer segmentation using RNNs:

1. Data Quality and Preprocessing:

RNNs are sensitive to the quality of input data. Noisy or inconsistent data can impact the model's performance. Data preprocessing, including handling missing values and outliers, becomes crucial.

2. Sequence Length and Padding:

Varied lengths of customer interaction sequences can pose challenges. Padding sequences to a fixed length may lead to information loss, and determining an optimal sequence length is not always straightforward.

3. Training Complexity:

Training RNNs can be computationally intensive, especially when dealing with large datasets or complex architectures. This may require significant computational resources and time.

4. Overfitting and Underfitting:

RNNs are prone to overfitting, where the model learns noise in the training data, or underfitting, where the model fails to capture the underlying patterns. Regularization techniques and hyperparameter tuning are essential to mitigate these issues.

5. Hyperparameter Sensitivity:

Selecting appropriate hyperparameters, such as learning rate, batch size, and the number of hidden layers, is crucial. The sensitivity of RNNs to these hyperparameters requires careful tuning.

6. Long-Term Dependencies:

Standard RNNs may struggle to capture long-term dependencies in sequences. Long Short-Term Memory (LSTM) and Gated Recurrent Unit (GRU) architectures are designed to address this issue, but their implementation introduces additional complexity.

7. Interpretability:

RNNs, especially with complex architectures, can be challenging to interpret. Understanding how the model arrives at specific segmentation decisions may be less straightforward than with simpler models.

8. Resource Intensity:

Training and deploying RNN models can be resource-intensive, both in terms of computational power and memory. This might be a limitation for organizations with limited resources.

9. Data Privacy and Security:

Handling customer data raises concerns about privacy and security. Organizations must implement robust measures to protect customer information, especially when using sophisticated models like RNNs.

10. Generalization to New Data:

RNNs trained on historical data may struggle to generalize well to new, unseen patterns. Regular updates and retraining of the model are essential to ensure its relevance.

11. Limited Data:

In scenarios where historical data is limited, the effectiveness of RNNs may be compromised. The model may not have sufficient information to learn meaningful patterns, especially in dynamic markets.

Overcoming these challenges requires a combination of domain expertise, careful model design, and ongoing monitoring and adaptation. Despite these challenges, the potential benefits of utilizing RNNs for customer segmentation make them an area of interest for organizations seeking more advanced and accurate 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 KMeans 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] 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.

[8] 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.