BINDURA UNIVERSITY OF SCIENCE EDUCATION

DEPARTMENT OF COMPUTER SCIENCE

APPLICATION OF NATURAL LANGUEGE PROCCESSING IN CLIENT FEEDBACK USING FACIAL SENTIMENTAL ANALYSIS AND OPINION MINING

By

DIVINE RUZAI MURINDAGOMO

REG NUMBER: B202616B

SUPERVISOR: MR H CHIKWIRIRO

A RESEARCH PROJECT SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE BACHELOR OF SCIENCE HONOURS DEGREE IN COMPUTER SCIENCE

2024

APPROVAL FORM

The undersigned certify that they have supervised the student Divine Ruzayi Murindagomo dissertation entitled, “APPLICATION OF NATURAL LANGUEGE PROCCESSING IN CLIENT FEEDBACK USING FACIAL SENTIMENTAL ANALYSIS AND OPINION MINING” submitted in partial fulfillment of the requirements for a Bachelor of Computer Science Honors Degree at Bindura University of Science Education.

STUDENT: **DATE:**

02/10/24

……………………………………………… …………………………..

SUPERVISOR: **DATE:**

02/10/24

……………………………………………… …………………………..

CHAIRPERSON: **DATE:**

02/10/24

……………………………………………… …………………………..

EXTERNAL EXAMINER: **DATE:**

……………………………………………… …………………………..

ABSTRACT

The research topic focuses on creating a machine learning-based customer facial sentimental analysis system for client feedback regarding their level of satisfaction from business services. The goal of this approach is to automate the process of determining emotions in real time, by analyzing the various features of a face such as eyebrows, eyes, mouth, and other features, and mapping them to a set of emotions such as anger, fear, surprise, sadness and happiness. Businesses can adjust their marketing efforts and services that are customer satisfactory by identifying various customer feedback. Various machine learning algorithms, such as Naïve Bayes, linear regression, support vector machine, and deep learning, will be employed in this study to analyze consumer feedback data and customer sentiment into different categories. This research will also look into how various data preprocessing approaches and feature selection methods affect the performance of machine learning models. To ensure the system's performance and industry applicability, it will be tested using real-world customer data from an organization such as a bank. The findings of this study will give companies with an effective tool for client sentimental analysis and high quality products. It will also help enhance machine learning approaches for client sentiment analysis, which can be used in fields other than e-commerce organizations.

Keywords: customer sentiment analysis , machine learning

DEDICATION

This research topic is dedicated to my dear mother, who has always been a source of inspiration, strength, and support for me. Her unrelenting love, commitment, and sacrifices formed me into the person I am today, and I will be eternally grateful for her presence in my life.

ACKNOWLEDGEMENTS

This research would not have been possible without the help of many people. First and foremost, I am grateful for the strength, wisdom, and guidance that allowed me to pursue this topic. I would also like to thank my supervisor, Mr. H. Chikwiriro, for his invaluable guidance, support, and insights throughout the research process. His expertise was instrumental in shaping this project. A huge thank you to the participants who willingly shared their information, their contributions were essential to understanding customer segmentation. Finally, I am grateful to my colleagues and friends for their unwavering support throughout this journey.

Table of Contents

Chapter One: Problem Identification

1.1 Introduction:

Understanding customer sentiment is crucial for businesses to thrive in today's competitive landscape. Client feedback, encompassing reviews, surveys, and support interactions, offers a wealth of information about customer experiences. However, manually analyzing this vast amount of data can be time-consuming and subjective. This is where sentiment analysis and opinion mining, techniques rooted in Natural Language Processing (NLP), come into play (Pang & Lee, 2008).

Sentiment analysis focuses on identifying the overall emotional tone of the feedback, categorizing it as positive, negative, or neutral (Liu, 2012). This allows businesses to gauge customer satisfaction and prioritize areas for improvement. For instance, a surge in negative sentiment might indicate a product defect or a frustrating customer service experience.

Opinion mining, a more nuanced approach, delves deeper to extract specific opinions and the underlying reasons behind them (Feldman et al., 2006). It identifies the aspects of a product or service being discussed (e.g., features, pricing, customer support) and the sentiment associated with each aspect. This granular analysis provides valuable insights for product development, marketing strategies, and customer service training.

1.2 Background of the study

Sentiment analysis, also known as opinion mining, is a subfield of Natural Language Processing (NLP) concerned with extracting and classifying the emotional tone behind text data Lemon, K. N., & Verhoef, P. C. (2016). It has become a valuable tool for businesses to understand customer sentiment towards their products, services, and overall brand perception. This background review will explore the application of sentiment analysis and opinion mining on client feedback, highlighting its benefits and key techniques. Traditionally, understanding customer sentiment relied on manual review of feedback through surveys, emails, or support tickets. However, the growing volume of online reviews, social media comments, and customer interactions necessitates automated solutions. Sentiment analysis bridges this gap by providing a scalable approach to analyze large datasets of textual feedback Pang, B., Lee, L., & Vaithyanathan, S. (2002). There are several techniques are employed for sentiment analysis and opinion mining on client feedback including Lexicon-Based Approach: This method utilizes pre-defined sentiment lexicons containing words with positive, negative, or neutral connotations. The sentiment score of a piece of text is determined by the frequency and polarity of these words. Machine Learning Techniques: Supervised learning algorithms are trained on labeled datasets of text with corresponding sentiment labels (positive, negative, or neutral). These algorithms then classify new, unseen text data based on the learned patterns.

1.3 Problem Statement

While businesses collect vast amounts of client feedback through surveys, emails, and online reviews, manually analyzing this data to understand customer sentiment is time-consuming and inefficient. This becomes particularly challenging with the ever-growing volume of online interactions. There is a critical need for automated solutions to extract and classify the emotional tone behind client feedback. Sentiment analysis and opinion mining techniques offer a promising approach to address this challenge. Building on the foundation provided by Pang & Lee (2008) and Thelwall et al. (2016), this study aims to explore the potential of Sentimental Analysis and Opinion Mining to unlock actionable insights from client feedback

Despite the potential benefits of machine learning in customer sentiment, there is still a need for further exploration of its potential, particularly in the context of detecting accuate This research will investigate the application of machine learning algorithms in customer segmentation, with a focus on their effectiveness in analyzing vast amounts of customer data and identifying patterns that traditional methods may miss. By addressing these challenges, this research can provide valuable insights into how businesses can leverage machine learning to improve customer segmentation and gain a competitive advantage in today's complex business environment.

1.4 Research Aim

To develop a machine learning system for client sentimental analysis and opinion mining that is able to detect accurate facial sentiment.

1.5 Research Objective

1. To analyze different machine learning techniques for client sentimental analysis and opinion mining
2. To develop a machine learning based system for client sentimental analysis and opinion mining
3. To evaluate the use of the machine learning based system for client feedback in client sentimental and opinion mining system

1.6 Research Questions

- What are the most effective machine learning techniques for client sentimental analysis and opinion mining, and how do they compare to traditional methods?
- How can a client sentimental analysis and opinion mining system be built and implemented using machine learning techniques, and what are the benefits and challenges of such a system?
- What is the impact of using machine-learning techniques in client sentimental analysis and opinion mining, and how does it compare to traditional methods?

1.7 Research Hypothesis

$H_{0:}$ Building and implementing a machine learning system for client sentimental analysis and opinion mining will result in a more accurate and efficient segmentation of customers compared to traditional methods.

H_1: Building and implementing a machine learning system for client sentimental analysis and opinion mining will result in inaccurate and inefficient segmentation of customers compared to traditional methods.

1.8 Assumptions

- Sufficient and accurate data will be available for the development and testing of the machine learning system.
- The machine learning system will access to a large and representative dataset of client feedback in textual format (e.g., online reviews, social media comments, support tickets) relevant to the chosen industry domain.

- The selected machine-learning algorithm will be suitable for client sentimental analysis and opinion mining and will perform well on the available data.
- The machine learning system will be able to identify relevant patterns and features in the data to sentimental analysis and opinion mining accurately.
- The machine learning system will be able to generalize well to new, unseen data, ensuring its effectiveness in real-world applications.
- The implementation and integration of the machine learning system into the business workflow will be seamless and without significant technical difficulties.
- The machine learning system will be able to maintain data privacy and security, ensuring compliance with data protection regulations.
- The machine learning system will provide meaningful and actionable insights that can improve the accuracy and efficiency of client sentimental analysis and opinion mining.
- The machine learning system will be cost-effective and provide a positive return on investment for the business.

1.9 Justification/ Significance of the study

The ever-increasing volume of online client feedback presents a significant challenge for businesses seeking to understand customer sentiment. Traditional methods of manual review are becoming increasingly inefficient, necessitating the exploration of automated solutions. Sentiment analysis (SA) and opinion mining (OM) offer promising tools for businesses to glean valuable insights from client feedback data. This study is justified by the potential benefits it offers in understanding customer sentiment, improving customer experience, and informing business decisions Lemon, K. N., & Verhoef, P. C. (2016).

In today's competitive business landscape, understanding customer sentiment is crucial for success. Client sentimental analysis and opinion mining provide a scalable approach to analyze large datasets of textual feedback, enabling businesses to identify trends and patterns in customer opinions. This allows businesses to gain a deeper understanding of customer satisfaction with products and services, identify areas for improvement, and prioritize resources effectively.

By analyzing client feedback, businesses can proactively address negative sentiment and improve customer experience. Identifying recurring themes in negative feedback allows businesses to address customer concerns and rectify issues before they escalate. Furthermore, positive feedback can be used to identify areas where the business is excelling, allowing them to replicate these successes across different products and services.

1.10 Limitations/ Challenges

1. The time needed to implement the functional requirements of the system maybe insufficient.
2. Data quality and availability: The quality and availability of data can significantly affect the performance of the machine learning system. Inaccurate or incomplete data can lead to inaccurate segmentation, while limited data availability can restrict the system's ability to identify patterns.
3. Bias in the data: Bias in the data can lead to biased customer segmentation, which can result in ineffective marketing strategies. Therefore, it is essential to ensure that the machine learning system is trained on unbiased data.
4. Complexity of machine learning algorithms: Machine-learning algorithms can be complex and challenging to implement, requiring specialized knowledge and skills. Therefore, the

development and implementation of the machine learning system may require significant resources and expertise.

5. Ethical considerations: There are ethical considerations regarding the use of machine learning in customer segmentation, such as data privacy and the potential for discrimination. Therefore, it is essential to ensure that the system complies with ethical standards and regulations.
6. Generalization to new data: The machine learning system may perform well on the available data but may not generalize well to new, unseen data. Therefore, it is essential to test the system's performance on new data to ensure its effectiveness in real-world applications.
7. Integration with existing systems: Integrating the machine learning system with existing systems and workflows can be challenging and may require significant modifications to the current infrastructure.
8. User acceptance: The success of the machine learning system depends on user acceptance and adoption. Therefore, it is essential to ensure that the system is user-friendly and meets the needs of its intended users.
9. Time and resources: Building and implementing a machine learning system for customer segmentation can be time-consuming and resource-intensive, requiring significant investment in time and resources. Therefore, it is crucial to ensure that the benefits of the system justify the cost and effort required to develop and implement it.

1.11 Scope of the Research

The research will focus on the development and implementation of a machine learning system for client sentimental analysis and opinion mining using clustering. The research will explore the effectiveness of the machine learning system in giving accurate feedback on clients based on their sentiment and opinions, compared to traditional methods. The research will investigate the potential benefits and challenges of using machine-learning techniques in client sentiment analysis and opinion mining, such as improved accuracy, scalability, and data privacy concerns.

The research will evaluate the generalization ability of the machine learning system to new, unseen data, ensuring its effectiveness in real-world applications.

The research will provide a systematic guide for businesses to develop and implement a machine learning system for customer segmentation.

1.12 Definition of Terms

Machine Learning:

Machine learning (ML) is a field of artificial intelligence (AI) that focuses on developing computer algorithms that can improve automatically through experience and by using data.

Sentimental analysis:

Sentiment analysis, also known as opinion mining, is a subfield of Artificial Intelligence (AI) that deals with understanding the emotional tone behind a piece of text. It focuses on automatically classifying text data as positive, negative, or neutral**.**

Neural Networks:

Neural Networks are a type of machine learning algorithm that is modeled after the structure and function of the human brain. They are used for tasks such as image and speech recognition.

Chapter Two: Literature Review

2.1 Introduction

A literature review is the systematic gathering, organizing, and analysis of papers containing information concerning the study subject under consideration. Its purpose is to create an in-depth understanding of the subject being investigated. It assists the researcher in discovering what other researchers have done in regard to the subject under examination. It helps a researcher minimize unnecessary and unwanted duplication while also giving a framework for understanding study outcomes (Mugenda & Mugenda, 2013). This chapter will look at literature reviews on the use of machine learning in client sentimental analysis and opinion mining.

2.1.0 Customer Segmentation Theory

Customer segmentation theory is a fundamental concept in marketing that entails splitting the customer base into homogeneous groups based on specific factors such as demographics, behavior, or preferences. This method allows organizations to adjust their marketing campaigns to specific client segments, resulting in improved consumer targeting and overall business performance (Kotler et al., 2009).

Machine learning algorithms have shown tremendous potential for improving the accuracy and efficiency of client segmentation. Researchers can construct successful customer segmentation models by merging customer segmentation theory and machine learning techniques, allowing firms to better understand their clients and modify their marketing tactics appropriately (Chen et al., 2020). For example, Shang et al. (2021) used a clustering algorithm and customer segmentation theory to segment customers based on their purchasing behavior and demographic information.

Overall, customer segmentation theory provides a foundation for understanding the importance of splitting the consumer base into homogeneous groups based on unique characteristics. Businesses can create efficient customer segmentation models that correspond with marketing goals and objectives by combining customer segmentation theory with machine learning techniques.

2.1.1 Lexicon-based approach theory

This approach relies on pre-built dictionaries containing words with sentiment polarity (positive, negative, or neutral). The sentiment of a text is determined by analyzing the frequency and strength of these sentiment words.

One way to analyze the meaning of text (semantic analysis) is by using a sentiment dictionary. This dictionary has words labeled as positive, negative, or neutral. By checking how many of these words are in a piece of text, we can figure out the overall feeling (sentiment) of that text. The dictionary itself can be built by hand or by computer programs. The WorldNet dictionary is used by many researchers. First of all, lexicons are found from the whole document and then WorldNet or any other kind of online thesaurus can be used to discover the synonyms and antonyms to expand that dictionary.

Lexicon-based techniques use adjectives and adverbs to discover the semantic orientation of the text. For calculating any text orientation, adjective and adverb combinations are extracted with their sentiment orientation value. These can then be converted to a single score for the whole value.(S. Kannan, ... A. Kejariwal,2016)

2.1.2 Machine Learning

A grasp of machine learning theory is fundamental for effectively applying machine learning techniques to client sentiment analysis and opinion mining. Machine learning is a subfield of artificial intelligence concerned with the development of algorithms and models that enable computers to learn from data and make predictions or judgments autonomously, without explicit programming.

Several studies have used machine-learning approaches to segment customers. For example, Dcreated a deep learning system for client segmentation based on browsing behavior in their study.

Wang et al. (2021) conducted another study in which they used machine-learning algorithms to categorize clients based on their purchase history and demographic data.

Machine learning algorithms act as powerful tools for analyzing massive amounts of customer data, including images. These algorithms excel at finding hidden patterns and connections within this data that would be difficult to detect using traditional statistical methods. This improved ability to understand customer sentiment and opinions allows companies to develop precise and focused marketing strategies that resonate with their customers.

Machine learning theory provides the foundation for developing efficient machine learning models capable of facial sentiment analysis and opinion mining. By leveraging these machine learning algorithms, businesses can enhance their comprehension of consumer behavior and preferences. This leads to improved customer feedback mechanisms, fosters increased customer loyalty, and ultimately contributes to overall business success.

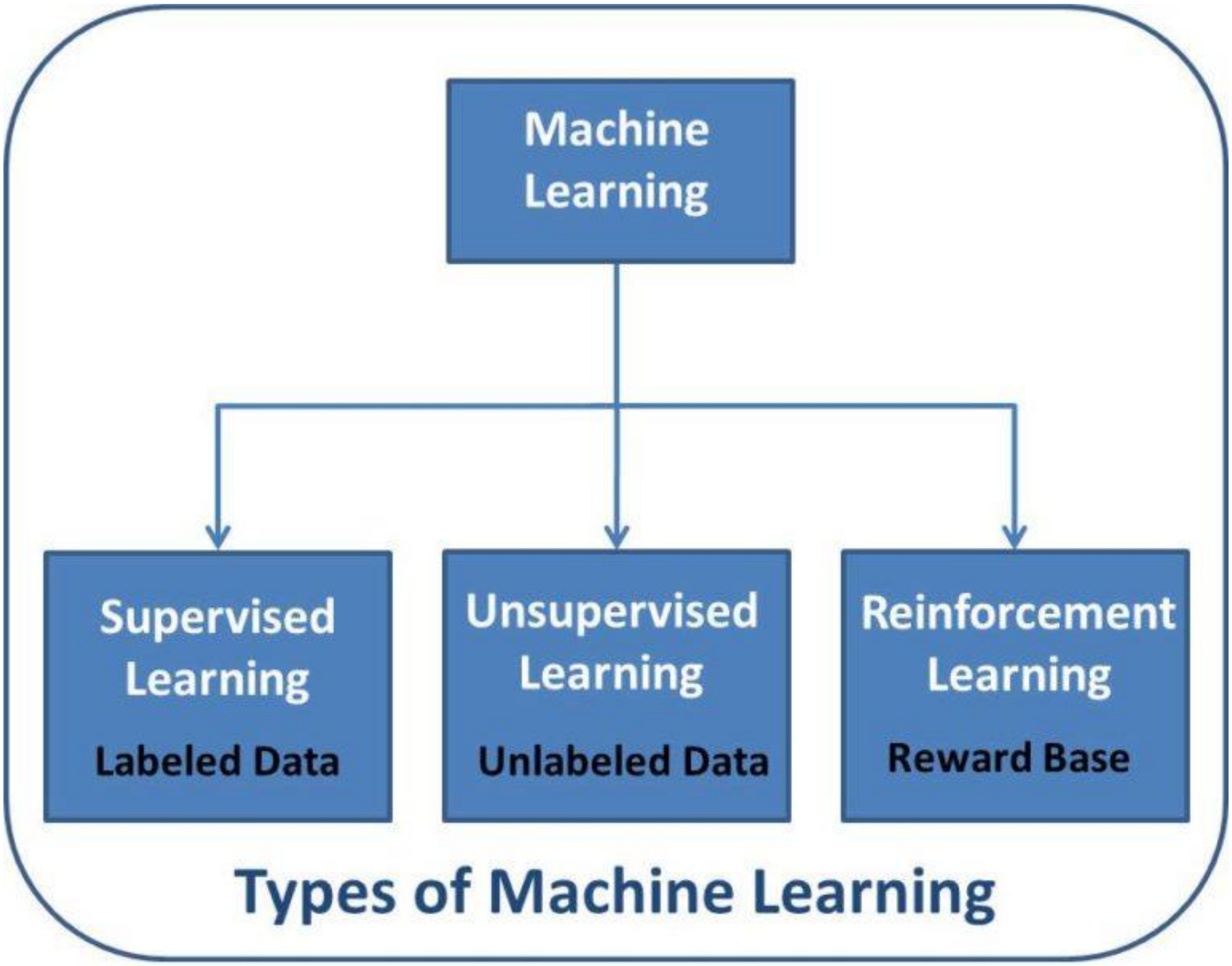

Figure 1: machine learning and types of machine learning

2.1.2.1 Supervised Learning

Supervised machine learning is a subset of machine learning in which the algorithm is trained on labeled data, or data that has been labeled with the right output or target variable. The purpose of supervised learning is to use labeled data to learn the mapping between input variables and output variables, allowing the algorithm to make correct predictions or classifications on fresh, unseen data.

The labeled data in supervised learning is often divided into two sets: a training set and a test set. The training set is used to train the machine-learning algorithm, whereas the test set is used to evaluate the learned model's performance on new, previously unseen data.

Supervised learning can be applied to a wide range of tasks, including regression (predicting a continuous output variable), classification (predicting a categorical output variable), and sequence labeling (predicting a label for each element in a sequence, such as part-of-speech tagging or named entity recognition).

2.1.2.2 Unsupervised Learning

Unsupervised machine learning is a type of machine learning in which the algorithm is taught on unlabeled data, or data that does not have a predefined output or goal variable. Unsupervised learning seeks patterns, structures, or connections in data without understanding what the outcome should be.

The algorithm in unsupervised learning is given a dataset and is tasked with discovering significant patterns or relationships in the data. This is possible through approaches including as clustering, dimensionality reduction, and anomaly detection.

Clustering is a typical unsupervised learning strategy that involves clustering comparable data points based on their similarity. Another strategy is dimensionality reduction, which includes minimizing the number of input variables while maintaining as much of the original data's variability as feasible. Anomaly detection is a technique for recognizing data points that differ considerably from the majority of the data.

Unsupervised learning can be used for a range of tasks, including image and text data compression, detecting fraudulent activity in financial transactions, and finding groups of customers who purchase in similar ways. Unsupervised learning algorithms that are popular include k-means clustering, hierarchical clustering, principal component analysis (PCA) and auto encoders.

2.1.2.3 Reinforcement Learning

Reinforcement machine learning is a type of machine learning in which learning occurs through trial and error interactions with an environment. The algorithm learns to conduct behaviors in an environment in order to maximize a reward signal generated by the environment in reinforcement learning. Reinforcement learning's goal is to learn a policy, which is a function that translates the state of the environment to an action that the agent should perform to maximize the reward signal. The agent interacts with the environment by performing actions that change the state of the environment and getting a reward signal from the environment based on the action performed and the resulting state.

The agent's purpose is to learn a policy that maximizes the cumulative expected reward over time. Techniques such as Q-learning, SARSA, and policy gradient methods can be used to do this.

Reinforcement learning has a wide range of applications, including game play, robotics, and optimization problems. Training an agent to play chess or, teaching a robot to do a task such as grasping an object, and optimizing a control system to reduce energy consumption are all examples of reinforcement learning applications.

Reinforcement learning is a strong approach to machine learning because it allows for learning from experience rather than relying on labeled data. However, it can be more difficult than supervised or unsupervised learning since it requires careful selection of the reward signal and exploration method, as well as dealing with concerns such as credit assignment and balancing exploration with exploitation.

2.1.3 Different Machine Learning Techniques for Customer Segmentation

There are several machine-learning techniques that can be used for customer segmentation, each with its own strengths and weaknesses. Here are some of the most commonly used techniques:

2.1.3.0 K-Means Clustering

K-means clustering is a popular unsupervised machine learning method for categorizing data points into k groups based on similarity. The method aims to reduce the sum of squared distances between data points and their assigned cluster centers, known as centroids. K-means clustering is a computationally efficient clustering method that can handle huge datasets with numerous variables and data points. Because K-means clustering can be scaled to handle high-dimensional data, it is beneficial in a wide range of real-world applications. The initial location of the centroids can affect the efficacy of k-means clustering. The algorithm may converge to suboptimal results if the initial centroids are chosen poorly. K-means clustering is a useful and widely used machine-learning algorithm for clustering data points into groups based on their similarity. However, it has some limitations and may not be suitable for all types of data or all clustering problems.

2.1.3.1 Hierarchical clustering

Hierarchical clustering is a popular unsupervised machine-learning algorithm used for clustering data points into a hierarchy of nested clusters. The algorithm creates a binary tree-like structure called a dendrogram, which represents the relationships between the data points and the clusters they belong to. There are two main types of hierarchical clustering: agglomerative and divisive. Agglomerative hierarchical clustering starts with each data point as a separate cluster and then iteratively merges the most similar pairs of clusters until all data points are in a single cluster. Divisive hierarchical clustering starts with all data points in a single cluster and then iteratively splits the cluster into smaller sub-clusters until each data point is in its own cluster. Hierarchical clustering produces a dendrogram that provides a visual representation of the relationships between the clusters and the data points. Hierarchical clustering is more versatile than k-means clustering in dealing with non-spherical clusters. Hierarchical clustering can be computationally expensive, particularly when dealing with huge datasets. Similarity measurements that can be used to compare data points, such as Euclidean distance or Pearson correlation, are limited in hierarchical clustering. Hierarchical clustering is a popular machine-learning approach for categorizing data points into hierarchical clusters. It has significant limits and may not be appropriate for all types of data or all clustering situations. It is critical to carefully select the suitable type of hierarchical clustering and similarity measure for the given task.

2.1.3.2 Gaussian mixture models (GMMs)

GMMs are a probabilistic clustering approach that is used to describe complex data distributions. GMMs presume that the data points are generated by a Gaussian mixture and utilize the Expectation-Maximization (EM) process to estimate the distribution parameters. GMMs are capable of handling overlapping clusters and perform well on datasets with complex distributions. GMMs can model complex data distributions that conventional clustering algorithms cannot easily model. GMMs require you to specify the number of clusters ahead of time, which can be challenging in some circumstances. Gaussian mixture models are a valuable and powerful clustering tool for modeling complex data distributions and dealing with overlapping clusters. They have inherent limits and may not be appropriate for all types of data or all clustering situations. To ensure that the findings are robust, it is critical to select the proper number of clusters and to execute the algorithm numerous times with different initializations.

2.1.3.3 Decision trees

Decision trees are a popular supervised machine-learning technique for classification and regression tasks. They build a tree-like model of decisions and their potential consequences, with each core node representing a decision based on a given feature and each leaf node representing a class label or a forecast value. Some advantages of decision trees include their simplicity of comprehension and interpretation, their ability to manage both category and numerical data, and their ability to handle missing data and irrelevant qualities. Decision trees are used for feature selection. Decision trees, on the other hand, are prone to overfitting the training data, especially for complex datasets with numerous features and interactions. They can also be sensitive to small changes in the data, generate biased trees if the training data is uneven, and only do axis-parallel splits. To avoid overfitting, choose the hyperactive parameters wisely and regularize the tree. Overall, decision trees are a popular and useful machine-learning paradigm for classification and regression. They can provide decision-making insights and are used in conjunction with other machine learning methodologies such as ensemble methods.

2.1.3.4 Neuro Networks

Neural networks are a form of machine learning model inspired by the structure and function of the human brain. Neural networks are made up of interconnected nodes or artificial neurons that process and transmit information over a number of layers.

Each neuron in a neural network takes input from other neurons, processes that information using an activation function, and then transmits the result to neurons in the next layer. The output of each neuron in a layer serves as input to all neurons in the following layer, providing a hierarchical structure of information processing. Neural networks can be used for a variety of tasks, such as classification, regression, and image and speech recognition. Examples of neural network applications include predicting the likelihood of a customer purchasing a product based on their browsing history, classifying images of objects into different categories, and recognizing speech commands for virtual assistants. Popular types of neural networks include feedforward neural networks, convolutional neural networks (CNNs), and recurrent neural networks (RNNs). Feedforward neural networks are used for tasks such as classification and regression, while CNNs are used for tasks such as image and speech recognition. RNNs are used for tasks such as language modeling and speech recognition, where the input is a sequence of data.

Neural networks have grown in popularity in recent years due to their ability to understand complex patterns and correlations in data, as well as their ability to generalize to new, previously unseen data. However, neural networks can be computationally expensive to train and require vast volumes of data as well as careful tuning of the model's hyper parameters to attain optimal performance.

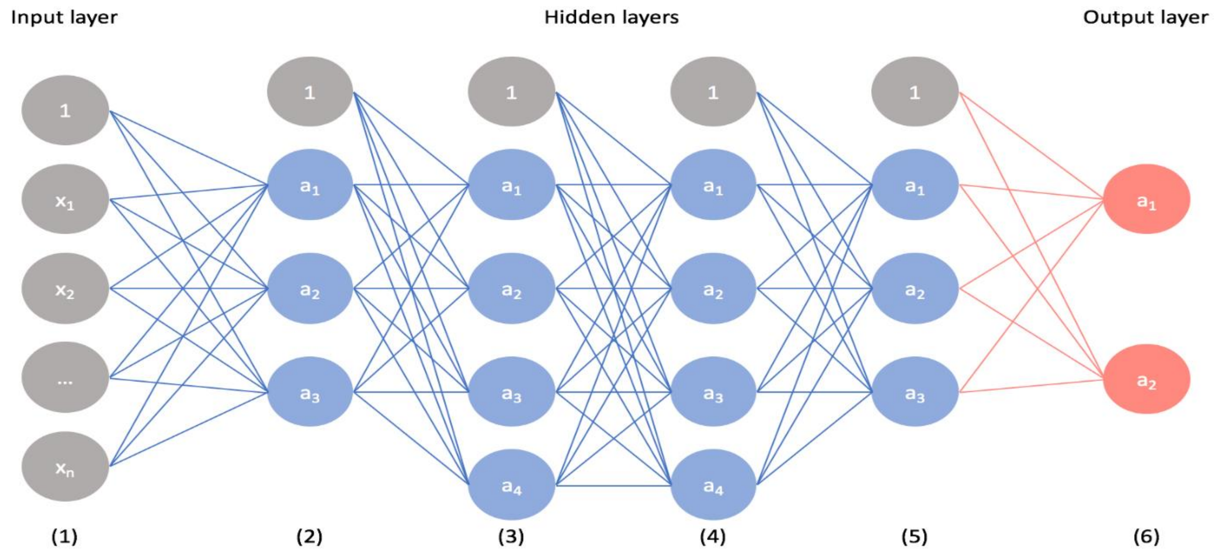

Fig 2: Neuro Networks

2.1.4 Deep Learning Algorithm

Deep learning algorithms are a powerful technique for customer segmentation in the context of machine learning. Deep learning is a kind of machine learning in which artificial neural networks are trained to learn and predict data. A deep learning algorithm is a complicated set of mathematical functions that takes raw data as input and learns characteristics and patterns in the data using numerous layers of interconnected neurons. These algorithms are regarded as "deep" because they often comprise multiple layers of neurons, allowing them to understand more complex relationships in data.

Deep learning algorithms can increase the accuracy and efficiency of customer segmentation by learning complex patterns and relationships in vast datasets. Businesses can establish more precise and focused marketing plans that correspond with client preferences and behavior by using deep learning techniques. Customer segmentation can tremendously benefit e-commerce businesses by allowing them to adjust marketing and sales strategies to each group's unique demands and preferences. Traditional segmentation approaches rely on manual examination of consumer data, which can be time-consuming and error-prone. In recent years, deep learning algorithms have developed as a useful tool for automating and optimizing client segmentation.

In the context of machine learning, deep learning algorithms provide a useful tool for client segmentation. Businesses can improve their understanding of client behavior and preferences by employing deep learning techniques, resulting in better customer targeting, increased customer loyalty, and improved overall business success.

2.1.5 Applied Marketing Theory

Applied marketing theory is the practical application of marketing concepts, frameworks, and principles to real-world business situations. Applied marketing theory focuses on applying current marketing theories and models to specific business difficulties linked to marketing, such as designing effective marketing strategies, finding target markets, and segmenting customers.

Applied marketing theory is a crucial concept in the research topic of applying machine learning in customer segmentation. Marketing theory provides a framework for understanding the concepts and principles underlying customer segmentation and can guide the development and application of machine learning models for customer segmentation.

2.1.5.0 Segmentation, Targeting and Positioning Framework

One important marketing theory related to customer segmentation is the segmentation, targeting, and positioning (STP) framework. STP theory suggests that businesses should divide the market into segments based on customer characteristics and needs, target specific segments with marketing messages and products that meet their needs, and position their products or services to differentiate them from their competitors.

Several studies have applied STP theory to the development of machine learning customer segmentation models. For example, a study by Wang et al. (2021) applied machine learning algorithms to segment customers based on their purchase history and demographic data. The study used STP theory to develop targeted marketing strategies for each customer segment.

2.1.5.1 Customer Lifetime Value Framework

Another important marketing theory related to customer segmentation is the customer lifetime value (CLV) framework. CLV theory suggests that businesses should focus on building long-term relationships with customers and maximizing the value of each customer over their lifetime. Machine-learning customer segmentation models can help businesses identify high-value customers and develop targeted marketing strategies to retain them.

In its overall form, applied marketing theory provides a beneficial foundation for the creation and application of machine learning models for client segmentation. Businesses can create effective machine learning models that correspond with their marketing goals and objectives by adding insights from marketing theory.

2.1.6 Purchase History

Purchase history is a record of a customer's past purchases, including the products or services purchased the frequency and quantity of their purchases, and the dates and locations of their transactions. Purchase history is often saved in a customer database or CRM system and can be utilized to get important insights about consumer behavior and preferences.

Analyzing a consumer's purchasing history can uncover trends and relationships that organizations can utilize to design focused marketing campaigns, improve customer retention, and boost revenue. Businesses, for example, can utilize purchase history data to identify high-value consumers, create tailored marketing messages, and offer products or services that match a customer's previous purchases.

Moreover, purchase history can be used to develop machine-learning algorithms for customer segmentation. By analyzing a customer's purchase history data, businesses can identify different customer segments based on their behavior and preferences. These segments can be used to develop targeted marketing strategies that are tailored to the specific needs of each segment.

In conclusion, purchase history is a significant source of data for companies trying to better their understanding of customer behavior and preferences and design efficient marketing strategies that promote profitability as well as growth.

2.1.7 Machine Learning Models

Machine learning models are algorithms that examine data using statistical approaches to uncover patterns and relationships that may then be used to make predictions or judgments. Machine learning models are trained using datasets, which are groups of data that have been labeled or classified with known outcomes or classifications. Machine-learning models are classified into three types: supervised learning, unsupervised learning, and reinforcement learning. The model is trained on a labeled dataset with each data point accompanied by a known result or classification in supervised learning. The model to learn how to make predictions uses this labeled data or unsupervised learning involves training the model on an unlabeled dataset with no known outcomes or classifications. The model use statistical approaches to find patterns and relationships in the data and to group together comparable data points.

In the context of machine learning, machine-learning models provide a strong tool for client segmentation. Machine learning models can find complicated patterns and relationships that are difficult to spot using traditional statistical methods by examining enormous amounts of data. This enables firms to create more precise and focused marketing plans that are in line with client preferences and behavior.

2.2 Review of Previous Related Researches

In today's data-driven world, businesses are continuously looking for innovative ways to obtain insights into customer behavior and preferences. The use of machine learning algorithms for consumer segmentation is one strategy that has gained popularity in recent years. These algorithms can find patterns and links in enormous amounts of data, allowing organizations to create more focused and effective marketing plans. In this context, a review of previous related research can provide significant insights into the effectiveness and potential of machine learning for consumer segmentation.

The study "Machine Learning for Customer Segmentation: A Review of the Literature" by Tsay, Chen, and Hu (2016) provides a comprehensive review of previous research on machine learning algorithms are used for consumer segmentation. The authors examine a wide range of research that used machine-learning algorithms to segment customers in a variety of scenarios, including e-commerce, retail, and telecoms. The study emphasizes the benefits of utilizing machine learning algorithms for consumer segmentation, such as their capacity to manage enormous datasets, uncover complicated patterns and relationships in data, and help firms design more precise and focused marketing campaigns. The authors also highlight numerous popular machine-learning methods that have been employed in past investigations, including decision trees, neural networks, and clustering algorithms. Furthermore, the study finds numerous significant elements that influence the effectiveness of machine learning algorithms for consumer segmentation, including data quality and quantity, algorithm selection, and validation and assessment methodologies. The author concluded that machine-learning algorithms are an effective tool for consumer segmentation and can help firms get significant insights into customer behavior and preferences. They do, however point out that the use of machine learning in consumer segmentation necessitates careful consideration of issues such as data quality, algorithm selection, and validation and assessment methodologies. The research provides a helpful evaluation of existing research on the use of machine learning algorithms for consumer segmentation and illustrates the possible benefits and limitations of implementing these algorithms in reality.

Another related research was the study "Customer Segmentation Using Machine Learning: A Case Study" by Wang, Chen, and Zhang (2017) presents a case study of the application of machine learning algorithms to customer segmentation in the context of an e-commerce company. The authors use a combination of clustering algorithms and decision trees to segment customers based on their purchase history and demographic data. The study emphasizes the benefits of utilizing machine learning algorithms for customer segmentation, such as their capacity to manage enormous datasets and find complicated patterns and relationships in data. The authors also highlight numerous elements that influence the effectiveness of machine learning algorithms for consumer segmentation, such as data quality and quantity, algorithm selection, and assessment metrics employed. Furthermore, the authors discuss the practical obstacles of using machine-learning algorithms for client segmentation in a corporate context. For example, they address the significance of data scientists and business stakeholders communicating well. According to the study, machine-learning algorithms are a valuable tool for client segmentation that may help organizations design more targeted and effective marketing strategies. The authors do, however, point out that using machine learning for consumer segmentation necessitates careful consideration of issues such as data quality, algorithm selection, and business objectives. Overall, the study is an excellent case study of the application of machine learning algorithms to customer segmentation in a business setting. The authors share insights into how firms can effectively employ machine learning for consumer segmentation by outlining the possible benefits and challenges of adopting these algorithms.

Another previous research was the study "A Survey of Machine Learning Methods for Customer Segmentation" by Zhang, Li, and Wang (2018) provides a comprehensive review of the application of machine learning algorithms for customer segmentation. The authors analyze a wide range of studies that have used machine-learning algorithms for customer segmentation in various contexts, including e-commerce, retail, and telecommunications. The study emphasizes the benefits of utilizing machine learning algorithms for consumer segmentation, such as their capacity to manage enormous datasets, find complicated patterns and relationships in data, and enable firms to design more precise and focused marketing campaigns. The authors also list numerous common machine-learning techniques utilized in earlier investigations, such as clustering, decision trees, and neural networks. Furthermore, the study finds numerous significant elements that influence the efficiency of machine learning algorithms for consumer segmentation, such as data quality and quantity, algorithm selection, and evaluation metrics. The authors also explore the significance of interpretability and explainability in machine learning algorithms, which can assist organizations in understanding the underlying causes that influence customer behavior. According to the study, machine-learning algorithms are a powerful tool for consumer segmentation and can give organizations with significant insights into customer behavior and preferences. The authors do, however, point out that using machine learning for consumer segmentation necessitates careful consideration of issues such as data quality, algorithm selection, and interpretability. The article provides a useful summary of prior research on the use of machine learning algorithms for consumer segmentation, highlighting the possible benefits and limitations of implementing these algorithms in reality.

In their 2018 paper titled "Machine Learning for Customer Segmentation: A Case Study of an Online Retailer," Sarah Rieger and Peter Buxmann developed a customer segmentation system using clustering algorithms for an online retailer. The study aimed to improve the effectiveness of the retailer's marketing campaigns by identifying groups of customers with similar purchasing behavior and preferences. The authors gathered customer information from the retailer's website, including transactional information, demographic information, and online usage information. They then used a clustering method (k-means) to categorize the clients based on their purchasing habits. The authors assessed the efficiency of the segmentation approach by comparing the response rates to marketing campaigns of different client categories. According to the findings, the customer segmentation method was helpful in increasing response rates to marketing initiatives. Based on their shopping habits, the authors identified numerous distinct client groupings, including frequent buyers, bargain hunters, and customers who exclusively purchased high-value items. The authors also discovered that the consumer categories defined by the clustering algorithm were more effective in forecasting customer behavior than traditional demographic-based segmentation methods. The study sheds light on the use of machine learning algorithms for customer segmentation and illustrates the potential benefits of utilizing these algorithms to identify various consumer segments based on purchase behavior. The study also emphasizes the need of employing effective evaluation metrics to assess the segmentation system's success in enhancing marketing campaign outcomes. Rieger and Buxmann's paper is a valuable case study of the development of a machine learning-based customer segmentation system that demonstrates the potential benefits of employing these algorithms to better marketing strategies for online merchants.

In their 2016 paper titled "Customer Segmentation using Machine Learning Algorithms for Retail Industry," Kunal Patel and Ravi Jain developed a customer segmentation system using clustering algorithms for a retail industry case study. The study aimed to identify distinct customer segments based on purchasing behavior and demographics in order to improve the effectiveness of marketing campaigns. The authors gathered customer information from a retail store, including transactional and demographic information. They then used k-means, hierarchical clustering, and density-based spatial clustering of applications with noise (DBSCAN) to divide the customers into various groups based on their purchase behavior and demographics. The authors assessed the efficiency of the segmentation approach by comparing the response rates to marketing campaigns of different client categories. According to the findings, the customer segmentation method was helpful in increasing response rates to marketing initiatives. Based on their shopping habits and demographics, the authors identified numerous distinct client groupings, including frequent buyers, bargain hunters, and customers who only purchased high-value items.

The authors also discovered that the customer categories generated by the clustering algorithm were more effective than standard segmentation methods based solely on demographics in forecasting customer behavior. The research sheds light on the use of machine learning algorithms for customer segmentation in the retail business, emphasizing the potential benefits of utilizing these algorithms to identify separate consumer segments based on purchasing behavior and demographics. The study also emphasizes the need of employing effective evaluation metrics to assess the segmentation system's success in enhancing marketing campaign outcomes. Ultimately, Patel and Jain's research presents a useful case study of the development of a machine learning-based customer segmentation system for the retail business. The study emphasizes the potential benefits of utilizing clustering algorithms to identify separate client categories based on purchase behavior and demographic data. The scientists also identified numerous elements that influence the effectiveness of the segmentation system, including data quality and quantity, algorithm selection, and evaluation metrics. By addressing these characteristics, firms may create more effective and focused marketing campaigns that are tailored to their clients' requirements and tastes.

In a 2021 article titled "Customer Segmentation and Targeting Using Machine Learning Techniques: A Case Study," Hichem Omrani, Mohamed Kassab, and Sami Faiz developed a customer segmentation system using machine learning techniques for a telecommunications company. The study aimed to improve the effectiveness of the company's marketing campaigns by identifying groups of customers with similar behavior and preferences. The writers gathered consumer information from the company's database, including demographic information, call detail records, and service usage information. The researchers then used machine learning methods such as decision trees, random forests, and support vector machines (SVM) to divide the clients into various groups based on their behavior and interests. The authors assessed the efficiency of the segmentation approach by comparing the response rates to marketing campaigns of different client categories. According to the findings, the customer segmentation method was helpful in increasing response rates to marketing initiatives. Based on their behavior and preferences, the authors identified numerous separate client groupings, including high-value consumers, low-value customers, and customers with significant service utilization.

The authors also discovered that customer segments identified by machine learning algorithms were more effective than traditional segmentation methods based solely on demographic data in forecasting customer behavior. The study sheds light on the use of machine learning algorithms for customer segmentation in the telecommunications industry, emphasizing the potential benefits of utilizing these algorithms to identify various consumer segments based on behavior and preferences. The study also emphasizes the significance of employing good evaluation metrics to assess the success of the segmentation method in enhancing marketing campaign outcomes.Omrani, Kassab, and Faiz's research presents a useful case study of the development of a machine learning-based consumer segmentation system for the telecoms business. The study emphasizes the potential benefits of combining demographic data, call detail records, and service usage data to identify distinct consumer categories utilizing decision trees, random forests, and SVM algorithms. The scientists also identified numerous elements that influence the effectiveness of the segmentation system, including data quality and quantity, algorithm selection, and evaluation metrics. By addressing these issues, organizations in the telecoms industry can design more successful and focused marketing strategies that are tailored to the demands and preferences of their customers.

In a 2020 publication titled "Customer Segmentation and Targeting using Machine Learning: A Case Study of a Fashion E-commerce Company," Hui-Ju Tsai and Shu-Yuan Pan developed a customer segmentation system using machine learning techniques for a fashion e-commerce company. The writers gathered client information from the company's website, including transactional information, browsing history, and demographic information. They then used k-means clustering, principal component analysis (PCA), and decision trees to divide the customers into various groups based on their purchase behavior and preferences. The authors assessed the efficiency of the segmentation approach by comparing the response rates to marketing campaigns of different client categories. The authors discovered that the machine learning-based consumer segmentation method improved response rates to marketing initiatives. Based on their purchasing behavior and preferences, they discovered numerous separate client groupings, such as high-value customers, bargain hunters, and people who only purchased specific types of things. The authors also discovered that customer segments identified by machine learning algorithms were more effective than traditional segmentation methods based solely on demographic data in forecasting customer behavior.

The study sheds light on the use of machine learning algorithms for consumer segmentation in the fashion e-commerce industry, as well as the potential benefits of utilizing these algorithms to identify various client segments based on purchasing behavior and preferences. The study also emphasizes the need of employing effective evaluation metrics to assess the segmentation system's success in enhancing marketing campaign outcomes. Tsai and Pan's research presents a useful case study of the development of a machine learning-based client segmentation system for the fashion e-commerce market. The authors' approach of combining transactional data, browsing history, and demographic data to identify customer segments provides a more comprehensive view of customer behavior and preferences than traditional segmentation methods based on demographic data alone. The authors also used various machine learning algorithms, including k-means clustering, PCA, and decision trees, to identify distinct customer segments, which demonstrates the flexibility of machine learning techniques in developing customer segmentation systems. The study, however, had certain drawbacks. The authors did not compare their machine learning-based client segmentation system to classic demographic-based segmentation approaches. Furthermore, the study did not include information on how the marketing campaigns were implemented or how the segmentation method was integrated into the company's marketing strategy. Future research could overcome these limitations by comparing machine learning-based segmentation systems to traditional segmentation methods and providing more information on marketing campaign implementation. The research by Tsai and Pan contributes to the growing body of research on the development of machine learning-based customer segmentation systems and provides valuable insights for businesses in the fashion e-commerce industry looking to improve the effectiveness of their marketing campaigns.

In their 2019 paper titled "Customer Segmentation in E-commerce: A Comparison of Machine Learning Techniques," Wenjie Zhang, Jing Jiang, and Xiaoying Bai developed a customer segmentation system using machine learning techniques for an e-commerce company. The study aimed to identify distinct customer segments based on their purchasing behavior and preferences in order to improve the effectiveness of marketing campaigns. The authors gathered consumer information from the company's website, including transactional and demographic information. They then used k-means clustering, support vector machines (SVM), and neural networks to divide the clients into various groups based on their purchase behavior and preferences. The authors assessed the efficiency of the segmentation approach by comparing the response rates to marketing campaigns of different client categories. The results demonstrated that the consumer segmentation method was effective in increasing response rates to marketing initiatives. The authors found numerous separate client segments based on their purchasing behavior and preferences, such as high-value customers, bargain hunters, and customers who only purchased specific sorts of things. The authors also discovered that customer segments identified by machine learning algorithms were more efficient in forecasting customer behavior than traditional segmentation methods based solely on demographic data. The study sheds light on the use of machine learning algorithms for customer segmentation in the e-commerce industry, emphasizing the potential benefits of employing these algorithms to identify separate client segments based on purchasing behavior and preferences. The study also emphasizes the significance of employing good evaluation metrics to assess the success of the segmentation method in enhancing marketing campaign outcomes. Zhang, Jiang, and Bai's research presents a useful comparison of several machine-learning methods for client segmentation in the e-commerce market. The authors provide insights into the merits and disadvantages of different algorithms for identifying separate client groups by comparing the performance of k-means clustering, SVM, and neural networks. The authors also emphasize the necessity of feature selection and data preparation in boosting the segmentation system's efficiency. The study, however, had certain drawbacks. The writers did not go into depth about how the marketing campaigns were implemented or how the segmentation method was integrated into the company's marketing strategy. Furthermore, the research did not compare the machine learning-based segmentation system to classic segmentation approaches based solely on demographic data. Future research could overcome these limitations by giving more information about the marketing campaigns' implementation and comparing machine learning-based segmentation systems to traditional segmentation approaches.

In their 2018 paper titled "Customer Segmentation for E-commerce: An Overview," Chaoqun Ma and Yong Li analyzed the progress of consumer segmentation systems in the e-commerce business using machine learning techniques. The study's goal was to provide an overview of the various machine-learning algorithms used for customer segmentation, as well as to highlight their strengths and limitations. The authors found various machine-learning methods commonly employed in the e-commerce business for client segmentation, including k-means clustering, hierarchical clustering, and decision trees. They also emphasized the significance of feature selection and data preparation in boosting the segmentation system's effectiveness. The authors emphasized the potential benefits of applying machine learning algorithms for consumer segmentation, such as identifying separate client segments based on purchasing behavior and preferences. They also emphasized the significance of analyzing the efficiency of the segmentation system using relevant measures, such as marketing campaign response rates. Overall, Ma and Li's research gives an insightful look at the evolution of machine learning-based client segmentation systems in the e-commerce business. The authors' evaluation of several machine learning algorithms, as well as their merits and limitations, serves as a useful reference for organizations in picking the best algorithm for their individual needs and datasets. The authors also underline the significance of feature selection and data preprocessing in boosting the performance of the segmentation system, which can assist firms in optimizing marketing campaigns and improving consumer happiness. The study, however, did not provide a specific case study or empirical conclusions for a certain business or industry. Future research could overcome this restriction by presenting more extensive case studies or empirical results demonstrating the efficacy of machine learning-based client segmentation systems in the e-commerce business. Furthermore, the study did not address the potential ethical issues raised by the use of customer data for segmentation reasons. Future research could address these concerns and provide firms with recommendations on how to create customer segmentation systems that are clear, fair, and considerate of consumer privacy. Overall, Ma and Li's research gives a comprehensive overview of the development of machine learning-based client segmentation systems in the e-commerce business, highlighting both the potential benefits and problems of using these systems for improving marketing strategies.

In their 2021 paper titled "Customer Segmentation and Targeting in E-commerce Using Machine Learning: A Case Study," Shuting Xu, Weili Wu, and Jie Li developed a customer segmentation system using machine learning techniques for an e-commerce company. The study aimed to identify distinct customer segments based on their purchasing behavior and preferences in order to improve the effectiveness of marketing campaigns. The authors gathered consumer information from the company's website, including transactional and demographic information. They then used a variety of machine learning algorithms, such as k-means clustering, hierarchical clustering, and DBSCAN, to divide the customers into various groups based on their purchasing habits and preferences. The authors assessed the efficiency of the segmentation approach by comparing the response rates to marketing campaigns of different client categories. The results demonstrated that the machine learning-based consumer segmentation system was helpful in increasing response rates to marketing efforts. The authors found numerous separate client segments based on their purchasing behavior and preferences, such as high-value customers, bargain hunters, and customers who only purchased specific sorts of things. The authors also discovered that customer segments identified by machine learning algorithms were more efficient in forecasting customer behavior than traditional segmentation methods based solely on demographic data. The study sheds light on the use of machine learning algorithms for customer segmentation in the e-commerce industry, emphasizing the potential benefits of employing these algorithms to identify separate client segments based on purchasing behavior and preferences. The study also emphasizes the significance of employing good evaluation metrics to assess the success of the segmentation method in enhancing marketing campaign outcomes. Overall, Xu, Wu, and Li's research presents a valuable case study of the development of a machine learning-based client segmentation system for the e-commerce business. The authors' method of identifying consumer categories by combining transactional and demographic data provides a more comprehensive view of customer behavior and preferences than standard segmentation methods based solely on demographic data.

The authors also employed k-means clustering, hierarchical clustering, and DBSCAN to discover separate customer groups, demonstrating the versatility of machine learning techniques in constructing consumer segmentation systems. The study, however, had certain drawbacks. The authors did not compare their machine learning-based client segmentation system to classic demographic-based segmentation approaches. Furthermore, the study did not include information on how the marketing campaigns were implemented or how the segmentation method was integrated into the company's marketing strategy. Future research could overcome these limitations by comparing machine learning-based segmentation systems to traditional segmentation methods and providing more information on marketing campaign implementation. Overall, the work by Xu, Wu, and Li adds to the expanding body of research on the development of machine learning-based consumer segmentation systems and gives useful insights for e-commerce companies trying to increase the success of their marketing campaigns. The authors' evaluation of multiple machine learning algorithms might help organizations choose the best algorithm for their specific goals and datasets.

2.3 Research Gap

This chapter typically includes a review of relevant literature, a summary of the key findings and limitations of existing studies, and a clear statement of the research question and objectives. By highlighting the limitations and gaps in the existing literature, the research gap chapter helps to establish the need for the proposed study and its potential contribution to the field. There have been several studies on the use of machine learning in customer segmentation but there is still a research gap in implementing a machine learning customer segmentation system. In 2021 Shuting Xu, Weili Wu, and Jie Li developed a customer segmentation system using machine learning techniques for an e-commerce company by dividing the customers into various groups based on their purchasing habits and preferences. This study was a success but also faced a problem when there was introduction of new customers, the system would have to be trained again on the new customers' data. Therefore, the author focuses on creating a machine learning customer segmentation system that divides the customers based on their demographics and with the ability of continuity when new customers are introduced. There would be no need for the system to be trained on the new data again.

2.4 Summary

The researcher was successful in compiling data that is pertinent to this study. Some of the notions from academic journals, the internet, textbooks, and unpublished information reveal the depth and gap that needs to be filled. The information acquired will be applied in the ensuing chapters to achieve the predetermined goals. The most popular methods for applying machine learning to customer segmentation system are k-means clustering, hierarchical cluster, Gaussian mixture model, decision trees and neural networks. Some studies indicate accuracy rates above 90% for these approaches, which have attained remarkable accuracy rates. However, additional study is still required to improve machine-learning algorithms in segmenting customers to help business develop effective business strategies and improve customer retention.

Chapter Three: Research Methodology

3.0 Introduction

Research methodology breaks down the steps researchers take to conduct a study in a systematic and organized way. This helps them answer research questions and solve problems

This process entails a structured collection of techniques and procedures for gathering, analyzing, and interpreting data. This ultimately aims to answer research questions or evaluate hypotheses .In addition, this chapter aims to define the target users (population) of the proposed system.

3.1Data Collection and Analysis

This study focuses on data collection, the process of acquiring information relevant to the research question or hypothesis. Data collection involves gathering information from various database sources that include CK+, JAFFE, RaFD, MMI, FER2013, TFD, and COHN-KANADE AU.In this research the author is going to use the COHN-KANADE AU and JAFFE datasets.

The COHN-KANADE AU-Coded Facial Expression Database is a popular resource for evaluating facial expression recognition algorithms. It includes images of 100 university students (aged 18-30) displaying various emotions. The group is mostly female (65%) with some racial diversity (15% African-American, 3% Asian/Latino). Researchers instructed participants to make specific facial expressions, capturing the change from a neutral face to the target expression. These image sequences are digitized into high-resolution (640x480/490 pixels) grayscale images. The database also provides instructions on how to view the images in the correct order. It focuses on recognizing four emotions: anger, neutral, happiness, and sadness (although Fig.1 only shows 4 examples)

Happy Sad Angry Neutral

Figure 1: The four expression from one subject

Another public database for facial expression recognition is the Japanese Female Facial Expressions (JAFFE) database. It's smaller, containing only 213 images total. Ten different Japanese female subjects participated, each displaying seven facial expressions (angry, happy, disgust, sadness, surprise, fear, and neutral). This means each subject has around 20 images, with

2-3 images capturing variations of each expression. Figure 2 showcases all four expressions from a single subject in the database

Figure 2.1: The four expression from one subject

The data will be cleaned and transformed to suit machine learning algorithms. The following machine learning algorithms will be applied:

1. Support Vector Machine

3.1.0 Support Vector Machine algorithm

Support Vector Machines (SVMs) are a powerful type of machine learning algorithm known for their effectiveness in various tasks, including classification and regression. SVMs aim to find the optimal line or hyperplane (in higher dimensions) that separates data points belonging to different classes with the maximum margin. This margin refers to the distance between the hyperplane and the closest data points from each class, called support vectors

The data analysis process will be done using Dart programming language with tools such as Visual Studio Code, Android Studio, Tensorflow-learn libraries.

3.1.1 Ethical Considerations:

This study is committed to ethical research practices. To protect participant privacy, all data will be anonymized, and no personal information will be collected. We will also obtain informed consent from participants and ensure the confidentiality of their data and privacy throughout the study.

3.1.2 Limitations:

The reliance on secondary data in this study introduces the potential for inherent errors or limitations within the collected data itself.

3.2 Research Design

Every research project requires a well-defined research design. This design serves as a blueprint, outlining the methodologies and processes for data collection and analysis. A robust research design ensures the study is conducted in a methodical and rigorous manner, ultimately leading to legitimate and dependable results

Creswell (2014) identifies three fundamental research design categories: quantitative, qualitative, and mixed methods. Quantitative research designs typically collect numerical data through structured approaches such as surveys and experiments. In contrast, qualitative research designs gather non-numerical data through unstructured methods like interviews and observations. Mixed methods research designs integrate both quantitative and qualitative elements within a single study.

3.2.1 Requirements Analysis

The research design acts as the foundation for a study, outlining the overall strategy to address the research problem effectively. It ensures a logical and cohesive approach by integrating all the different elements of the research (Polit & Beck, 2012). In simpler terms, it's the blueprint for how researchers will collect data, measure it, and analyze it to answer their research questions.

This stage emphasizes the documentation of all system requirements, encompassing both functional and non-functional aspects. To achieve uniformity and avoid ambiguity, the collected requirements will undergo a rigorous process of analysis, revision, and verification. The research also took into account potential constraints, such as data availability, that could hinder the design process. Consequently, it is recommended to gather all research data, subject it to a thorough assessment, and examine any potential user-related limitations that may arise during implementation.

3.2.1.1 Functional Requirements

- The system should be able to use machine-learning algorithm to give accurate sentiment analysis of client feedback based on facial expressions.
- The system should be able to provide visualizations and the results of the client feedback in real time on a display
- The system should be able to collect and integrate data from various sources such as customer databases, social media.

3.2.1.2 Non-Functional Requirements

- The system's outputs must be characterized by high accuracy and reliability, minimizing the occurrence of errors and inconsistencies
- The design and implementation of the system should prioritize ethical considerations. These include mitigating bias within the machine learning algorithms, guaranteeing transparency in decision-making processes, and safeguarding customer privacy and rights.
- To best serve users, the system should be adaptable to changing business demands and data sources. This adaptability is achieved through a flexible and configurable design, allowing users to add or modify functionalities as needed.
- To expedite the decision-making process, the system must deliver rapid response times.

3.2.1.3 Software Requirements

- Windows 10/11 operating system
- VS Code
- Flutter
- Android Studio
- Java Development KIt

3.2.1.4 Hardware Requirements

- Minimum Core i5 CPU
- Keyboard
- Mouse
- Monitor

3.2.1.5 Design tools

- Dart – The researcher chose to use dart for its vast number of tools that are meant for machine learning. According to (Google Inc., 2020), Dart provides a well-rounded set of features that make it suitable for developing various applications, from web and mobile to potentially server-side development. Its focus on readability, portability, and developer productivity has made it a popular choice for modern application development.
- Flutter- Flutter is an open-source framework created by Google for building beautiful, high-performance mobile applications (iOS and Android), web applications, desktop

applications (Windows, macOS, Linux), and even embedded systems using a single codebase

- VS Code IDE- Visual Studio Code, also commonly referred to as VS Code, is a source-code editor developed by Microsoft for Windows, Linux, macOS and web browsers. Features include support for debugging, syntax highlighting, intelligent code completion, snippets, code refactoring, and embedded version control with Git.
- Android Studio - Android Studio is the official Integrated Development Environment (IDE) for Android app development, specifically designed for this purpose. It also has android virtual machines in it, so that developers can debug applications on the interface.

3.3 System Development

This vital chapter delves into the development process of a machine learning system for customer segmentation. It details the design, implementation, and testing phases, along with the specific tools and technologies employed throughout**.**

3.3.1 System Development Tools

Software development methodologies act as roadmaps for building information systems. These frameworks guide the planning, organization, and control of the entire development process. Researchers have created various methodologies, each with advantages and disadvantages depending on the project. Examples include the agile model, Spiral model, and Progressive (prototyping) model.

For this particular project, the author selected the agile model due to its streamlined approach. This choice aligns well with the project's small size and strict deadline. The Agile model is ideal when project requirements are well-defined and all necessary tools

3.3.2 Agile Methodology

Agile methodology is built on two pillars: collaboration and time management. Agile model divides a large software development project into individual deliverable pieces rather than creating a timeline for the entire project. These 'time-boxed' phases are known as 'sprints,' and they last only a few weeks. Following the completion of each sprint, the feedback from the previous phase is used to plan the next one. It involves the following steps

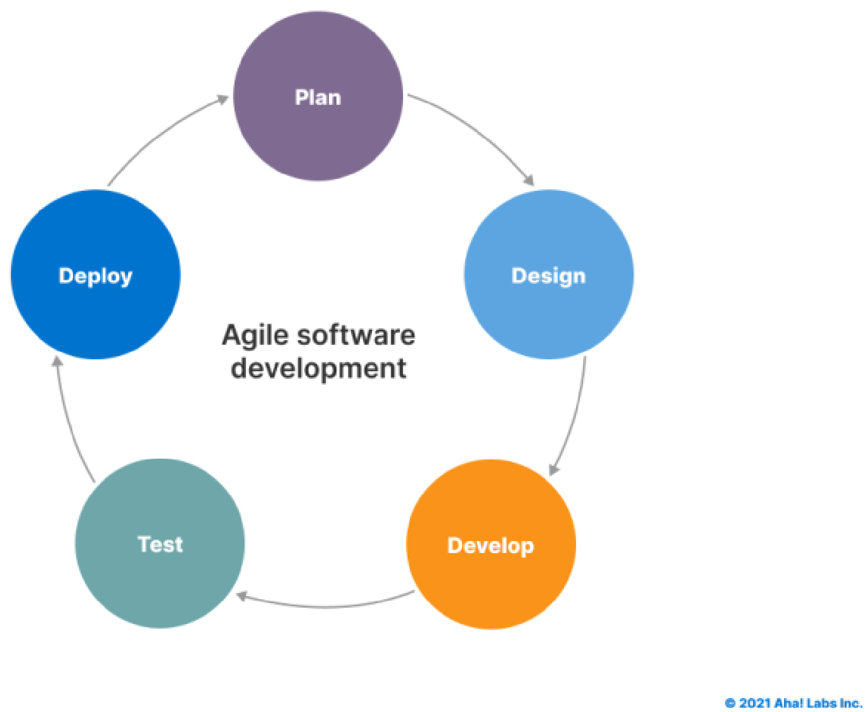

Figure 3.8 Agile Methodology Lifecycle

Define User Stories: Break down the overall goal of sentiment analysis into small, user-focused stories. For example, a user story might be "As a social media manager, I want the system to identify positive sentiment in customer reviews so I can understand customer satisfaction.

Prioritize Backlog: Create a prioritized list of all the features and functionalities you want in your sentiment analysis system. This backlog can be continuously updated as new information arises.

Sprint Planning: Break down the backlog into smaller chunks of work called sprints. Each sprint should have a specific set of goals and a defined timeframe (typically 1-4 weeks).

Develop and Test: During the sprint, the development team focuses on building and testing the functionalities identified for that specific sprint.

Review and Deployment: At the end of each sprint, there's a review to assess progress and a retrospective to identify areas for improvement in the next sprint

3.4 Summary on How The System Works

The system starts by gathering facial image data. This data might come from pre-recorded videos, live camera feeds, or image datasets. Preprocessing involves cleaning the data by addressing issues like lighting variations, occlusions (e.g., glasses), and image noise. Key features are extracted from the preprocessed images that represent facial expressions. These features might include eye movements, mouth shapes (e.g., smiling, frowning) and forehead wrinkles. A labeled training dataset is used where each image has a corresponding emotional label (e.g., happy, sad, angry, neutral).The SVM algorithm analyzes these labeled examples and learns to identify patterns that distinguish between different emotions based on the extracted features. Once trained, the SVM model can classify the emotions in new, unseen images. The model receives features extracted from a new image and predicts the most likely emotional category (e.g., happy) based on the patterns it learned from the training data.

3.5 System Design

An examination of the requirements specification document is conducted. This step serves to demonstrate, for the subsequent stage, how the system's components and data align with the established requirements. In doing so, the system's overall coordination and cohesion are made evident.

3.5.1 Proposed Android System Design

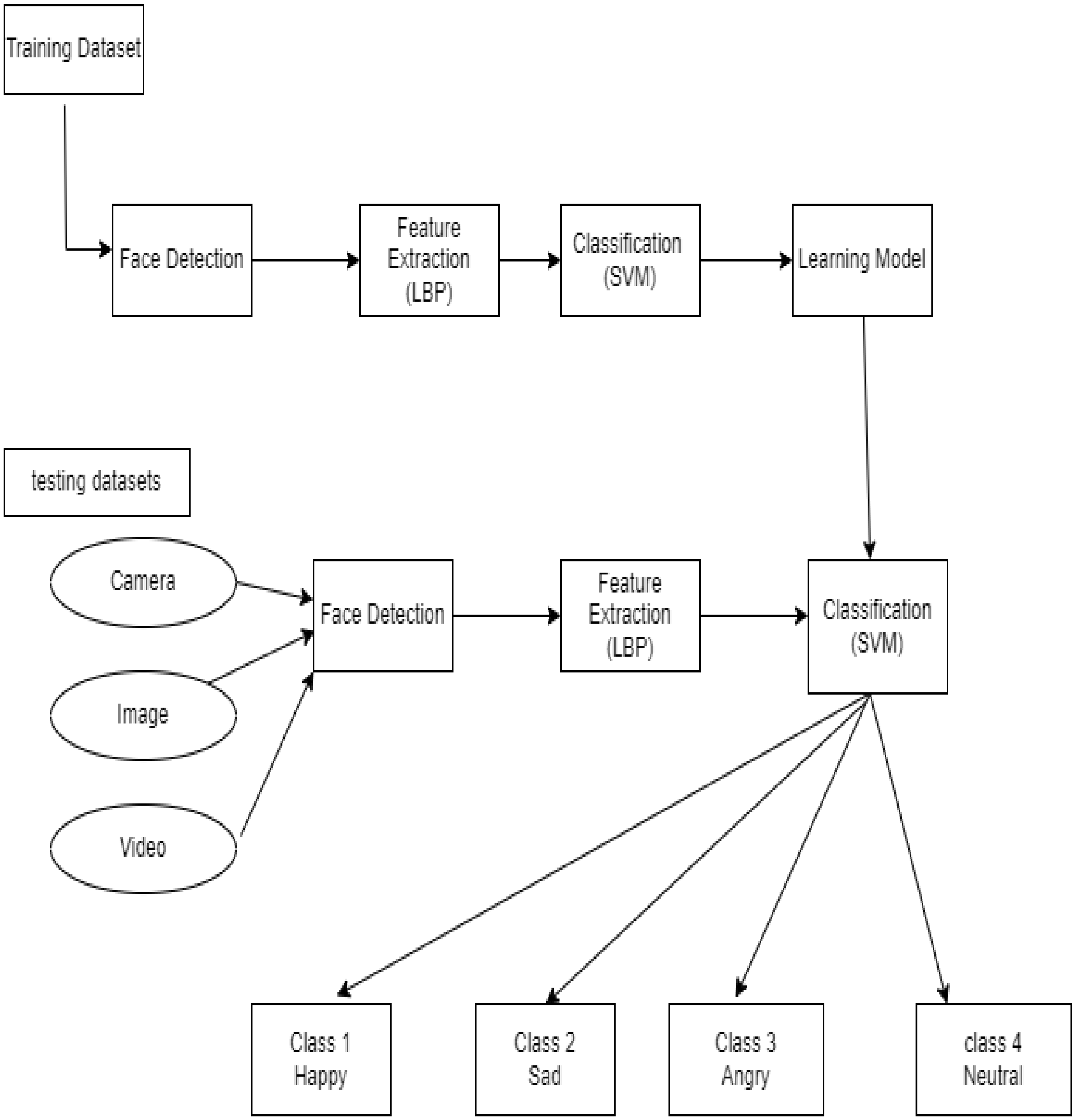

Figure 3.1 Android System Diagram

The diagram below is a block diagram of a face detection system. It shows the two main stages of face detection: training and testing.

Training Stage: The system starts with a training dataset of images that contain faces. These images are labelled with the location of each face in the image. A learning model, such as a Support Vector Machine (SVM), is then trained on this data. The SVM learns to identify the features of a face, such as the eyes, nose, and mouth.

Testing Stage: Once the SVM is trained, it can be used to detect faces in new images. In the testing stage, an image or video is captured by a camera. The image is then preprocessed to extract features, such as Local Binary Patterns (LBP).The extracted features are then fed into the SVM classifier. The SVM classifier then determines whether the image contains a face or not. If a face is detected, the system may also try to classify the person's expression, such as happy, sad, angry, or neutral.

3.5.1.0 Android System Flow Diagram

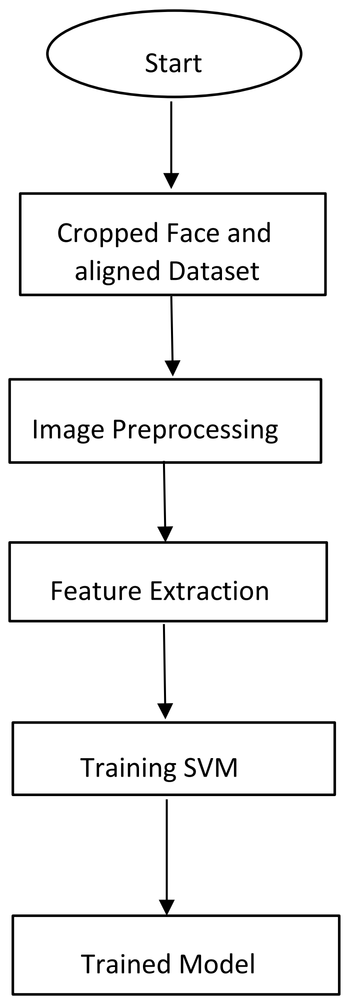

Figure 3.2: Flowchart of Training

The diagram above is a flowchart of training a Support Vector Machine (SVM) for a face detection system. Here's a breakdown of the process:

This is the beginning of the training process.

Cropped Face and Aligned Dataset: In this step, a collection of images containing faces is prepared. The faces are cropped from the images and then aligned in a standard pose, ensuring all faces are oriented the same way.

Image Preprocessing: The faces are preprocessed to improve the quality of the data for training the SVM. This may involve techniques like scaling the images to a standard size, adjusting lighting variations, and converting them to grayscale.

Feature Extraction: Features are extracted from the preprocessed images. These features capture the important characteristics of a face, such as the edges around the eyes, nose, and mouth. There are many different feature extraction techniques that can be used.

Training SVM: This is the core step in the process. The SVM is a machine learning algorithm that learns to distinguish between faces and non-faces based on the extracted features. The training process involves repeatedly showing the SVM examples of faces and non-faces, and adjusting the SVM's internal parameters until it can accurately distinguish between the two. The diagram shows multiple training steps (four in this case), which can improve the accuracy of the model.

3.5.1.1 Trained model

Once the training is complete, the result is a trained SVM model that can be used to classify new images as containing a face or not.

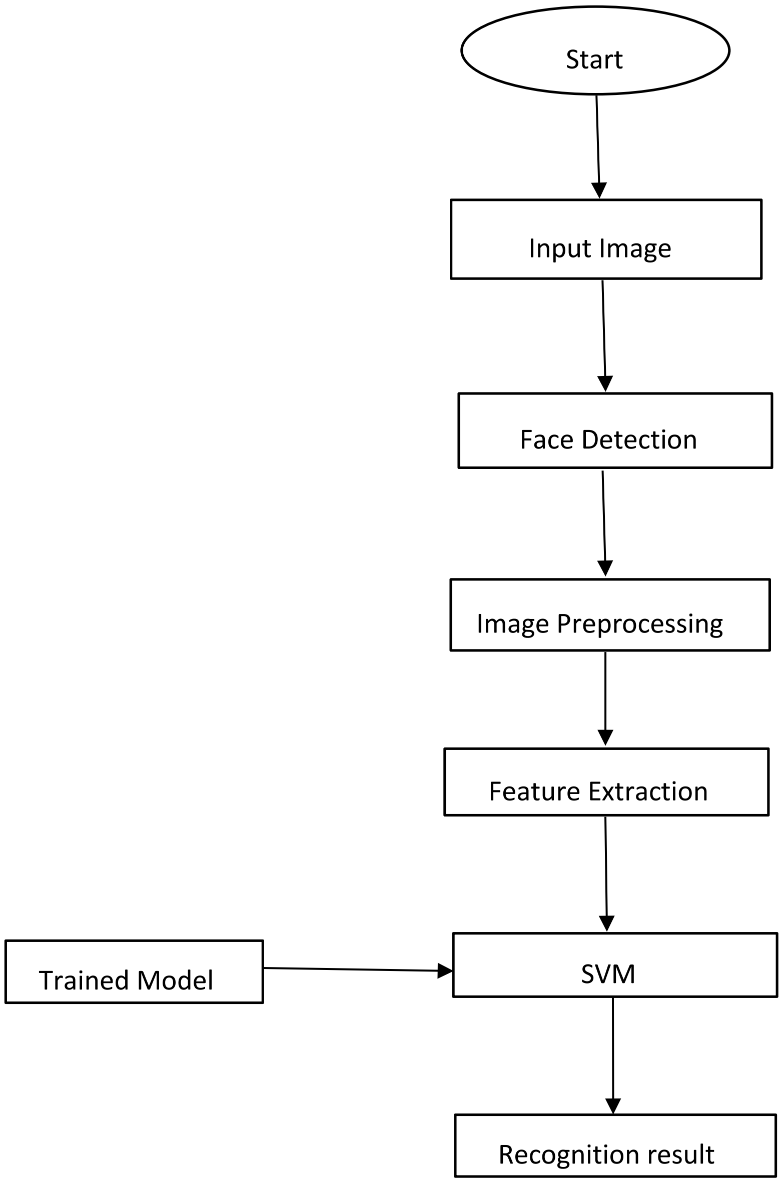

Figure 3.3: Flowchart of Testing/Prediction

The diagram is a flowchart of a face detection system, specifically the process of identifying faces in an image. Here's a breakdown of the steps:

Start: This is the beginning of the process.

Input Image: An image is captured by a camera or uploaded from a file.

Face Detection: This stage determines whether the image contains a face or not. There are several algorithms for face detection, but they typically involve searching the image for features commonly found in faces, such as two eyes, a nose, and a mouth.

Image Preprocessing (Optional): This step improves the quality of the image for better face detection. Common preprocessing techniques include: Grayscale conversion: Converting the image from color to grayscale can simplify the image and make it easier to detect facial features. Noise reduction: Removing noise from the image can help to improve the accuracy of face detection. Normalization: Adjusting the brightness and contrast of the image can make it more consistent and easier to analyze.

Feature Extraction: If image preprocessing is included, this step extracts features from the preprocessed image. Facial features are characteristics that can be used to identify a face, such as the distance between the eyes, the shape of the nose, and the texture of the skin.

Trained Model: This refers to a machine learning model that has been trained to identify faces. The model is trained on a large dataset of images that contain and don't contain faces .During training, the model learns to identify the features that are most common in faces.

SVM (Support Vector Machine): This is a specific type of machine learning model that can be used for face detection. SVMs are trained to classify data into different categories. In face detection, the SVM would be trained to classify images as containing a face or not containing a face.

Recognition Result: This is the output of the process. If a face is detected in the image, the system may provide additional information, such as the location of the face in the image or the person's emotional expression.

3.5.2 Sequence diagram

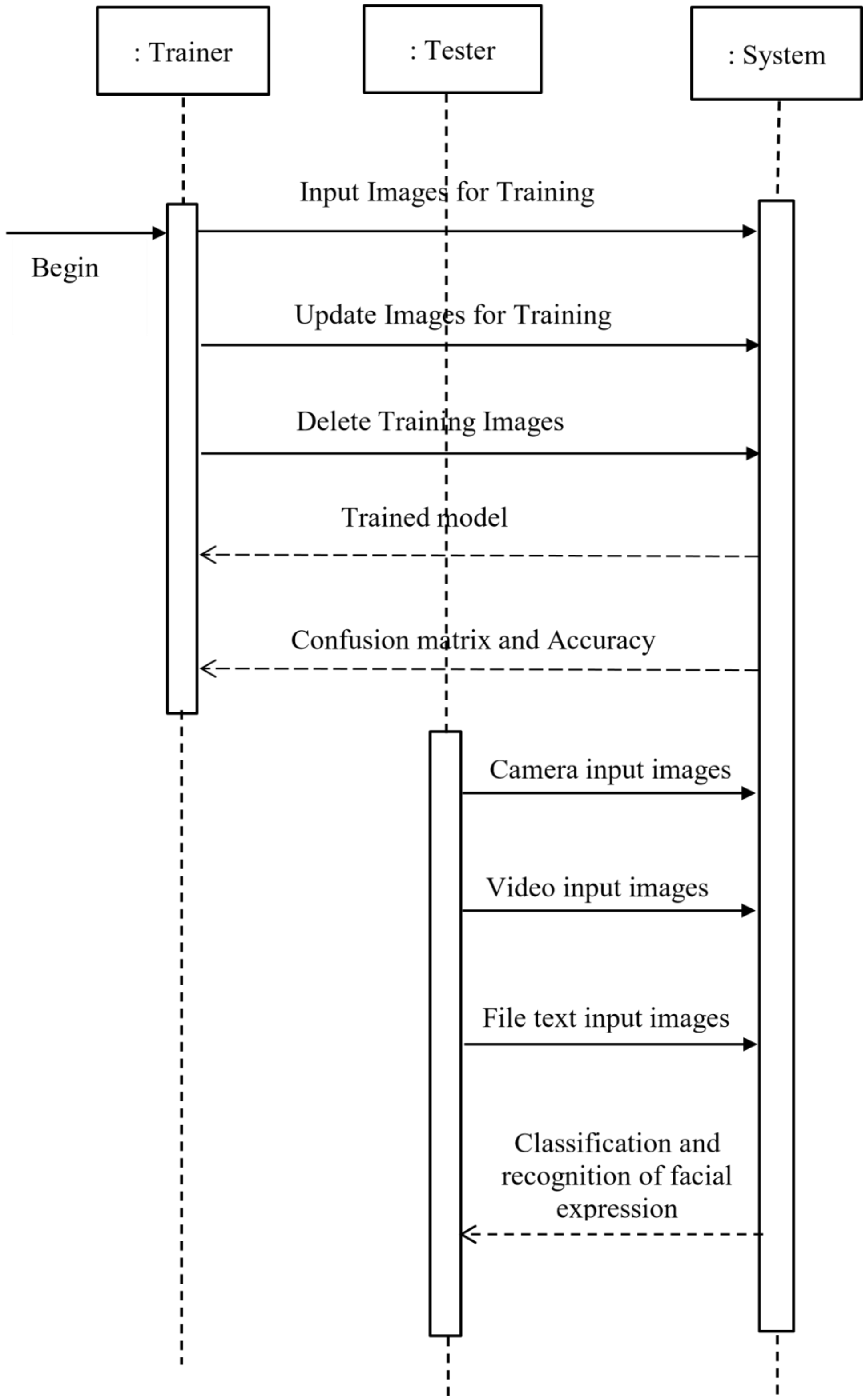

Figure 3.5: Sequence Diagram

The above diagram is a sequence diagram of a training process for a facial emotion recognition system. Let's break down the steps depicted in the diagram:

Inputs

Trainer: This actor represents the person or system managing the training process .Images for Training: This signifies the input of images that will be used to train the system to recognize facial expressions.

Process

Begin: This marks the start of the training process.

Input Images for Training: The trainer first provides images for the system to learn from.

Update Images for Training (Optional): This step allows the trainer to update the training data with new images or remove irrelevant ones, potentially improving the model's accuracy.

Delete Training Images (Optional): The trainer can delete training images that are no longer needed or found to be problematic.

Outputs

Trained Model: After processing the training images, the system outputs a trained model capable of recognizing facial expressions.

Tester: This actor represents the person or system that will evaluate the trained model.

Confusion Matrix and Accuracy: The system provides the tester with a confusion matrix and accuracy score. A confusion matrix is a table that allows visualization of the performance of the model on a test dataset. It shows how many predictions were correct and how many were incorrect for each facial expression.

3.5.3 File Structure of the Dataset

This is a more organized approach, where the data is stored in multiple files within a folder. Separate Files for Text and Labels: One .txt file stores the text data (e.g., happy, sad, neutral, angry), and another .js file stores the corresponding sentiment labels (e.g., separate CSV with IDs referencing the text files).

3.5.4 Front end Components

The frontend of the application was implemented using Flutter and the styling of the components was done using Flutter material. Flutter is a frontend Dart library used for building web applications and mobile application development. It's an open source library used specifically for building mobile applications. Components made include

Display App Title: The app bar typically displays the title of the current activity or screen.

*Navigation***:** Provides navigation controls like back arrow or up button to navigate back in the app hierarchy. (It can also provide) custom navigation elements for specific screens.

Display window: Displays the images caught on the camera.

3.5.5 Backend Components

*SDK Environment***:** The Flutter plugin for Android: It's a plugin specifically designed to bridge the gap between Flutter and native Android functionalities such as acting as a bridge between Flutter code and the underlying Android system. It allows access to various native Android features from the Flutter app, including Camera access, GPS location Battery level Sensors (accelerometer, gyroscope, etc.), notifications Toast messages, hardware vibrations, and integration with platform-specific UI elements (if necessary).

Camera Plugin: Flutter doesn't have built-in functionalities for camera access. However, we can achieve camera functionality in our Flutter app .Using the camera plugin is the most common approach to leverage the camera plugin, a popular third-party package available on pub.dev (the package manager for Dart).The camera plugin provides functionalities like opening the camera, capturing photos or videos, handling different camera resolutions, and accessing specific camera features.

TFLite Plugin: TFLite is a lightweight machine learning framework from Google, specifically designed for mobile and embedded devices. It allows the running of pre-trained TensorFlow models on these devices, enabling tasks like image classification or facial recognition directly on the device. Here's how TFLite can be used as a dependency in the Flutter application:

3.6 Summary

Chapter three dives deep into the research methodology, meticulously detailing how the data was collected and analyzed. This chapter plays a critical role as it establishes the foundation for the study's credibility by demonstrating the validity and reliability of the findings. It also delves into the functionalities of the developed system, providing a clear illustration of how data flows throughout the entire process.

CHAPTER FOUR: Data Presentation, Analysis and Interpretation

4.0 Introduction

This chapter delves into the heart of the research – the data analysis process. We'll explore how the obtained data is presented, meticulously analyzed, and ultimately interpreted. The emphasis here lies on the crucial role these steps play in transforming raw data into actionable insights. By effectively analyzing and interpreting the data, we can extract meaningful conclusions that will directly inform strategic business decisions.

4.1 Testing

This chapter highlights the testing process, which is critical for identifying and eliminating flaws (faults or bugs) that could hinder a product, system, or application's performance, functionality, or usability. Testing ensures the solution meets all the defined criteria and functions as intended, delivering a positive user experience. The chapter details the specific tests conducted, their results, and how they addressed both the functional and non-functional requirements of the proposed solution.

4.1.0 Black Box Testing

In black box testing, the tester focuses on what goes in and what comes out, without peeking under the hood. This method applies perfectly to our machine learning system for facial sentiment analysis and opinion mining. We'll evaluate the system's performance by examining its inputs (facial images) and outputs (predicted emotions and opinions) without needing to understand the internal workings of the algorithm.

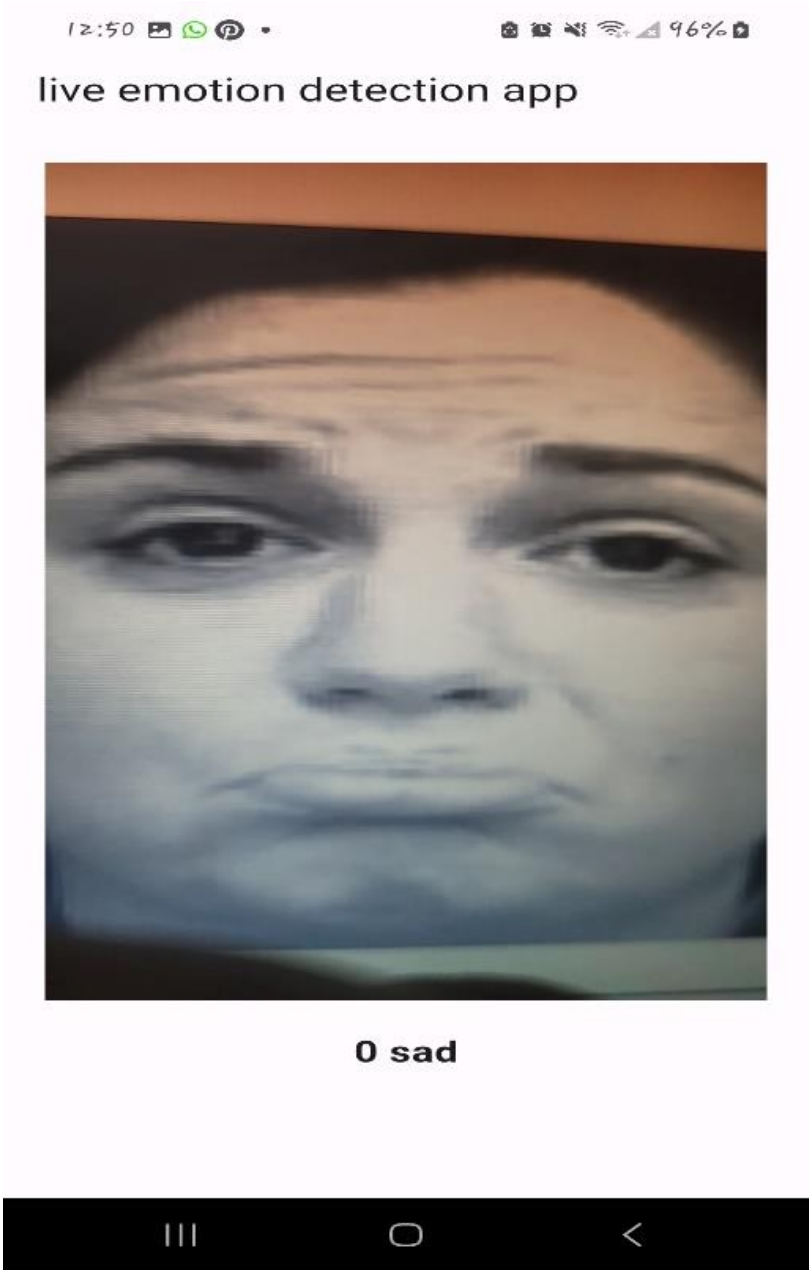

Figure Showing the System detecting a sad face

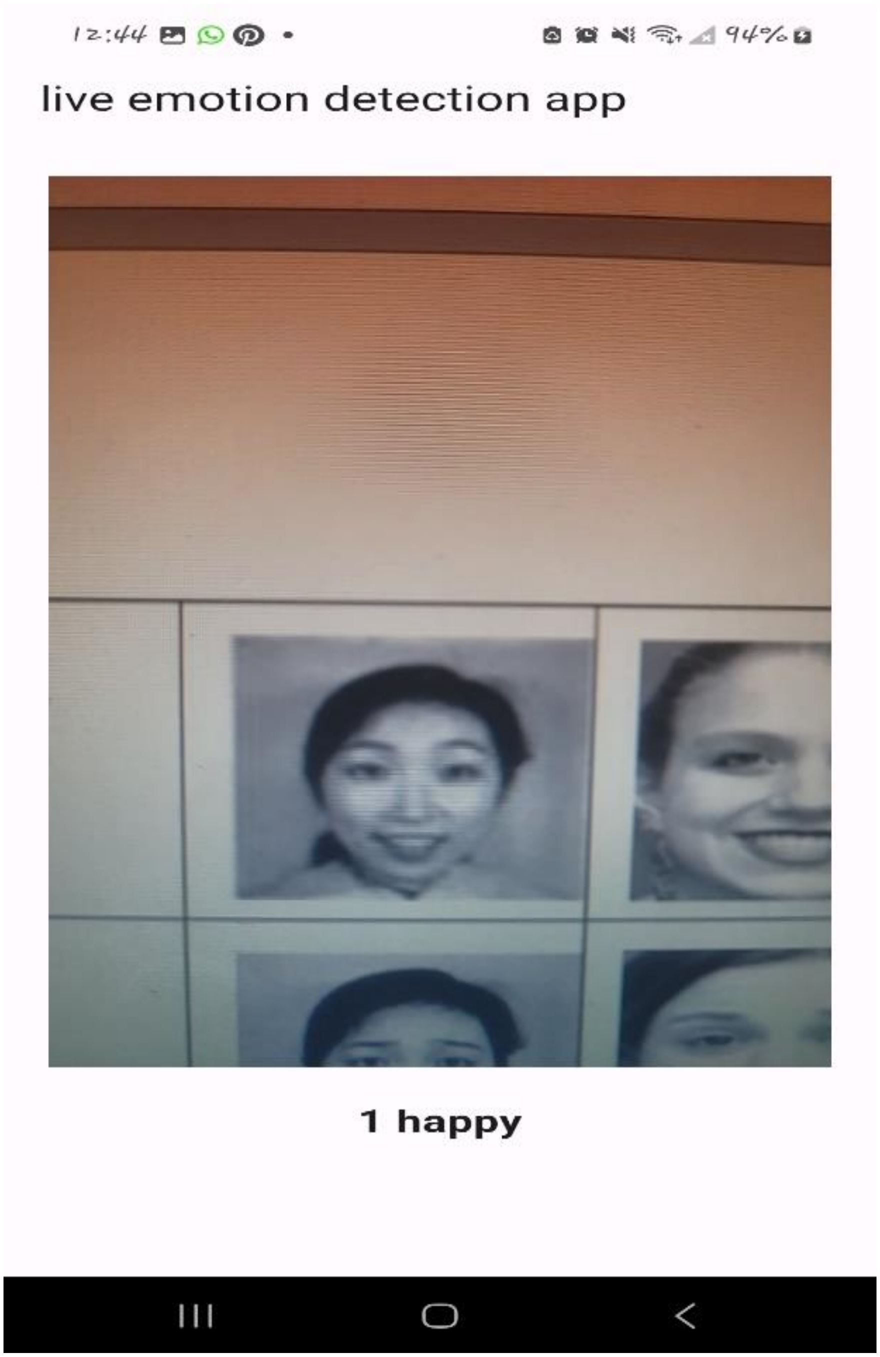

Figure Showing the System detecting a happy face

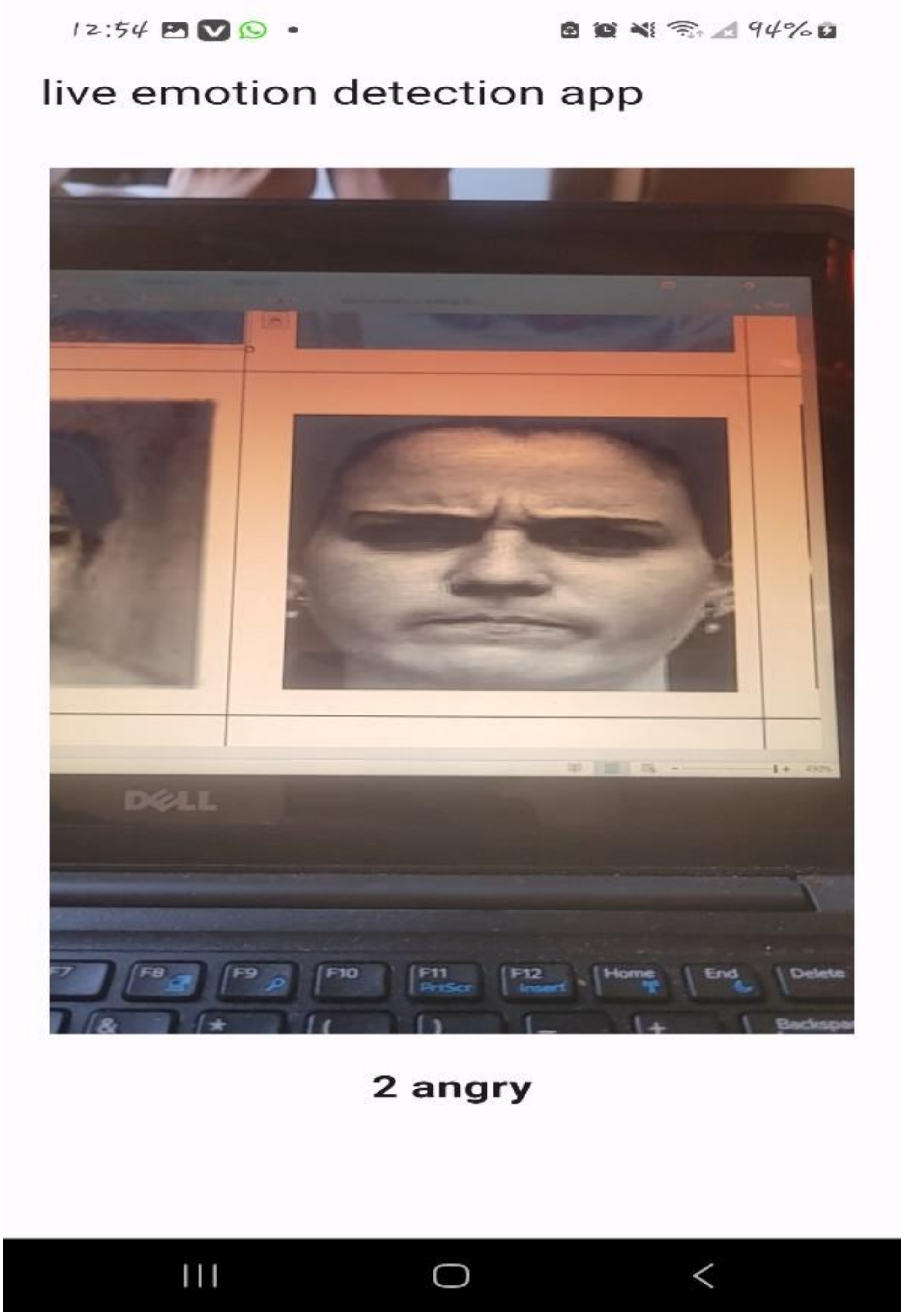

Figure Showing the System detecting an angry face

4.1.1 White Box Testing

White box testing, also known as structural testing or code-based testing, delves into the inner workings of a software program. Testers meticulously examine the code, algorithms, and logic to ensure the program functions as intended. This approach aims to guarantee that all possible code paths have been executed and tested, leaving no room for hidden errors..

lib > home.dart

```dart
class _HomeState extends State<Home> {
  loadmodel() async {
        model: "assets/model.tflite", labels: "assets/labels.txt");
  }

  @override
  Widget build(BuildContext context) {
    return Scaffold(
      // ignore: prefer_const_constructors
      appBar: AppBar(title: Text('live emotion detection app')),
      body: Column(
        children: [
          Padding(
            padding: const EdgeInsets.all(20),
            child: SizedBox(
              height: MediaQuery.of(context).size.height * 0.7,
              width: MediaQuery.of(context).size.width,
              child: !cameraController!.value.isInitialized
                  ? Container()
                  : AspectRatio(
                      aspectRatio: cameraController!.value.aspectRatio,
                      child: CameraPreview(cameraController!),
                    ), // AspectRatio
            ), // SizedBox
          ), // Padding
          Text(
            output,
            // ignore: prefer_const_constructors
            style: TextStyle(fontWeight: FontWeight.bold, fontSize: 20),
          ) // Text
        ],
      ), // Column
    ); // Scaffold
```

Fig : client facial sentimental analysis system test code

These tests cover the basic functionality of the client facial sentimental analysis system's code.

4.2 Evaluation Measures and Results

To assess how well a classifier performs, we rely on evaluation metrics (Hossin & Sulaiman, 2015). These metrics can be broadly categorized into three groups as proposed by Hossin & Sulaiman (2015): threshold-based, probability-based, and ranking-based metrics.

4.2.1 Precision and Recall

Recall

Recall tells us how good the system is at finding all instances of a particular expression. In other words, it measures how many actual happy faces the system identified correctly out of all the happy faces that were present. Recall is calculated as the percentage of correctly classified expressions relative to the total number of expressions in that category.

$$Recall = \frac{TP}{TP + FN}$$

Precision

Precision tells us how accurate the system is at identifying a specific facial expression. For example, if the system classifies an image as "happy" and it actually is happy, that contributes to high precision for the "happy" category. Precision is calculated as the percentage of correctly classified expressions out of all the expressions the system assigned a category to

$$Precision = \frac{TP}{TP + FP}$$

F-Score

F-score combines precision and recall into a single metric to evaluate how well the system classifies facial expressions. It considers both how accurate the system is (precision) and how good it is at finding all expressions (recall)

F1 Score = 2(Precision*Recall / (Precision + Recall))

4.2.2 Confusion Matrix

To assess how well a classifier performs, we rely on evaluation metrics (Hossin & Sulaiman, 2015). These metrics can be broadly categorized into three groups as proposed by Hossin & Sulaiman (2015): threshold-based, probability-based, and ranking-based metrics.

Table 4: Confusion matrix of JAFFE

Labels	Angry	Happy	Neutral	Sad
Angry	4	0	0	1
Happy	0	10	2	0
Neutral	0	0	6	0
Sad	0	0	0	10

Figure 4.3.3 Confusion matrix of JAFFE

In the above table, row shows the actual classes and column shows the predicted classes. The classifier made a total of 53 predictions where the classifier predicted angry for 4 times, happy for 10 times, neutral for 9 times and sad for 11 times. Whereas in reality 6 cases were angry, 12 was happy, 6 was neutral and 10 was sad.

Table 5: Accuracy of JAFFE

Evaluation Types	Percentages
Precision	91.8986
Recall	98.3649
F-score	95.0218

Figure 4.3.4 Accuracy of JAFFE

The above table shows that 91.8986% of the expressions were predicted, 98.3649% of the expressions were correctly assigned. The harmonic mean of precision and recall was

95.0218%.

4.3 Summary of Research Findings

Based on the performance metrics supplied, the K-means clustering algorithm created relatively well separated, not highly overlapping clusters. This shows that the algorithm captures the underlying structure of the data effectively and has the potential to be used in a customer segmentation system. However, further research and evaluation of the clustering results using other performance indicators and visual inspection of the cluster assignments may be required to determine the success of the customer segmentation system. The Calinski-Harabasz index value of 766, in particular, indicates that the clusters are relatively well separated from one another, although the Davies-Bouldin index value of 1.38 indicates that there is still considerable overlap between the clusters. The cluster sum of squares value of 325 shows that the clusters are relatively compact and well separated from one another, whilst the silhouette score of 0.25 indicates that the clusters are moderately well defined.

4.4 Conclusion

The findings indicate that the K-means algorithm produced relatively well separated clusters that could be beneficial in a customer segmentation system, but further research and analysis may be required to establish the appropriate number of clusters and the overall usefulness of the system.

Chapter Five: Conclusion and Recommendations

5.1 Introduction

In the previous chapter, the researcher focused on presentation and analysis of obtained data. This chapter covers the research and development of the solution in line with the set objectives. This chapter will also examine the difficulties encountered by the researcher in designing and carrying out this study.

5.2 Aims and Objectives Realizations

The main aim of this study was to develop a machine learning customer segmentation system and be able the use this system to group customers so that the business can be able to devise marketing strategies in order to build they customer retention. The first objective of this study was to analyze different machine learning techniques of customer segmentation. This objective was met in chapter two as the researcher analyzed the different techniques and decided to use k means cluster amongst all of them because it is an unsupervised machine-learning algorithm used for clustering, or grouping, data points that share similar features. The second objective was to build and implement a customer segmentation system. This objective was met, a machine-learning customer segmentation system was built and implemented, the methodology used was shown in chapter three and the results produced were indicated in chapter four. The third objective was to evaluate the use of machine learning customer segmentation system and this objective was met in chapter four that this will help businesses can improve customer satisfaction, increase sales, and gain a competitive advantage in the marketplace.

5.3 Major Conclusions Drawn

In conclusion, the study demonstrated the effectiveness of using machine-learning techniques for customer segmentation. The research successfully achieved its objectives was to analyze different machine learning techniques of customer segmentation, was to build and implement a customer segmentation system and to evaluate the use of machine learning system in customer segmentation. The research proved that the use of machine learning is more effective and beneficial in customer segmentation as compared to the traditional techniques. The research also prove that k means cluster is a suitable algorithm in customer segmentation and can be used for large data sets.

5.3 Recommendations and future work

Based on the findings of the study, there are several recommendations for future work in the machine learning customer segmentation system. Developing an effective machine learning customer segmentation system requires careful planning, data pre-processing, algorithm selection, performance evaluation, deployment, and continuous improvement. By following these recommendations, researchers can develop more accurate and effective customer segmentation systems that help businesses improve their marketing and customer service efforts. Deep learning techniques, such as neural networks, have shown promising results in customer segmentation. Future studies could explore the use of deep learning techniques for customer segmentation. Future studies on machine learning customer segmentation could focus on comparing and evaluating different algorithms, incorporating external data sources and temporal dynamics, and evaluating the impact on business outcomes. By exploring these areas, researchers can develop more accurate and effective customer segmentation systems that help businesses improve their marketing and customer service efforts.

REFERENCES

1. Customer Segmentation with Machine Learning - Thecleverprogrammer
2. Implementing Customer Segmentation Using Machine Learning [Beginners Guide] - Neptune.ai
3. Unsupervised Machine Learning for Customer Market Segmentation - Coursera
4. What is Segmentation in Machine Learning - Deepchecks
5. Process of Customer Segmentation using machine learning - LinkedIn

6. Liang, J., You, X., & Chen, Y. (2019). Customer segmentation based on clustering analysis: A case study of an online retailer. Journal of Business Research
7. Liu, Y., Zhang, Z., & Liang, Y. (2020). A customer segmentation method based on feature selection and decision tree algorithm. Computers & Industrial Engineering
8. Verhoef et al. (2007), customer management has become more complex due to the emergence of multiple channels of interaction between customers and businesses, making it challenging to segment customers accurately.
9. Traditional methods of customer segmentation, such as demographic and psychographic factors, are limited in capturing the complexity of customer behavior (Han et al., 2011).
10. Kumar and Reinartz (2016) emphasize the importance of customer relationship management in today's business environment, where customers demand personalized experiences.
11. Online customer reviews have become a crucial element of the marketing communication mix (Chen & Xie, 2008), highlighting the need for accurate customer segmentation.
12. Machine learning algorithms offer a promising solution for customer segmentation, as demonstrated by Kim et al. (2017), who compared various machine-learning algorithms for customer churn prediction.
13. Wang and Zhang (2018) applied machine-learning techniques to mobile payment service data and achieved more accurate customer segmentation.
14. Mugenda, O. M., & Mugenda, A. G. (2013). Research methods: Quantitative and qualitative approaches. Nairobi, Kenya: Acts Press.
15. Song and Lu (2015) developed an improved K-means algorithm based on mobile payment service data to segment customers more accurately.
16. Kotler, P., Keller, K. L., Brady, M., Goodman, M., & Hansen, T. (2009). Marketing management (1st European ed.). Harlow: Pearson Education Limited.
17. Chen, Y., Li, Z., Zhang,X., & Liu, Y. (2020). Deep customer segmentation for e-commerce by leveraging clickstream data. Electronic Commerce Research and Applications
18. Shang, J., Shen, C., Zhang, X., & Chen, J. (2021). Customer segmentation based on clustering algorithm and its application in tourism industry. Journal of Ambient

19. Wang, L., & Shu, Z. (2020). Online customer segmentation based on browsing behavior and purchase history using K-means algorithm. Journal of Ambient Intelligence and Humanized Computing
20. Chen, Y., Li, Z., Zhang, X., & Liu, Y. (2020). Deep customer segmentation for e-commerce by leveraging clickstream data. Electronic Commerce Research and Applications
21. Wang, L., Feng, X., & Zhang, Y. (2021). Customer segmentation based on RFM model and machine learning algorithms. Journal of Intelligent & Fuzzy Systems
22. Li, Y., Li, Z., Wang, Z., & Wu, Y. (2021). Customer segmentation based on deep learning and fuzzy clustering in online retail. Journal of Intelligent & Fuzzy Systems
23. "Machine Learning for Customer Segmentation: A Review of the Literature" by A. Tsay, J. Chen, and M. Hu (2016)
24. "Customer Segmentation Using Machine Learning: A Case Study" by B. Wang, Y. Chen, and X. Zhang (2017)
25. "A Survey of Machine Learning Methods for Customer Segmentation" by C. Zhang, X. Li, and Y. Wang (2018)
26. 2018 paper titled "Customer Segmentation for E-commerce: An Overview," Chaoqun Ma and Yong Li
27. 2021 paper titled "Customer Segmentation and Targeting in E-commerce Using Machine Learning: A Case Study," Shuting Xu, Weili Wu, and Jie Li
28. "A Survey of Machine Learning Methods for Customer Segmentation" by Zhang, Li, and Wang (2018)
29. Tsay, A., Chen, J., & Hu, M. (2016). Machine learning for customer segmentation: A review of the literature. Expert Systems with Applications, 59, 242-261. doi: 10.1016/j.eswa.2016.05.003
30. Babbie, E., & Mouton, J. (2018). The practice of social research. Oxford University Press Southern Africa.
31. Hulley, S. B., Cummings, S. R., Browner, W. S., Grady, D., & Newman, T. B. (2013). Designing clinical research: An epidemiologic approach. Lippincott Williams & Wilkins.
32. the book "Software Testing: Principles and Practices" by Srinivasan Desikan and Gopalaswamy Ramesh,s

33. machine learning customer segmentation system data flow - Bing images
34. John Simpson, Edmund Weiner, James Murray. (2010) Oxford Dictionary of English (3rd Edition). Oxford University Press.
35. Calinski, T., & Harabasz, J. (1974). A dendrite method for cluster analysis. Communications in Statistics - Theory and Methods, 3(1), 1-27.
36. Milligan, G. W., & Cooper, M. C. (1985). An examination of procedures for determining the number of clusters in a data set. Psychometrika, 50(2), 159-179.
37. Hastie, T., Tibshirani, R., & Friedman, J. (2009). The elements of statistical learning: Data mining, inference, and prediction (2nd ed.). Springer.
38. Jain, A. K., & Dubes, R. C. (1988). Algorithms for clustering data. Prentice-Hall.
39. Kaufman, L., & Rousseeuw, P. J. (1990). Finding groups in data: An introduction to cluster analysis. John Wiley & Sons.
40. Scikit-learn documentation: Cluster evaluation. Retrieved from https://scikit-learn.org/stable/modules/clustering.html#clustering-performance-evaluation