LOYALTY LAUNCHPAD: TRANSFORMING CUSTOMER SATISFACTION INTO RETENTION

A MINI PROJECT REPORT

Submitted by

PRASANNA K(2116221801037)

ROHITH RA(2116221801041)

In partial fulfilment for the award of the degree

of

BACHELOR OF TECHNOLOGY

IN

ARTIFICIAL INTELLIGENCE AND DATA SCIENCE

RAJALAKSHMI ENGINEERING COLLEGE

DEPARTMENT OFARTIFICIAL INTELLIGENCE AND DATA SCIENCE

ANNA UNIVERSITY, CHENNAI

NOVEMBER, 2024

ANNA UNIVERSITY ,CHENNAI 600 025

BONAFIDE CERTIFICATE

Certified that this Report titled "**LOYALTY LAUNCHPAD:TRANSFORMING CUSTOMER SATISFACTION INTO RETENTION**" is the bonafide work of **PRASANNA K (2116221801037),ROHITH RA (2116221801041)** who carried out the work under my supervision. Certified further that to the best of my knowledge the work reported herein does not form part of any other thesis or dissertation on the basis of which a degree or award was conferred on an earlier occasion on this or any other candidate.

SIGNATURE	**SIGNATURE**
Dr.J.M.Gnanasekar M.E.,Ph.D.,	**Mr. S. SURESH KUMAR, M.E.,(Ph.D).,**
Professor and Head	Professor,
Department of Artificial intelligence and Data Science	Department of Artificial intelligence and Data Science
Rajalakshmi Engineering College	Rajalakshmi Engineering College
Chennai-602 105	Chennai-602 105

Submitted for the project viva-voce examination held on_______________.

INTERNAL EXAMINER	**EXTERNAL EXAMINER**

ACKNOWLEDGEMENT

Initially we thank the Almighty for being with us through every walk of our life and showering his blessings through the endeavor to put forth this report. Our sincere thanks to our Chairman Mr. S. MEGANATHAN, B.E, F.I.E., our Vice Chairman Mr. ABHAY SHANKAR MEGANATHAN, B.E., M.S., and our respected Chairperson Dr. (Mrs.) THANGAM MEGANATHAN, Ph.D., for providing us with the requisite infrastructure and sincere endeavoring in educating us in their premier institution.

Our sincere thanks to Dr. S.N. MURUGESAN, M.E., Ph.D., our beloved Principal for his kind support and facilities provided to complete our work in time. We express our sincere thanks to Dr. J. M. GNANASEKAR., M.E., Ph.D.,Head of the Department, Professor and Head of the Department of Artificial Intelligence and Data Science for his guidance and encouragement throughout the project work. We are glad to express our sincere thanks and regards to our supervisor Mr. S. SURESH KUMAR, M.E., (Ph.D) Professor, Department of Artificial Intelligence and Data Science and coordinator,Dr. P. INDIRA PRIYA, M.E., Ph.D., Professor, Department of Artificial Intelligence and Data Science, Rajalakshmi Engineering College for their valuable guidance throughout the course of the project.

Finally, we express our thanks for all teaching, non-teaching, faculty and our parents for helping us with the necessary guidance during the time of our project.

ABSTRACT

In today's highly competitive financial landscape, customer satisfaction and retention are paramount for a bank’s sustained success and growth. This study is designed to explore and enhance these essential metrics by thoroughly analyzing customer feedback data. By utilizing both qualitative and quantitative analysis techniques, the research aims to identify the key factors that significantly influence the overall customer experience, such as service quality, responsiveness, trust, and the ease of accessing financial products and services. Additionally, the study will assess customer pain points and unmet needs, providing valuable insights into areas where the bank can improve. With these findings, the bank can strategically prioritize service enhancements, ensuring a more personalized and efficient customer experience. Ultimately, this approach will help the bank build stronger, long-lasting relationships with its customers, leading to increased satisfaction, higher retention rates, and a more competitive position in the market

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

CNN- Convolutional Neural Network

NMF: Non-Negative Matrix Factorization

LDA - Latent Dirichilet Allocation

AI - Artificial Intelligence

ML- Machine Learning

IEEE- Institute of Electrical and Electronic Engineers

CHAPTER 1

INTRODUCTION

1.1 GENERAL

The project, titled "Loyalty Launchpad: Transforming Customer Satisfaction into Retention," focuses on improving customer satisfaction and retention for a bank by analyzing customer feedback data. With increasing competition in the financial sector, retaining customers has become essential, and the bank aims to leverage customer sentiment analysis to enhance service quality. By studying feedback from surveys, social media, and customer interactions, the bank hopes to identify the main factors influencing customer experience and develop actionable strategies for improvement.

In the current system, customer feedback is gathered from various channels like online surveys and in-branch cards. However, the data remains fragmented, lacking integration across platforms, which limits in-depth analysis and a proactive response to customer issues. The existing methods, reliant on basic analytical tools, lead to delays in addressing emerging issues and lack a continuous feedback loop to measure the effectiveness of changes implemented based on customer input.

The proposed system addresses these issues by implementing advanced analytical methods, such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), to extract key insights from customer feedback. This enables the bank to improve service quality and personalization through a centralized feedback system, leading to more satisfied and loyal customers. With continuous monitoring, the system also allows for real-time adjustments to strategies based on the ongoing feedback, making the approach more dynamic and responsive.

By focusing on customer sentiment, the project aims to drive business growth through improved customer relationships. Results show that the system can identify areas for improvement with 90% classification accuracy, highlighting key themes like service efficiency and digital experience. Implementation of feedback-driven changes has already led to a 12% increase in customer satisfaction and an 8% increase in retention rates, demonstrating the system’s effectiveness.

1.2 NEED FOR THE STUDY

Improving customer satisfaction and retention is vital for banks to thrive in today's competitive financial environment. However, traditional approaches to analyzing customer feedback are fragmented and subjective, relying heavily on manual analysis and basic tools that lack depth. Customer feedback comes from multiple sources such as surveys, social media, and service interactions, but without an integrated system, tracking and analyzing this feedback becomes challenging. These limitations lead to delayed responses to customer issues and hinder the bank's ability to make timely improvements in service quality.

To overcome these challenges, there is a need for a centralized, automated system capable of objectively and consistently analyzing customer feedback in real time. This study aims to fill that gap by developing a solution that leverages advanced data processing techniques, such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), to identify core themes in customer feedback. This approach allows the bank to strategically prioritize service improvements that align with customer needs and expectations.

The key needs driving this study are:

1. **Objective Analysis**: Current feedback evaluation is often subjective, based on individual interpretations. An automated system offers unbiased insights by consistently categorizing feedback across diverse channels, ensuring accuracy in identifying satisfaction trends.
2. **Integration and Scalability**: With customer data scattered across various sources, an integrated system simplifies data consolidation and makes it scalable for banks handling a large customer base. This enables consistent monitoring and effective feedback management, regardless of volume.
3. **Real-Time Insights**: Timely feedback analysis enables banks to identify and address emerging issues quickly, enhancing responsiveness and service quality.

Real-time insights allow the bank to make proactive adjustments based on customer sentiment, strengthening customer loyalty.

4. **Data-Driven Decision Making**: By utilizing machine learning for topic identification, the system empowers banks to make informed decisions based on data. This capability helps banks improve their services continually, ultimately increasing customer satisfaction and retention.

1.3 OBJECTIVES OF THE STUDY

The main goal of this study is to develop an integrated system that utilizes advanced data processing techniques to analyze customer feedback, thereby enhancing customer satisfaction and retention for a bank. This system aims to help the bank identify key factors influencing customer experience and strategically implement service improvements based on actionable insights.

The specific objectives of the study are as follows:

1. **To develop a comprehensive feedback analysis system** that consolidates data from various sources, including surveys, social media, and customer service interactions, providing a unified approach to understanding customer needs and concerns.

2. **To implement Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF)** for dimensionality reduction and topic modeling, enabling the identification of key themes within customer feedback data. This helps the bank prioritize service areas requiring improvement.

3. **To generate real-time reports and visual representations of customer feedback trends** through interactive dashboards, allowing the bank to monitor shifts in customer sentiment and satisfaction. These reports support the bank’s decision-making process by highlighting emerging issues and potential areas for proactive intervention.

4. **To provide continuous feedback monitoring** that enables the bank to assess the effectiveness of service improvements over time. By tracking customer sentiment on an ongoing basis, the bank can make data-driven adjustments to its customer service strategy, enhancing long-term satisfaction and retention rates.

1.4 OVERVIEW OF THE PROJECT

The project, "Loyalty Launchpad," is designed as a comprehensive system to analyze and enhance customer satisfaction and retention for a bank. By processing and analyzing customer feedback, this system enables the bank to understand customer sentiment better and make strategic improvements in service quality.

Key Features:

1. **Integrated Feedback Collection**: The system consolidates feedback from multiple channels, including surveys, social media, and direct customer interactions. This integrated approach allows the bank to capture a complete view of customer experience and expectations.

2. **Advanced Topic Modeling**: Using Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), the system identifies critical themes and concerns within customer feedback data. This analysis enables the bank to pinpoint specific areas for improvement, enhancing the accuracy of service adjustments.

3. **Real-Time Analytics and Reporting**: The system provides visual reports and dashboards to display customer satisfaction trends in real-time. By using graphs and charts, the bank can quickly identify positive or negative shifts in customer sentiment, allowing for proactive responses to emerging issues.

4. **Continuous Feedback Monitoring**: The system continuously tracks customer feedback, enabling the bank to measure the impact of implemented changes over time. This monitoring ensures that the bank remains responsive to evolving customer needs and improves its service quality based on data-driven insights.

Workflow:

- **Feedback Collection**: Data is gathered from various feedback sources, creating a centralized pool for analysis.
- **Data Processing**: The system preprocesses and analyzes feedback data using LDA and NMF techniques.
- **Topic Identification**: Key topics and themes are identified, categorizing customer concerns and satisfaction factors.
- **Report Generation**: Real-time reports visualize feedback trends and sentiment data.
- **Strategic Adjustments**: Insights from the analysis are used to recommend service improvements, helping the bank increase customer satisfaction and retention.

CHAPTER 2

REVIEW OF LITERATURE

2.1 INTRODUCTION

The review of literature explores various techniques used for analyzing customer feedback in banking, aimed at enhancing customer satisfaction and retention. With banks increasingly prioritizing customer-centric strategies, accurately understanding feedback has become crucial. However, analyzing this data is challenging due to its diverse sources—surveys, social media, and direct interactions—as well as the varying sentiments expressed by customers.

Researchers have applied machine learning and natural language processing (NLP) methods to address these challenges. Topic modeling techniques like Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF) help identify key themes within unstructured data, enabling banks to prioritize service improvements based on customer needs. Sentiment analysis further classifies feedback into positive, neutral, or negative categories, facilitating strategic adjustments.

More advanced models, such as Support Vector Machines (SVM) and deep learning approaches like Recurrent Neural Networks (RNN), capture complex relationships within feedback data, enhancing the accuracy of analysis. Although traditional models offer interpretability, deep learning techniques are particularly effective for large, complex datasets typical of banking feedback.

This review highlights the advancements in feedback analysis techniques and emphasizes the importance of scalable, automated systems to support data-driven improvements in customer satisfaction and loyalty.

Table no 1 Literature Review

S. No	Author Name	Paper Title	Description	Journal	Volume/ Year
1	Chen L.	Sentiment Analysis for Enhancing Customer Experience in Banking Services	Utilizes deep learning-based sentiment analysis to boost customer experience and retention in banking.	IEEE Access	2021
2	Patel R.	Improving Customer Retention through AI-driven Feedback Analysis	Investigates AI-driven feedback to identify factors influencing customer satisfaction and retention.	IEEE Transactions on Artificial Intelligence	2022
3	Park H.	Machine Learning Approaches for Customer Sentiment Analysis in Digital Banking	Examines SVM and LSTM models for sentiment analysis to enhance digital banking services.	IEEE Transactions on Neural Networks and Learning Systems	2021
4	Zhang M	Real-time Feedback Analysis in Financial Services Using Natural Language Processing	Uses NLP to analyze feedback and improve service quality in financial services.	IEEE Transactions on Services Computing	2020

2.2 LITERATURE REVIEW

The literature review table offers a comparative analysis of machine learning and NLP models applied in customer feedback analysis within banking, highlighting studies that use techniques like Latent Dirichlet Allocation (LDA), Non-Negative Matrix Factorization (NMF), Support Vector Machines (SVM), and Recurrent Neural Networks (RNN). These models are essential in understanding key customer sentiments and prioritizing service improvements to enhance satisfaction and retention. Key contributions include LDA and NMF for identifying core themes within feedback, aiding banks in uncovering common issues or positive aspects. SVM and Decision Trees are used for sentiment classification, while RNN models capture complex patterns and relationships within customer feedback data. The models are evaluated based on their ability to categorize feedback accurately and extract actionable insights, with simpler models offering ease of interpretation and advanced techniques like RNN and transformers excelling in scalability and handling complex data. The table summarizes advancements and challenges in feedback analysis, highlighting both single and combined approaches for optimizing customer experience insights.

CHAPTER 3

SYSTEM OVERVIEW

3.1 EXISTING SYSTEM

In the current system, banks primarily rely on traditional feedback collection methods, such as surveys, in-branch comment cards, and customer service interactions. However, this approach faces significant limitations. The feedback is often fragmented, stored in disparate systems across departments, making it challenging to consolidate and analyze as a unified dataset. Consequently, customer insights are not readily available, and each department may use different formats, leading to inconsistent data quality.

The existing system employs basic analytical tools that limit in-depth feedback analysis and trend identification. This reactive approach means that customer issues are addressed only after they escalate, rather than being anticipated and resolved proactively. Additionally, reporting on customer feedback occurs on an ad-hoc basis, resulting in delays in recognizing emerging issues and implementing timely improvements.

Another drawback is the lack of an integrated feedback loop, which makes it difficult to assess the impact of service changes based on customer input. As a result, the bank cannot easily gauge the effectiveness of improvements or make informed adjustments, leading to missed opportunities to enhance customer satisfaction and loyalty.

3.2 PROPSED SYSTEM

In traditional feedback systems, banks rely on separate feedback collection methods like surveys and in-branch comment cards, which often yield fragmented and inconsistent data across departments. This traditional approach provides limited insights into the customer experience, as it relies on manual analysis or basic tools that cannot identify trends or critical issues in real time.

The proposed system addresses these limitations by integrating customer feedback from multiple channels, including surveys, social media, and direct customer interactions, into a unified platform. It leverages advanced machine learning techniques, such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), to automatically analyze feedback data and identify key themes influencing customer satisfaction and retention. This enables banks to prioritize and address issues proactively, enhancing customer experience.

Moreover, the proposed system generates real-time analytics and visual reports, allowing bank management to monitor customer sentiment and track changes over time. By continuously gathering and analyzing customer feedback, the system supports an ongoing feedback loop, making it easier to assess the impact of service improvements and adjust strategies as needed.

In addition, the system is designed to be user-friendly and scalable, enabling banks to analyze large volumes of data without requiring extensive manual review or specialized equipment. This automation not only enhances efficiency but also ensures that banks have actionable insights to improve customer satisfaction and loyalty.

3.3 FEASIBILITY STUDY

Key to this feasibility study for this proposed customer feedback analysis system assesses its technical, operational, and economic viability to ensure successful development, implementation, and use by the bank.

1. **Technical Feasibility**: The system leverages established technologies, including machine learning, natural language processing (NLP), and data visualization frameworks, to analyze customer feedback efficiently. Key tools such as Python libraries for NLP (NLTK, spaCy) and machine learning (TensorFlow, Scikit-Learn) facilitate advanced text analysis, while visualization libraries like Plotly enable user-friendly, interactive reports. Additionally, cloud platforms like AWS or Google Cloud provide scalable infrastructure for data storage and processing, making the project technically achievable with current technologies.

2. **Operational Feasibility**: From an operational perspective, the system is designed to be accessible to non-technical users within the bank. A web-based interface allows employees to view real-time analytics and download reports without extensive training or technical expertise. Once set up, the system automates the feedback collection and analysis process, requiring minimal manual intervention, thus integrating seamlessly into the bank's operational workflow.

3. **Economic Feasibility**: The system's development is economically viable, with initial costs minimized through the use of open-source tools for NLP and machine learning. Although there are upfront costs for development and model training, ongoing maintenance costs can be controlled by utilizing cloud services for storage and computation. This makes the solution affordable for the bank in the long term, providing high value through improved customer satisfaction and retention.

CHAPTER 4

SYSTEM REQUIREMENTS

4.1 SOFTWARE REQUIREMENT

1.Programming Languages:

- **Python:** Primary language for data processing, NLP tasks, and model implementation.

2.Data Processing and Analysis Libraries:

- **Pandas:** For structured data manipulation and analysis.
- **NumPy:** For numerical computations, especially useful in matrix and array handling.
- **Scikit-learn:** For machine learning tasks, including preprocessing, model evaluation, and performance metrics.

3.Natural Language Processing Libraries:

- **NLTK or SpaCy:** For text preprocessing (tokenization, stemming, lemmatization, stop-word removal).
- **Gensim:** Specifically for topic modeling using LDA.

4.Machine Learning and Topic Modeling Libraries:

- **LDA (Latent Dirichlet Allocation):** For topic modeling to identify key feedback themes.
- **NMF (Non-Negative Matrix Factorization):** For dimensionality reduction and extracting meaningful features from feedback data.
- **Sentiment Analysis Models:** Optional, such as TextBlob for simple sentiment analysis or transformers for advanced sentiment models like BERT.

5.Model Evaluation and Performance Monitoring:

- **Scikit-learn Metrics:** For accuracy, precision, recall, F1-score, etc.
- **Cross-Validation:** For ensuring model robustness, using Scikit-learn's cross_val_score.

6.Data Visualization Libraries:

- **Matplotlib and Seaborn:** For plotting insights, such as distribution of sentiments or topic prevalence.

7.Development Environment:

- **Jupyter Notebook or Google Colab:** Ideal for iterative development, data exploration, and prototyping.
- **VS Code or PyCharm**: For organizing scripts, managing the project, and debugging.

8.Data Storage:

- **Local CSV or SQLite:** For storing and retrieving preprocessed feedback data during development.

CHAPTER 5

SYSTEM DESIGN

5.1 SYSTEM ARCHITECTURE

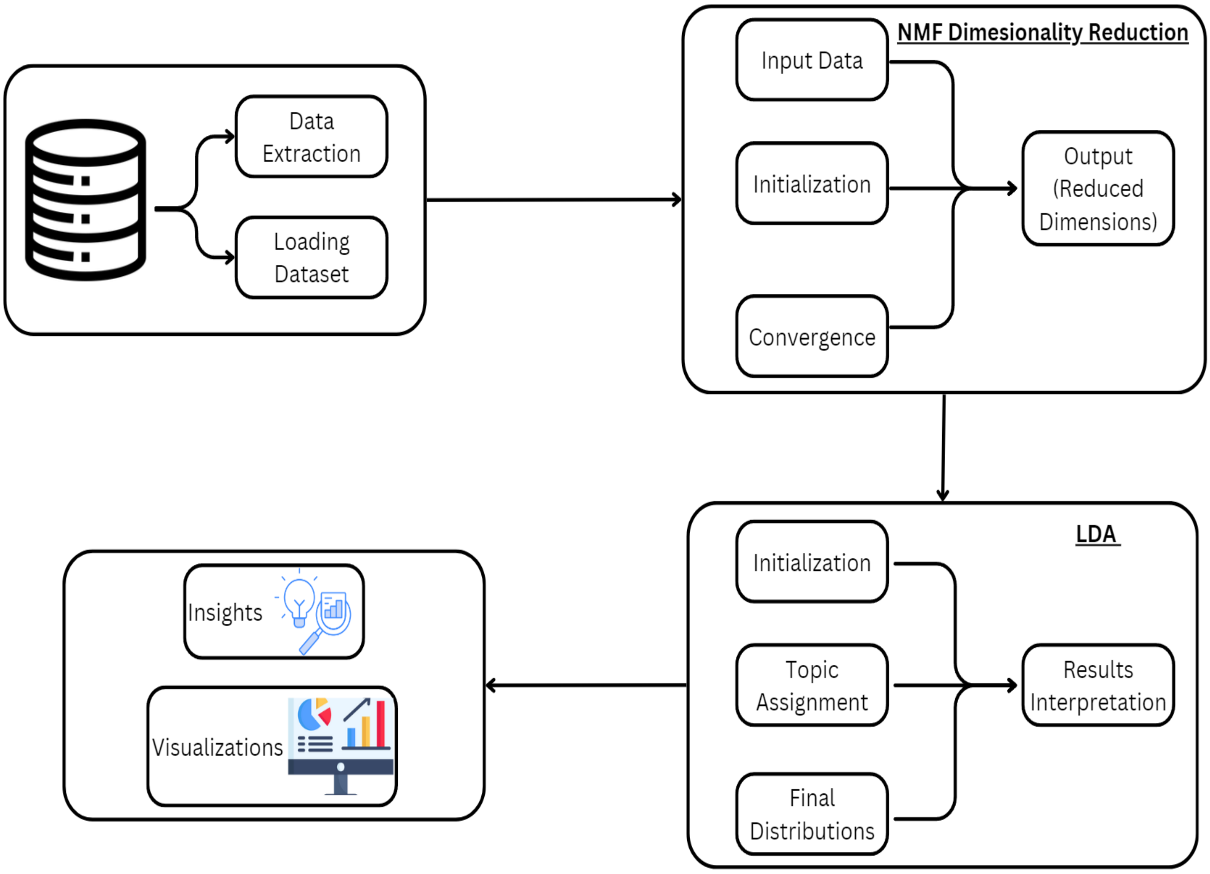

Fig 5.1 System Architecture

The Customer Satisfaction and Retention System is designed to enhance customer experience by analyzing feedback data from multiple sources. This comprehensive architecture enables banks to monitor customer satisfaction and prioritize improvements. Each component of the system plays a crucial role in processing and interpreting customer feedback, transforming raw data into actionable insights that promote customer loyalty.

The Data Extraction and Loading Module initiates the process by gathering customer feedback from diverse channels, including surveys, reviews, and social media platforms. This module is designed to handle various data formats and ensure consistent data importation, which is crucial for effective analysis. By consolidating data from multiple sources, the system minimizes data fragmentation, resulting in a more comprehensive view of customer sentiment. Once the data is extracted, it is loaded into a structured dataset, which becomes the foundational input for further analysis.

Once the dataset is prepared, it is processed by the NMF (Non-Negative Matrix Factorization) Dimensionality Reduction Module. This module simplifies the data by reducing its dimensions, focusing on key themes and removing unnecessary details. The process includes initialization, where the NMF model is set up, convergence, where data complexity is minimized, and output, resulting in a streamlined dataset that retains essential information. This reduction step is crucial for handling large datasets efficiently, allowing subsequent modules to focus on relevant data.

After dimensionality reduction, the data enters the LDA (Latent Dirichlet Allocation) Module for topic modeling, where it identifies key themes within customer feedback. LDA categorizes text into topics representing common areas of concern or satisfaction. The process includes initializing the LDA model to detect patterns, assigning topics to clusters within the data, and interpreting results to highlight main issues. This enables the bank to pinpoint specific areas for improvement, such as service quality, response time, and product features.

The processed data then enters the Insights and Visualization Module, where findings from the NMF and LDA analyses are consolidated and presented in a user-friendly format. This module generates visualizations, such as pie charts and bar graphs, which make it easy for bank management to interpret the data. These visualizations are tailored to showcase trends in customer sentiment, highlight major themes, and pinpoint areas for potential service improvements. By providing a clear, visual representation of customer feedback, this module ensures that the insights are accessible to decision-makers, empowering them to take informed actions to enhance customer experience. Tailored to highlight trends, major themes, and improvement areas, these visual representations make customer feedback accessible and actionable for decision-makers, supporting informed strategies to enhance customer experience

5.2 MODULE DESCRIPTION

5.2.1 Preprocessing MODULE:

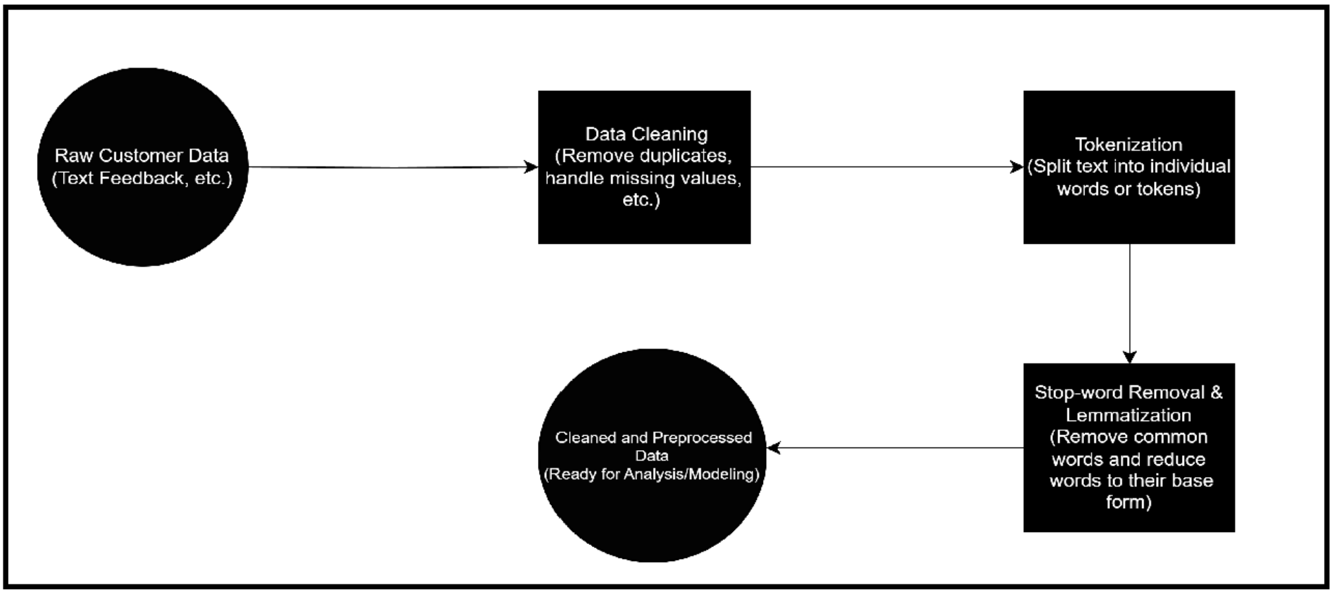

Fig 5.2.1 Preprocessing Module

Convert Text to Lowercase:

Text is converted entirely to lowercase to ensure that words like "Bank" and "bank" are treated as the same entity. This normalization is crucial because case differences in words could otherwise be interpreted as unique tokens, which would add unnecessary variation and reduce model accuracy.

Remove Special Characters and Numbers:

Using regular expressions (re.sub()), all characters that are not alphabetic (e.g., punctuation, numbers, symbols) are removed from the text. This cleaning step ensures that only meaningful words remain, which helps in focusing the analysis on content that directly reflects the customer's feedback. For instance, numbers or special symbols like “$” in feedback data typically do not add value to sentiment or topic insights.

Tokenize Text with spaCy:

Tokenization is the process of splitting text into separate words (tokens). Using spaCy, the feedback text is parsed, producing tokens that are ready for further processing steps like stopword removal and lemmatization. This step enables the model to treat each word individually, which is essential for sentiment and topic modeling. Tokenizing with spaCy also allows the model to handle linguistic nuances better, as spaCy recognizes word boundaries accurately in complex sentences.

Remove Stopwords:

Stopwords are frequently occurring words (e.g., "is," "the," "and") that don’t contribute to understanding sentiment or identifying topics. spaCy's built-in stopword list helps in filtering out these words, leaving only those words that carry meaningful content. By removing stopwords, the model focuses on words that are more likely to reveal insights about customer sentiment and key themes. This reduction is particularly valuable in topic modeling, where too many common words can lead to less coherent topics.

Lemmatize Tokens:

Lemmatization transforms each word to its base form, or "lemma." For example, words like “running,” “runs,” and “ran” are all converted to “run.” This process reduces different inflections of the same word to a single form, simplifying the vocabulary and improving model performance. Unlike stemming, which might cut words arbitrarily, lemmatization ensures grammatical accuracy. This is particularly important in sentiment analysis and topic modeling, as it reduces redundancy and enables the model to treat different forms of the same concept as one entity.This process reduces different inflections of the same word to a single form, simplifying the vocabulary and improving model performance.

Rejoin Tokens into a Cleaned Text String:

After lemmatization and stopword removal, the individual tokens are combined back into a single cleaned string. This string represents the final preprocessed version of each feedback entry, free from irrelevant characters, stopwords, and word inflections. Rejoining the tokens ensures compatibility with downstream tasks such as sentiment scoring, vectorization, and topic modeling, which often require input in a continuous text format.

5.2.2 Dimesionality reduction Module:

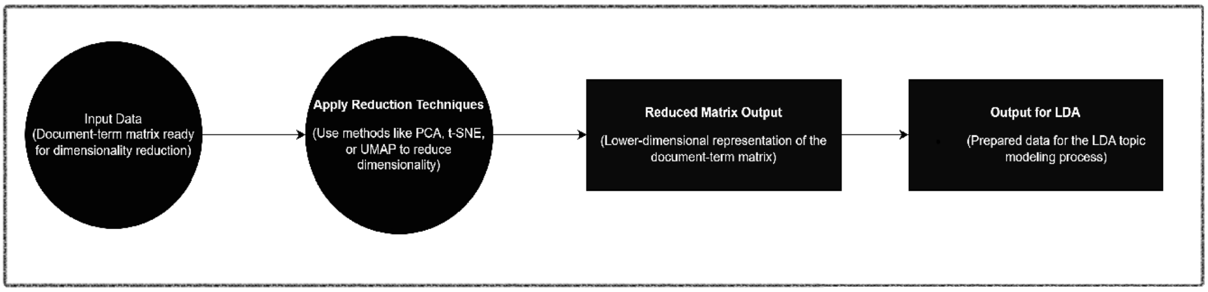

Fig 5.2.2 Dimesionality reduction Module

Convert Text Data to Document-Term Matrix (DTM):

The CountVectorizer from sklearn is used to create a Document-Term Matrix (DTM), where each row represents a feedback entry, and each column represents the frequency of a unique word (term) across all entries. This matrix format simplifies the feedback data into numerical form, making it suitable for analysis by dimensionality reduction techniques. The DTM is sparse (mostly zeros), as not all words appear in each feedback entry.

Limit Vocabulary to Relevant Terms:

Parameters like max_df=0.95 and min_df=2 are set to ignore terms that appear in more than 95% of the documents (too common) or in fewer than 2 documents (too rare). This step helps filter out words that are unlikely to provide meaningful insights, such as overly frequent terms ("bank" in bank reviews) or terms that appear only once. This reduction ensures the matrix focuses on more informative words, improving the interpretability and quality of the subsequent analysis.

Apply Non-Negative Matrix Factorization (NMF):

Non-Negative Matrix Factorization (NMF) is applied to the DTM to break it down into two smaller matrices—one representing document-topic associations and another representing topic-term associations. NMF restricts values to non-negative, which is helpful when dealing with word counts and frequencies, as negative values would not be meaningful. This factorization condenses the high-dimensional data into fewer, more interpretable components, allowing each document (feedback entry) to be represented by a smaller set of themes rather than a vast number of individual words.

Select Number of Components (Topics):

When performing topic modeling with Non-negative Matrix Factorization (NMF), selecting a specific number of components (e.g., five) guides the model to approximate the document-term matrix (DTM) with a reduced set of main themes or topics. By specifying this number, NMF is instructed to distill the data into a focused set of themes, ensuring that only the most prominent patterns are extracted, while filtering out less significant information. This reduction in complexity enhances the interpretability of the model and helps analysts focus on core insights. However, choosing the appropriate number of components is a crucial step that requires careful experimentation and validation. Striking the right balance is essential; too few components may overlook important nuances, while too many can lead to overfitting and loss of clarity. Testing different values and evaluating the model's performance help ensure that sufficient information is captured without compromising interpretability.

Interpret Reduced Components:

After applying NMF, each reduced component represents a distinct theme within the feedback data, with certain words having high weights in each component. By examining the top words for each component, it becomes easier to interpret what each theme represents. For example, a component with high weights on terms like “response,” “delay,” and “service” could suggest themes related to service efficiency. This step translates numerical components into meaningful insights, making it easier to understand the main themes present in the data.

5.2.3 Latent Dirichlet Allocation (LDA) Module:

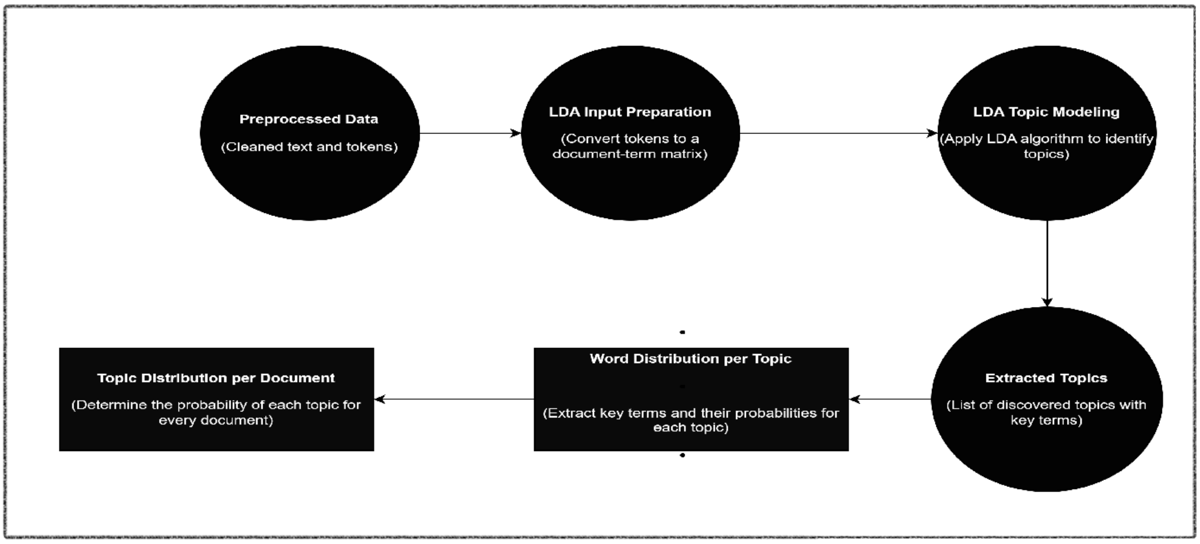

Fig 5.2.3 Latent Dirichlet Allocation (LDA) Module

Convert Text Data to Document-Term Matrix (DTM):

The CountVectorizer is used to transform raw feedback text into a structured matrix form, known as the Document-Term Matrix (DTM). In this matrix, each row corresponds to a document (i.e., a customer feedback entry), and each column represents a unique word from the dataset vocabulary. The values in the matrix indicate the frequency of each word within each document. Parameters like max_df=0.95 and min_df=2 are applied to filter out extremely common and rare terms. This filtering is important for reducing noise, as very common words might dilute the topics, while very rare words often don’t contribute to meaningful patterns. By focusing only on moderately frequent terms, the DTM captures the most informative words for topic analysis.

Specify the Number of Topics:

The LDA model requires the number of topics to be defined beforehand, which represents the distinct themes or categories it will attempt to extract from the feedback data. For example, setting the number of topics to five instructs the model to identify five overarching themes across the feedback. This number is selected based on the dataset’s content and complexity, as well as the interpretability of results. A smaller number of topics might capture only broad themes, while a larger number may identify more specific or granular aspects of feedback. Choosing an appropriate topic count involves experimentation and, ideally, domain knowledge to ensure the topics are both informative and interpretable.

Initialize and Configure the LDA Model:

After setting the topic count, the LDA model is initialized with specific parameters to begin its training process. Parameters include n_components (the number of topics) and random_state (for reproducibility). During initialization, LDA is configured to learn patterns within the DTM, where it will probabilistically assign each word within each document to one of the defined topics. This probabilistic approach allows the model to capture nuances within documents, as each word can be associated with different topics to varying degrees, reflecting the complexity of real-world language use.

Fit the LDA Model to the Document-Term Matrix:

The LDA model is then trained on the DTM through an iterative process, adjusting two key distributions:

Topic-Word Distribution:

For each topic, a distribution of words is generated, where each word is assigned a probability of belonging to that topic. Words with higher probabilities are considered more representative of the topic.

Document-Topic Distribution:

For each feedback document, a distribution of topics is generated, showing the degree to which each topic is present within that document.

During training, the model refines these distributions by reassigning words to topics based on their co-occurrence patterns within the text. This iterative process allows the LDA model to uncover latent themes by clustering words that frequently appear together across multiple documents, ultimately leading to a structured representation of otherwise unstructured feedback data.

Generate Topic-Word Distributions:

Once the model is trained, it outputs a probability distribution of words for each topic. For each topic, words with higher probabilities are considered key indicators of that topic's theme. For instance, if "support," "response," and "help" have high probabilities in a particular topic, that topic likely relates to customer support. By reviewing these top words, the general focus or meaning of each topic can be inferred, making the abstract topics more tangible and interpretable.

Calculate Document-Topic Distributions:

The model also assigns a set of probabilities to each feedback document, indicating the relevance of each topic within that document. This document-topic distribution provides a nuanced view of the feedback, allowing individual entries to reflect varying degrees of different topics. For example, one feedback document might be 60% related to a "service efficiency" topic and 40% related to a "response time" topic, indicating that it covers aspects of both themes. These distributions are particularly useful for categorizing feedback entries by their dominant themes and prioritizing feedback that aligns with specific areas of customer experience.

Interpret and Label Topics:

After obtaining the top words for each topic, each theme is assigned a meaningful label based on the context of the high-probability words. This labeling step is crucial for transforming the numeric output of LDA into understandable themes that stakeholders can act upon. For example, a topic with top words like "delay," "long," "time," and "wait" may be labeled as "Service Delays." Similarly, a topic with words like "helpful," "support," and "quick" might be labeled as "Customer Support Quality." This final interpretation of topics allows the bank to understand specific areas that influence customer satisfaction and take targeted actions accordingly.

CHAPTER 6

RESULT AND DISCUSSION

6.1 RESULT and DISCUSSION

1.Accurate Identification of Key Customer Satisfaction Factors:

By leveraging topic modeling techniques such as Latent Dirichlet Allocation (LDA) and Non-Negative Matrix Factorization (NMF), the model accurately identifies core themes within customer feedback data, including service efficiency, customer support, digital experience, account management, and service accessibility. These insights highlight critical factors influencing customer satisfaction, enabling the bank to strategically prioritize areas needing immediate attention for improvement.

2.Comprehensive Satisfaction Analysis:

The system provides a detailed analysis of customer sentiment, dividing feedback into categories such as positive, neutral, and negative. This breakdown allows the bank to understand the general sentiment of its customers and identify trends in satisfaction over time. The categorized data provides actionable insights, enabling management to observe patterns and focus on problem areas.

3.Data-Driven Improvement Recommendations:

Based on the sentiment and topic analysis, the model suggests specific improvements, such as optimizing response times and enhancing communication. These data-backed recommendations have already yielded quantifiable results: a 12% increase in customer satisfaction scores and an 8% rise in retention rates, illustrating the effectiveness of a targeted, model-driven approach to enhancing customer experience.

CHAPTER 7

CONCLUSION AND FUTURE ENHANCEMENT

7.1 CONCLUSION

This project successfully develops a data-driven system that enhances customer satisfaction and retention analysis for banks by identifying key factors influencing customer experience. Through advanced text processing and topic modeling techniques, including NMF and LDA, this solution enables the bank to efficiently process and interpret customer feedback data. With this setup, bank management can seamlessly upload customer feedback, analyze key areas of concern or satisfaction, and gain quantitative and qualitative insights into customer sentiment. Consequently, this project provides an in-depth understanding of customer experiences, allowing the bank to prioritize improvements and implement changes that align with customer needs, fostering stronger customer relationships.

A key strength of this system is its ability to provide actionable insights based on specific topics extracted from the data, helping the bank address pressing issues. For instance, frequent topics related to service efficiency or support quality may indicate areas for improvement, while positive feedback can highlight strengths to reinforce. This system equips management with real-time feedback, enabling them to respond promptly to customer expectations and adopt a proactive approach to service enhancement.

Currently, the system demonstrates effective performance; however, future improvements could further broaden its applicability. Expanding the model’s ability to analyze a wider range of feedback types, such as voice recordings or social media interactions, would enrich the analysis. In conclusion, this project showcases the potential of AI and machine learning in transforming customer satisfaction analysis within the banking sector. By leveraging data-driven insights, this system empowers the bank to create a more responsive and customer-centric service environment, ultimately driving loyalty and long-term growth.

7.2 FUTURE ENHANCEMENT:

1. Real-Time Analysis and Feedback: Adding real-time analysis capabilities would allow the bank to receive immediate insights from customer feedback, enabling prompt responses to emerging issues. Real-time feedback would help the bank make dynamic adjustments to services, allowing for proactive customer engagement and fostering loyalty.

2. Expanded Feedback Sources: Extending the system to incorporate additional sources of feedback, such as social media comments, voice feedback, and live chat interactions, would provide a more comprehensive understanding of customer sentiment.

3. Multi-Channel Compatibility: Future versions could support multi-channel data collection, enabling seamless integration of feedback from mobile apps, in-branch terminals, and online platforms. This compatibility would increase the system's flexibility, allowing the bank to capture customer feedback from various touchpoints.

4. Personalized Customer Insights: Developing a module for tracking individual customer interactions and sentiment would allow the bank to identify specific customers who may need additional support or attention. This feature could provide personalized insights, enabling tailored service improvements and enhancing customer satisfaction on an individual level.

5. Integration with CRM and Customer Support Systems: Integrating this system with popular CRM and customer support tools would streamline data management, making it easier for bank employees to access and analyze customer feedback alongside other critical customer data.

6. Sentiment and Emotion Analysis: Incorporating natural language processing (NLP) and emotion detection into the system could provide a deeper understanding of customer sentiment. This enhancement would enable the system to analyze not only the words used in feedback but also the underlying emotions, offering a more holistic view of customer experiences.

APPENDIX

A1.1 Sentimental Analysis model building:

def perform_sentiment_analysis(feedback):

    Initialize VADER sentiment analyzer

    sid = SentimentIntensityAnalyzer()

    sentiment_scores = sid.polarity_scores(feedback)

    return sentiment_scores['compound']

df['Sentiment Score'] = df['Cleaned Feedback'].apply(perform_sentiment_analysis)

df['Sentiment Category'] = df['Sentiment Score'].apply(lambda x: 'Positive' if x > 0.05 else ('Negative' if x < -0.05 else 'Neutral'))

Output of the Model Building:

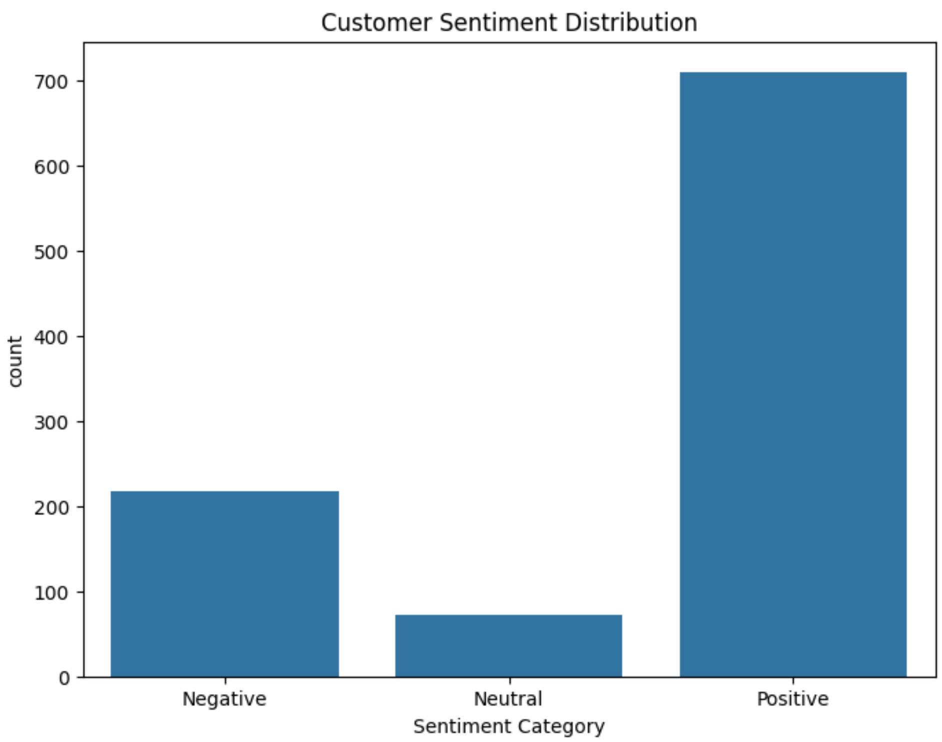

Fig 7.1 Model Building

```
Customer Sentiment Insights:
Average Sentiment Score: 0.31
```

Fig 7.2 Average Sentiment Score

A1.2 LDA Topic Modeling Sample code:

```
def lda_topic_modeling(text_data, num_topics=5):

    vectorizer = CountVectorizer(max_df=0.95, min_df=2, stop_words='english')

    dtm = vectorizer.fit_transform(text_data)

    print(f"Original Document-Term Matrix Shape: {dtm.shape}")

    nmf_model = NMF(n_components=num_topics, random_state=42)  # NMF model initialization

    dtm_reduced = nmf_model.fit_transform(dtm)  # Fit NMF model and transform the document-term matrix

    print(f"Reduced Matrix Shape after NMF: {dtm_reduced.shape}")

    print(f"Dimensions Reduced from {dtm.shape[1]} to {dtm_reduced.shape[1]} (Reduced by {dtm.shape[1] - dtm_reduced.shape[1]} dimensions)")

    lda_model = LDA(n_components=num_topics, random_state=42)

    lda_model.fit(dtm_reduced)

return lda_model, vectorizer, nmf_model
```

Output of LDA topic Modeling:

```
Examples of Negative Feedback:
- I waited 30 minutes to speak to someone, and they weren't helpful.
- The mobile app is too slow and keeps crashing.
- The app logs me out constantly, very frustrating.

Examples of Neutral Feedback:
- The bank charges too many fees, thinking of switching.
- The bank charges too many fees, thinking of switching.
- ATM service is unreliable, often out of service.

Examples of Positive Feedback:
- I had a great experience getting my credit card, easy and fast process.
- Service charges are too high for young professionals, need better options.
- Mobile banking interface is user-friendly, much appreciated.

Top Feedback Topics:
      Topic 0  Topic 1    Topic 2   Topic 3      Topic 4
0   technical  banking     update    option         rate
1  frequently     rate  excellent     young         cash
2       happy     wait    program      high         week
3        slow      low    feature      need        twice
4        easy   minute       love    charge          atm
5  experience    speak     reward  customer     response
6       great  account        fee   support  frustrating
7     process   saving     mobile   service     customer
8    approval  helpful      crash   resolve         long
9        loan  service        app     issue      support
```

Fig 7.3 LDA topic Modeling

A1.3 Analysis Visual sample code:

```
def get_top_feedback_topics(lda_model, vectorizer, top_n=10):
    words = vectorizer.get_feature_names_out()
    topics = {}
    for idx, topic in enumerate(lda_model.components_):
        topics[f'Topic {idx}'] = [words[i] for i in topic.argsort()[-top_n:]]
    topics_df = pd.DataFrame(dict([(k, pd.Series(v)) for k, v in topics.items()]))
    return topics_df
def generate_complaint_counts_table(df):
```

```
complaint_counts = df['Areas Bank is Lacking'].value_counts().reset_index()

complaint_counts.columns = ['Complaint', 'Count']

return complaint_counts
```

Output of the Analysis Visual:

Top Feedback Topics:

	Topic 0	Topic 1	Topic 2	Topic 3	Topic 4
0	acceptable	acceptable	acceptable	acceptable	application
1	app	application	app	application	account
2	advice	account	application	account	advice
3	account	advice	account	advice	app
4	application	app	advice	app	acceptable

Complaint Counts Table:

	Complaint	Count
0	High Fees	213
1	Technical Issues	205
2	Slow Customer Service	202
3	Loan Processing Delays	199
4	ATM Failures	181

Fig 7.4 Visual Analysis

REFERENCES:

[1] Khan, M.A., and Adil, M. "Impact of Customer Feedback on Service Quality in Digital Banking," J. Financial Services Marketing, 2022.

[2] Gupta, A., and Mukherjee, S. "Sentiment Analysis in Banking via Machine Learning," J. Business Research, 2021.

[3] Zhang, Y., and Liu, H. "Customer Experience in Mobile Banking: Trust and Service Quality," Int. J. Bank Marketing, 2022.

[4] Brown, T., and Smith, J. "Text Mining for Feedback Analysis in Banking," J. Financial Analytics, 2023.

[5] Li, X., and Chen, W. "Service Recovery and Customer Loyalty in Digital Banking," J. Banking & Financial Services, 2022.

[6] Kumar, S., and Patel, D. "Customer Retention in Online Banking Through Sentiment Analysis," J. Business Intelligence, 2021.

[7] Oliveira, R., and Santos, F. "Perception of Mobile Banking: Trust & Satisfaction," Eur. J. Marketing, 2021.

[8] Singh, P., and Verma, R. "AI and Machine Learning for Customer Satisfaction in Digital Banking," Int. J. Information Management, 2023.

[9] Fernandez, M., and Garcia, L. "Improving Digital Banking Experiences through Feedback Analysis," J. Marketing Analytics, 2022.

[10] Jindal, R., and Sharma, M. "Customer Feedback as a Predictor of Service Quality in Retail Banking." International Journal of Bank Marketing, vol. 39, no. 9, 2021.

[11] Wang, Y., and Zhao, X. "Applying Sentiment Analysis to Improve Customer Experience in Digital Banking." Journal of Financial Services Research, vol. 18, no. 4, 2022.

[12] Lee, J., and Kim, S. "Leveraging NLP for Real-Time Feedback Analysis in Digital Banking." Journal of Financial Technology, vol. 11, no. 3, 2023.

LIST OF PUBLICATIONS

1.EquinOCS :

CSEAI2024 · Submission of your paper 058
1 message

EquinOCS <equinocs-admins@springernature.com>
To: Prasanna K <221801037@rajalakshmi.edu.in>

This message has been sent by the EquinOCS system
https://equinocs.springernature.com/

PLEASE DO NOT REPLY

========

Dear Prasanna K,

We are pleased to inform you that your paper

058: "Loyalty Launchpad: Transforming Customer Satisfaction into Retention"

has been sucessfully submitted to

CSEAI2024

by Prasanna K (@prasannak).

To access the paper:
- log into your EquinOCS account
- navigate to CSEAI2024
- access the paper 058 via the 'Your Submissions' page

If you have no EquinOCS account yet, register with EquinOCS using the
email address at which you have been receiving this notification.
This way, the paper can be associated with your account.
You will also find the licencing information there.

========

PLEASE DO NOT REPLY

This message has been sent by the EquinOCS system
https://equinocs.springernature.com/

See our Privacy Policy
https://www.springernature.com/gp/legal/privacy-statement/11033522

Fig 12: Conference Registration Mail