CUSTOMER SENTIMENT ANALYSIS USING NLP

A project report submitted in partial fulfillment

of the requirements for the degree of

Bachelor of Technology

in

Electronics & Communication Engineering

by

NAME

21BEC1462 – Vikash Vijayan A
21BEC1625 – Ragul D
21BEC1655 – Niranjana D

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Electronics Engineering,

Vellore Institute of Technology Chennai,

Vandalur-Kelambakkam Road,

Chennai - 600127, India.

April 2025

VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

Declaration

I hereby declare that the report titled "***Customer Sentiment Analysis using NLP***" submitted by me to the School of Electronics Engineering, Vellore Institute of Technology, Chennai in partial fulfillment of the requirements for the award of **Bachelor of Technology** in **Electronics and Communication Engineering** is a bona-fide record of the work carried out by me under the supervision of ***Dr. Latha P***.

I further declare that the work reported in this report, has not been submitted and will not be submitted, either in part or in full, for the award of any other degree or diploma of this institute or of any other institute or University.

Sign: ____

Name & Reg. No.: ______________________________

Date: ____

School of Electronics Engineering

Certificate

This is to certify that the project report titled ***Customer Sentiment Analysis using NLP*** submitted by ***Vikash Vijayan A (21BEC1462), Ragul D (21BEC1625), Niranjana D (21BEC1655),*** to Vellore Institute of Technology Chennai, in partial fulfillment of the requirement for the award of the degree of **Bachelor of Technology** in **Electronics and Communication Engineering** is a bona-fide work carried out under my supervision. The project report fulfills the requirements as per the regulations of this University and in my opinion meets the necessary standards for submission. The contents of this report have not been submitted and will not be submitted either in part or in full, for the award of any other degree or diploma and the same is certified.

Supervisor		**Head of the Department**	
Signature:	....	Signature:	....
Name:	....	Name:	....
Date:		Date:	

Examiner

Signature: ....

Name: ....

Date:

(Seal of the School)

Abstract

Getting understanding of the sentiment of customers is important to businesses looking for ways to improve the customer experience and make strategic decisions. By using Natural Language Processing (NLP), the analysis of customer reviews is possible and transforms unstructured feedback into digestible data. The system successfully identifies customers' emotions, behaviour and pain points through machine learning, sentiment analysis, text analytics. Some of key methods used here are sentiment polarity detection, aspect based sentiment analysis, opinion mining for extracting meaningful insights from a variety of data sets. Furthermore, real time visualization and reporting functionalities enable businesses with simple decision support tools in form of business dashboards, NLP based sentiment analysis is extremely efficient for customer engagement and holds key to a scalable solution for business looking to get ahead with data driven insights.

Acknowledgements

We wish to express our sincere thanks and deep sense of gratitude to our project guide, Dr. Latha P, Professor, School of Electronics Engineering, for her consistent encouragement and valuable guidance offered to us in a pleasant manner throughout the course of the project work.

We are extremely grateful to Dr. Ravishankar A, Dean, Dr. Reena Monica, Associate Dean (Academics) & Dr. John Sahaya Rani Alex, Associate Dean (Research) of the School of Electronics Engineering, VIT Chennai, for extending the facilities of the School towards our project and for his unstinting support.

We express our thanks to our Head of the Department Dr. Mohanaprasad K for his support throughout the course of this project.

We also take this opportunity to thank all the faculty of the School for their support and their wisdom imparted to us throughout the course.

We thank our parents, family, and friends for bearing with us throughout the course of our project and for the opportunity they provided us in undergoing this course in such a prestigious institution.

Contents

List of Figures

Chapter 1

Introduction

Businesses utilize customer sentiment analysis to understand consumer emotions for enhancing both customer service quality and strategic domain decisions. A company gains a superior position in the market by utilizing structured customer opinion assessment to recognize new market directions while fixing vital problems and offering solutions that precisely match consumer demands. Fearful data analysts benefit from the combination of NLP technology and machine learning algorithms which allows them to extract essential insights from big and disordered data sources effortlessly. Modern technology helps companies acquire full insight into customer product assessments and market conditions which leads to the creation of solution plans for industry variations.

The processing of advanced methods converts unstructured customer input into relevant business-generated insights. To identify customer emotions properly there exist three fundamental sentiment analysis techniques: sentiment polarity analysis and aspect-based sentiment analysis and opinion mining. The classification of sentiments through polarity analysis either identifies opinions as positive or negative or neutral thereby revealing overall customer sentiment. A deeper review occurs through aspect-based sentiment analysis which evaluates particular elements of products or services to assist businesses in determining their advantages and development opportunities. Text data evaluation known as opinion mining enables analysts to extract subjective statements which reveal customer preferences and their expectations. Businesses achieve data-driven operational adjustments together with decision-making through real-time data visualization systems that track sentiment trends in dynamic ways.

NLP-based sentiment analysis serves as a revolutionary tool which enhances customer satisfaction through effective proactive customer outreach activities for businesses. Service offerings of companies improve as sentiment analysis solutions allow them to refine services while simultaneously building a positive brand image and market growth. Sentiment analysis uses its pivotal role to shape marketing directions and enhance product development and service quality which results in industry leadership with better brand perception.

Sentiment analysis delivers to businesses deeper insights about customer requirements together with their product preference standards along with their experience-related issues. Companies achieve targeted improvement identification through organization-level sentiment analysis when they explore different corporate aspects including product quality service delivery pricing and brand perception. Sentiment analysis reaches its maximum precision level when artificial intelligence (AI) systems are implemented because AI models detect faint emotional cues while interpreting relationship context and also recognize sarcasm in text. The analysis strengthens through predictive analytics because it enables businesses to foresee customer responses which allows them to act before problems grow out of control.

Organizations leverage sentiment analysis data to establish customer-centric strategic plans which deliver maximum brand loyalty while giving them market competitiveness. The combination of artificial intelligence with natural language processing and machine learning technology continuously improves sentiment analysis methods and brings new growth opportunities for business customer relationships and quick adjustments to consumer expectations. Computing industries continue to identify the worth of sentiment analysis which will drive expanded usage resulting in a more prominent operational place for contemporary enterprises.

Chapter 2

Literature Survey

2.1 A Novel Customer Perception Analysis System Using Natural Language Processing and Attribute Control Charting

This research paper develops an advanced methodology which combines sentiment analysis alongside topic modeling and statistical quality control charts for performing effective customer review assessment. The main purpose exists to discover emerging negative patterns while tracking essential customer complaints that modify satisfaction evaluations. In the regression testing phase the study uses analysis to determine which quality attributes produce specific measurable changes in customer satisfaction scores. The model underwent validation through research conducted on luxury hotel reviews collected from a big dataset. Statistical control methodologies together with natural language processing techniques substantially enhanced the process of detecting essential customer dissatisfaction areas. This strategy enables organizations to access crucial information that leads to better service improvements and improved customer satisfaction outcomes.

2.2 Aspect-Based Sentiment Analysis for Service Industry

Scientists explore how aspect-based sentiment analysis functions for mobile app review assessment in service-oriented businesses. A two-stage rule-based system works together with machine learning methods and fine-tuned BERT models in this proposed method. The model reaches enhanced sentiment classification by applying extra hyperparameter tuning which helps it recognize both direct and indirect sentiments. The assessment quality of customer feedback becomes more robust through added annotation methods. ABSA techniques deliver exceptional value to service-based businesses because they enable tracking of specific service features which influence customer perception. Commercial entities can enhance service quality and customer experience through better sentiment analysis because rule-based systems integrated with machine learning techniques improve accuracy according to research findings.

2.3 Comparative Analysis of Deep Natural Networks and Large Language Models for Aspect-Based Sentiment Analysis

The research study compares deep learning models to large language models (LLMs) in the analysis of aspect-based sentiment (ABSA). A performance evaluation of DeBERTa, PaLM and GPT-3.5-Turbo takes place while using

DATS, MAMS and SemEval16 benchmark datasets. DeBERTa proves itself as the optimal model for ABSA since it delivers better precision and recall metrics. The diverse sentiment-based datasets show that PaLM maintains excellent performance levels across all examined domains. Large language models demonstrate powerful capabilities in sentiment analysis according to the study which also notes that ABSA demands future improvement for dealing with domain-sensitive issues and more accurate contextual sentiment interpretation. The study supports the development process of highly accurate and scalable sentiment analysis technology.

2.4 EEBERT: An Emoji-Enhanced BERT Fine-Tuning on Amazon Product Reviews for Text Sentiment Classification

The research presents EEBERT which represents a sentiment classification enhancement by using BERT model fine-tuning with emoji-based emotional measurements. The Sentiment Adjustment Factor (SAF) within this analysis model distributes emojis through tokenization which strengthens the identification of textual sentiment expressions. Experimental results validate the superb accuracy achieved by EEBERT at 97.00% when it conducts five-fold cross-validation on Amazon product reviews. The model surpasses current sentiment classifiers because it delivers an outstanding 99.21% accuracy rating for the dataset. The research outcomes show that emojis play an essential part in sentiment analysis while providing valuable assistance to improve machine learning algorithms that identify customer emotions.

2.5 Data-Driven Decision-Making to Rank Products According to Online Reviews and the Interdependencies Among Product Features

The paper designs a hybrid decision system which combines sentiment analysis with multi-criteria decision-making techniques for ranking products according to customer reviews. The system applies association rule mining together with fuzzy cognitive maps and interval-valued intuitionistic fuzzy theory for product ranking using the MULTIMOORA ranking algorithm. Through this framework the system produces dependable product rankings because it detects successful relationships between different product elements. The proposed approach receives validation through an analysis of Amazon mobile phone reviews which proves its suitability for complicated decision-making situations. This combined method delivers superior product ranking precision to businesses while providing a data-based customer perception measurement system that optimizes product operational performance.

2.6 Sentiment Analysis on Feedback Data of E-commerce Products Based on NLP

The research examines how natural language processing techniques analyze customer feedback which primarily focuses on Amazon product reviews obtained from e-commerce platforms. Data pre-processing under the NLTK library framework within Python converts reviews into analytical input by creating a regime that produces sentiment analysis output through regression modeling. The developed model demonstrates exceptional performance by reaching an F1-score of 0.87 which signifies its successful sentiment classifying ability. Research-derived consumer preferences together with sentiment analysis plays a key role in helping business leaders understand their customers and enhance their product lines successfully. Advanced NLP methods foster better sentiment detection accuracy which enables e-commerce platforms to deliver increased customer need-responsive service.

2.7 A Deep Learning Framework for Sentiment Analysis in Finance Sector Reviews

The research develops a deep learning framework which analyzes investor reviews together with market sentiments within financial settings. Bi-directional LSTMs and RoBERTa operate as transformer-based architectures within the framework that analyzes sentiment trends from financial discussions. Financial data analysis through the model enables investors to make decisions based on real-time market sentiments that it detects successfully. Analysis of financial market sentiments produces meaningful results that help identify risks while building improved investment plans and revealing market instability before it occurs. AI-powered sentiment analysis tools have gained substantial importance in financial decisions according to research data.

2.8 Hybrid Sentiment Analysis Model for Analyzing Customer Reviews on Travel and Tourism Websites

The research presents a combined model of sentiment analysis through the integration of deep learning techniques and lexicon-based methods for assessing customer comments on travel websites. The research refrains feedback captured from TripAdvisor and Booking.com to understand which service features affect customer satisfaction the most. The hybrid model utilizes linguistic rules and machine learning algorithms to boost the sentiment classification process which results in superior accuracy for positive and negative feedback identification. Tourism service providers can optimize their products based on user feedback through precise sentiment insights which hybrid approaches deliver according to the research findings.

2.9 Enhancing Sentiment Analysis for Customer Support Systems Using Reinforcement Learning

The research suggests implementing sentiment analysis enhancement through reinforcement learning in order to improve customer support methods. A system which combines reinforcement learning models with sentiment classification techniques generates real-time prediction adjustments through customer interactions. The research shows that reinforcement learning decreases sentiment analysis misclassification rates and improves such models for customer service environments. The research demonstrates this technique improves the personalization of responses so automated customer support solutions offer better context-based help. The research outcomes show reinforcement learning provides promising potential for enhancing sentiment-based artificial intelligence systems

Chapter 3

Methodology

3.1 Modules

3.1.1 Web Scraper for Automated Data Collection

Accurate data collection requires a proper review extraction process from the Skytrax website. The application leverages requests library for secure HTTPS connection while BeautifulSoup extracts necessary data that includes user review content along with submission date and geographic origin. A systemized approach enables users to easily navigate through various review sections so they can obtain new data in a reliable continuous manner. A structured data processing system operates on collected data which is illustrated by Fig. 1. The sentiment analysis model processes review data obtained by the web scraping module to perform customer opinion classification. The processed data gets stored in cloud-based data warehouses to allow scalable access for users. The last module of the system displays graphical insights which help users make decisions and analyze trends.

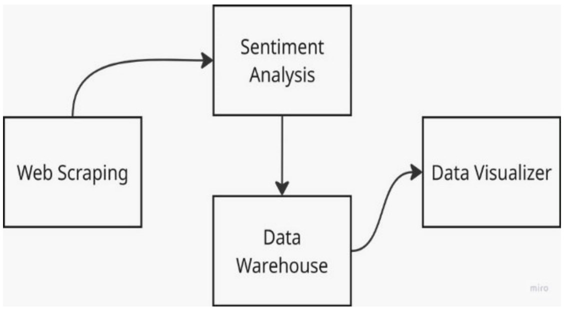

Fig.1: Proposed Model (To Run Locally)

3.1.2 Natural Language Processing (NLP)

The NLP module transforms customer review text structure through multiple operations which enables it to perform sentiment analysis accurately. The module performs essential operations through three preprocessing steps that use Regex proper expressions to cleanse text and tokenize the words followed by part-of-speech tagging to mark sentence grammar. Stop words removal during preprocessing helps eliminate words without bearing any influence on sentiment outcome while lemmatization normalizes word bases. A complete data preprocessing program increases sentiment classification precision and trustworthiness in the results.

3.1.3 Sentiment Scoring

Data preprocessed through sentiment analysis applies VADER [3] NLP model implementation for categorizing sentiments into positive and negative and neutral expressions. The system produces evaluative sentiment ratings which provide air travel service operators with unique customer satisfaction understanding capabilities. The sentiment scoring technique allows organizations to track market data alongside brand-awareness and service quality and spot existing weaknesses. Sentiment score analysis helps airlines find ways to tackle passenger concerns directly while improving all aspects of flight service quality by taking rapid feedback-oriented actions.

3.1.4 Database Management for Structured Data Storage

The organized storage system maintains processed reviews so they can be accessed and analyzed efficiently. The database management module establishes CVS file organization to arrange information through airline categories alongside review dates and user geographic positions. The structured storage system allows researchers and airline service providers to analyze customer feedback through seamless integration with analytical tools. A properly organized dataset enables the airline industry to detect patterns and visualized sentiments due to which they can make data-informed decisions to improve their service delivery. A combination of sentiment analysis with structured storage solutions enables airlines to improve operations while foreseeing customer requirements and stay competitive in aviation services.

3.2 Functionalities

3.2.1 Automated Review Data Scraping

System automation enables the application to retrieve airline reviews from Skytrax website in an efficient manner. The automated web scraper operates independently to pull data from multiple web pages which it combines with the review text content and user data and date information. The automated system makes workload more efficient while maintaining consistent data along with its capacity for gathering big volumes of customer feedback. Automated review extraction enables airlines to maintain ongoing customer opinion tracking and rapid changes to passenger demands without human execution and in real-time for feedback integration.

3.2.2 Text Cleaning and Standardization

Regular expressions (Regex) serve as the tool to achieve text standardization by enabling text cleaning operations. Special characters, symbols together with unnecessary white spaces undergo removal during the cleaning operation to generate a standardized input format. By adopting standardization methods the accuracy of sentiment analysis increases because the processed data becomes better organized for analysis purposes. Text cleaning procedures maintain essential information in the input data which leads to better model accuracy by minimizing unwanted sentiments in classification models. Standardized customer reviews enable companies to compare different airlines accurately for quality benchmarking.

3.2.3 Sentiment Scoring for Deeper Insights

A subsequent stage of sentiment analysis processing classifies customer reviews into three groups which include positive, negative and neutral categories. The process of sentiment scoring grants airline service providers sophisticated customer satisfaction insights through marker trend detection and brand perception evaluation while revealing necessary service improvement locations. The structured scoring method enables organizations to convert customer sentiments into measurable results through quantifiable methods. A thorough grasp of sentiment distributions assists business operations by letting them focus their resources on essential areas while developing marketing plans according to passenger opinion results. Through analysis airlines can use empirical data for systematized decision-making that helps them maintain competitive positions in their evolving sector.

3.2.4 Real-Time Updates and Scheduled Execution

Regular data cleaning operations within a scheduled execution system help maintain fresh and relevant data in the dataset. Regular updates allow the system to continuously track customer sentiment via integrated review publications. By using real-time reporting businesses and analysts gain the capability to track changing passenger feedback patterns thus enabling them to take immediate steps for adjusting their service methods and customer outreach tactics. Through real-time updates airlines detect new problems which they resolve before negative effects can hurt customer satisfaction levels. Scheduled execution integration creates an automated process for maintaining fresh data which enables companies to anticipate industry evolution and better meet changing customer demands.

3.3 Protocols

3.3.1 HTTPS Protocol for Secure Data Transfer

The system protects ethical processes of data extraction through its implementation of the HTTPS (Hypertext Transfer Protocol Secure) protocol. Web scrapers can establish a protected data exchange with Skytrax through HTTPS which encrypts their communications along with protecting transmitted information from unauthorized interception. Data traffic encryption through HTTPS shields information from unauthorized interception while sustaining the integrity of the obtained data. Businesses that use HTTPS standards follow ethical web scraping practices which helps them maintain compliance regarding security and privacy regulations. The integration of HTTPS operates at two levels; it secures data while building trust among users so they can rely on legitimate review acquisition from airline customers.

3.3.2 Regex Protocol for Text Standardization

The fundamental protocol for structuring standardization of extracted textual data which text preprocessing relies on is Regular Expressions (Regex). Text normalization employs Regex as a principle to detect and eliminate irrelevant characters while normalizing formats and symbols which exist in review texts. The application of Regex-based transformations enables the system to transform raw text into structured data that improves NLP model and sentiment analysis efficiency. The method creates standardization of textual data from different review sources which minimizes analytical result discrepancies. Businesses achieve better precision in customer feedback analysis through Regex applications in preprocessing text which enhances sentiment classification quality.

3.4 Data collection approaches

3.4.1 Scraping Reviews

A properly designed Python script uses the requests and BeautifulSoup libraries to extract airline reviews from the Skytrax website. The libraries provide smooth web page interaction which enables automatic extraction of review content. The scraper uses a precise detection mechanism to obtain vital review aspects through identification of review texts found in <h3> tags. The script tracks review timestamps stored in <time> elements through the datetime attribute to generate machine-readable times for better analysis of processing information.

The scraper extracts user location information from elements classified under text_sub_header which links every review to geographic metadata. This specific record provides essential insights into satisfaction patterns across different airline operating areas. The scraper operates through a loop-based operation to move between several pages which streamlines the automatic collection of extensive reviews spanning various airlines. The system applies multiple rounds of data collection to accomplish the building of a detailed review dataset including feedback from various customer experiences. The extraction system retains numerical data in prepared structures to provide an appropriate base for exclusive and rate-based analysis procedures. Data scraping maintains its stability through constant updates which keeps the dataset current with the newest customer feedback in the airline sector.

3.4.2 Data Storage and Management

Structured reviews undergo a successful extraction process that leads to their systematic storage in Amazon RDS (Relational Database Service) while using MySQL as the chosen database management system. The selection of MySQL brings optimized performance for structured datasets because it enables fast indexing and quick query execution. MySQL stands out as an ideal system for processing extended text-based information sets that match airline review datasets.

A careful data organization through the database schema allows accurate query retrieval by implementing key information elements including Review_ID, Airline Name, Review Date, User Location, and Sentiment Score. The methodology offers quick and reliable query execution so research becomes more efficient. The previously extracted reviews proceed to an initial processing phase which operates on Amazon EC2 (Elastic Compute Cloud) instances before entering the database. The cleaning and structuring procedures carried out by these instances serve an essential function that makes data ready for additional analysis.

The RDS system becomes more reliable and resilient through its implementation of automatic backup features which safeguard data from losses. The system features both redundancy features and replication capabilities which enable uncontested data recovery when the system experiences failures. The system reliability and performance receives enhancement through AWS CloudWatch integration which provides constant database operational monitoring. The real-time monitoring system identifies performance bottlenecks early so users can perform proactive optimizations which maintain optimal database response and efficiency. The system ensures both optimal performance and high availability through these implementation measures which makes it suitable for processing large airline review data sets.

3.4.3 Data Processing and Preprocessing

The preprocessing of raw text data includes a thorough process which produces consistent and accurate results for sentiment analysis purposes. The preprocessing workflow has been engineered to improve textual information by making it ready for sentiment classification purposes. The initial step in this workflow depends on Regular Expressions (Regex) for eliminating special characters while cleaning up redundant spaces as well as formatting anomalies. The standardized text improves analytical effectiveness because it removes all unneeded material.

The cleaning process for the initial phase is followed by tokenization and stopword removal functions. The analysis excludes stopwords because universal words like "the" and "and" along with "is" add no valuable insights to sentiment evaluation. The system can study words both separately and in their context after tokenization processes break down the text into its smaller units. The text quality receives improvement through the implementation of normalization approaches combining lemmatization with Part-of-Speech (POS) tagging. An application of lemmatization transforms words into basic forms which creates standardization and the use of POS tagging assigns grammatical types to words to enhance contextual interpretation.

The careful data preprocessing approach structures and optimizes the data to be ready for sentiment analysis. Standardizing the data input before running analysis allows the system to achieve better accuracy combined with reliability in sentiment classification leading to improved customer sentiment insights about airline services. The processed dataset delivers improved performance to sentiment analysis models which in turn produces exact and practical output results.

3.4.4 Scalability and System Optimization

The system addresses future scalability needs because it has been built with performance enhancement along with flexibility in mind. The system employs incremental updates because they ensure that only current reviews are processed for storage purposes. The method eliminates repeated information storage so it enhances resource optimization while reducing excessive data storage needs. The system maintains its efficiency through selective process of new reviews to avoid excessive storage of redundant data.

Load balancing through Amazon EC2 allows organizations to manage changes in workload efficiently. Half of the solid-state drives are either idle or occupied by inactive data so they can become immediately available for real-time data processing purposes during periods of increased activity. The adaptive scaling strategy prevents operational obstructions which allows seamless data operations and supplementing storage needs.

The performance of database queries improves because users create indexes strategically on important MySQL fields. Opportunistic query execution speeds become possible because of index organization for airline names and timestamps thus making real-time trend analysis much more efficient. System performance improves through the implementation of cache systems that decreases the number of database queries resulting in minimized response times. The optimized system operates at high speed and maintains dependable data processing through these efficiency methods which enables it to handle substantial airline review datasets properly.

3.4.5 Ethical Considerations in Data Collection

The ethical compliance standard serves as a basic principle for constructing and operating this system for analyzing airline reviews. The web scraping operations remain within legal parameters by following all terms of service provided by Skytrax for data collection purposes. The scraper contains programmed exclusion features which protect personal information to maintain ethical standards for data collection.

The application uses HTTPS encryption as an extra measure to protect both security and privacy of gathered information. The applied encryption protocol maintains data security transmission while stopping unauthorized access and defending data integrity. The system establishes powerful security protocols which actively fights against data break-ins as well as cyber attacks.

The system maintains ethical code in relation to its proper usage of the collected information. Research on sentiment analysis exclusively uses the airline review database which prevents unauthorized personal information access. Data operations follow legal and moral standards with transparent ethics to deliver important information about airline customer satisfaction. The system maintains its ethical standards to create credible research from airline evaluation data alongside developing trustful observer-client relationships.

3.5 Data analysis approaches

3.5.1 Sentiment Analysis

The systematic methodology of sentiment analysis enables big data strategy to utilize three collapsed labels (positive, negative and neutral) for categorizing customer sentiments. The measurement of customer satisfaction becomes efficient through this classification system which businesses in particular sectors like airlines find beneficial. The same process of lemmatization applies to the review data before classification through sentiment analysis algorithms. The system uses rule-based or machine learning-based natural language processing (NLP) to determine sentiment scores in customer reviews. Customer satisfaction scores function as crucial markers for assessment of customer happiness levels to enable airlines in monitoring service quality improvements while making strategic adjustments. Implementation of a sentiment analysis system with accurate techniques enables organizations to track customer sentiment shifts and detect future service improvement requirements for maintaining high customer satisfaction.

3.5.2 Keyword Extraction

The analysis of customer reviews becomes deeper by performing keyword extraction alongside sentiment assessment of tokenized linguistic data. The analytical process reveals main issues which occur throughout customer feedback allowing airlines to understand popular problems better. Reviews contain frequent terminologies that allow airlines to discover major problems through which passengers express their discomfort about their seating space while also expressing dissatisfaction about service standards and baggage procedures and inflight entertainment. The extracted key terms help airlines direct their improvements to the areas which customers find most important thus leading to increased satisfaction and customer loyalty. Airline operations could benefit from seating improvement measures if review data shows numerous customers expressing seating discomfort concerns. By receiving persistent reports about baggage mishandling airlines must improve their processes which track and handle luggage to lower the occurrence of delayed or missing bags.

3.5.3 POS Tagging Insights

The analysis of customer sentiment becomes more effective through the implementation of Part-of-Speech (POS) tagging because this method shows the grammatical structure in reviews. The approach mainly concentrates its analysis on descriptive linguistic elements that include verbs, adjectives and nouns because these words give direct information about customer encounters. Airlines who study the adjectives in customer feedback can discover which precise attributes these passengers link to their service through specific examples such as "comfortable seats" and "delayed flights" and "friendly staff." Nouns disclose the main targets of customer feedback while both nouns and verbs describe the events linked to airline service delivery. Through linguistic pattern classification companies can systematically enhance their individual service components. Toys with multiple passenger reports about unhelpful or rude behaviors of airline employees can direct their efforts into training staff and improving customer service to boost passenger satisfaction.

3.5.4 Frequency Analysis

Frequency analysis serves as a vital method for customer review study by producing statistical results on the most typically stated expressions found in airline customer feedback. A number of passenger reviews gets analyzed through a word and phrase frequency count which shows airlines which concerns repetitively appear and dominant issues throughout the reviews. The quantitative method enables enterprises to detect essential problem domains that need fixing by studying customer opinion data patterns. Frequent occurrences of specific issues signal future problems to airlines so they can solve them beforehand to prevent bigger affectment of their customer base. The highest occurring phrases in customer reviews mentioning "delayed flights" and "poor food quality" should prompt airlines to investigate these problems for implementing resolution strategies. Employee resource distribution becomes more efficient because businesses understand which customer service areas produce the most frequent negative comments.

3.5.5 Data Modelling

Data modeling serves as an essential method to arrange imported data for better analysis through its organizational and structural capabilities. The process combines separate tables into an integrated schema because it creates a single standardized database for efficient data retrieval. The central Fact Table in Figure 2 connected with five individual table entities representing major airlines Air India, British Airways, Emirates, Etihad and Qatar Airways. This fact table combines Airline, Airport, Country, Customer Review, Site and Date attributes to produce a wide-ranging analysis capability for various airlines. The established relationships in data modeling maintain logical data relationships which guarantee precision and standardization enabling analysts to discover significant insights at a higher pace. This structured data model receives enhanced analytical capabilities through Power BI because it allows effortless data querying and visualization capabilities. This dataset integration produces aggregate findings which create an all-encompassing view of customer sentiment along with service quality patterns for diverse airlines. A properly designed data model helps airlines both monitor evolving customer attitudes and perform competitor service comparisons to generate decisions that improve satisfaction rates.

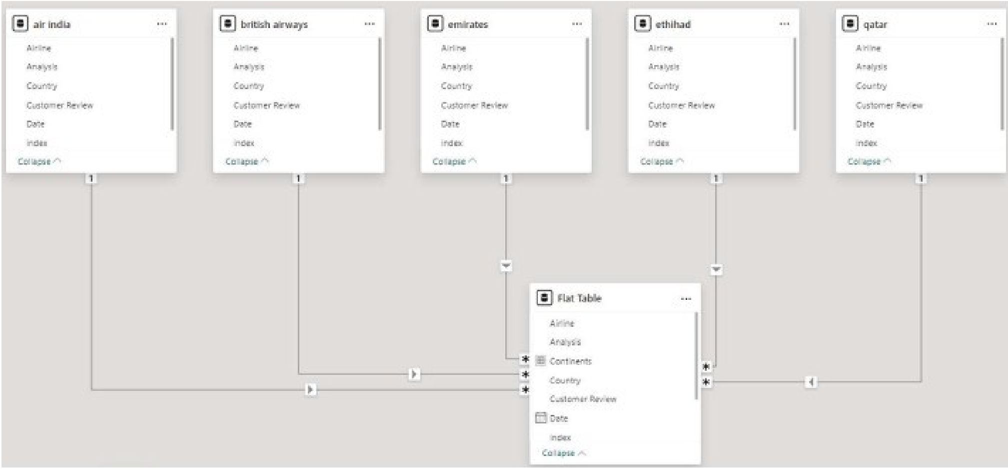

Fig.2 : Tables and calculated columns of all five airlines - relationship

3.6 System model:

3.6.1 Input Layer

The first component in the system framework identifies and retrieves data from Skytrax webpages with help from Amazon EC2 instances and their automation functions. The data harvesting process operates through Python scripts that implement Requests together with BeautifulSoup to fulfill data scraping operations. The Skytrax website review information gets extracted in a structured fashion through scripts which focus on particular HTML elements containing review content. The <h3> tags are used for retrieving review text while <time> tags provide machine-readable datetime attributes that reveal both time and date of each review. The text_sub_header class elements in the HTML code extract user locations accurately for proper data collection. A programmed scheduling system handles both automatic and time-based reviews extraction so human inputs are not necessary. The data collection system continuously obtains new reviews for analysis thus keeping the customer feedback database current.

Amazon CloudWatch operates as part of the system infrastructure to supply real-time monitoring of EC2 instances for the purpose of maintaining performance and reliability during data collection. The instances running in Amazon CloudWatch get inspected for performance which records operational information while scanning for errors that occur during the scraping operations. System administrators can respond swiftly to disruptions by receiving failure alerts from CloudWatch that enables them to keep data extraction consistent. The detailed monitoring system creates a resilient environment that helps reduce system downtime and improve airline review data retrieval performance.

3.6.2 Processing Layer

The collected Skytrax data moves to the processing layer for thorough preprocessing operations until it achieves quality standards for analysis purposes. The automated preprocessing functions operate through Amazon EC2 to process large archival sizes of airline review data efficiently. The first procedural point requires data cleaning which removes duplicate records alongside unnecessary symbols and standardizes inconsistent field formats to ensure proper data accuracy. Regex-based normalization standards allow the system to normalize text so all reviews follow consistent formats during preprocessing. The preparatory step serves to enhance data analysis capabilities because it removes unwanted information while creating a format suitable for sentiment classification methods.

The stage that follows cleaning and standardization within the processing layer is sentiment classification. The predefined sentiment analysis models deployed by the system identify sentiments in customer reviews which they label as either positive or negative or neutral. The analysis uses natural language processing models to determine sentiment classes by examining review content. Airlines can evaluate customer satisfaction through sentiment classification because it reveals their time-based positive and negative review percentages.

Amazon CloudWatch enables the smooth operation and scalable performance of the entire data processing pipeline by serving as its monitoring solution. CloudWatch delivers ongoing visualizations about execution durations along with faults identification and resources consumption assessments. System administrators gain the ability to find processing bottlenecks through this system which allows them to optimize their computational resource utilization. CloudWatch allows the system to operate with high reliability because it ensures efficient data processing of large airline review datasets along with performance stability.

3.6.3 Storage Layer

Amazon RDS (Relational Database Service) using MySQL stores the processed data from airlines and maintains a robust system for structured queries and scalable management of extensive text-based datasets. The airline review data stored in Amazon RDS by MySQL achieves a well-structured format that enables easy data access for analysis purposes. The database schema contains essential data elements which consist of review ID and user location together with review text and timestamp and sentiment scores and rating and sentiment labels. The defined structure helps analysts run advanced queries with speed to obtain detailed information about customer sentiments along with service quality data.

Excel Spreadsheet files operate as mid-data storage between applications to transfer data from Amazon RDS storage. The CSV files act as an interface for data visualization applications and analytical systems to access processed review information for analysis. The system maintains dual storage elements of database structures and CSV files which increases flexibility so customers can execute large data queries in addition to expedited data movement when analysts need external access.

The structured storage plan enables users to monitor customer sentiment patterns that evolved throughout airline history. Employee reviews allow analysts to restore historical data for determining long-term patterns between specific seasons and persistent service problems needing resolution. An efficiently managed storage layer enables airlines to use historical data for prediction analytics thus allowing them to anticipate customer needs before service quality decreases.

3.6.4 Output Layer

The output layer of the data pipeline utilizes Power BI to analyze processed data that has been stored. With Power BI users can develop interactive dashboards and reports which present complete information about sentiment patterns combined with keyword trends and airline service performance metrics. Busienses receive fast access to critical insights through visualized customer sentiment formats which allow them to measure satisfaction and detect service quality weaknesses.

Power BI dashboards present fundamental metrics which show both airline sentiment distributions and keyword frequency patterns and total review changes throughout time. The analysis enables airline management to base their choices on factual data so they can resolve frequent passenger issues while improving journey quality. The continuous tracking of sentiment trends enables airlines to respond immediately to developing customer problems and execute quick corrective actions upon detecting sentiment spikes.

Automatic dashboard updates are one of the main benefits obtained from this automation system for data processing. The Power BI dashboard system gets filled with fresh review data through automatic updates so businesses maintain real-time access to customer sentiment information. The automated processing system requires minimal human intervention to cut down operational mistakes and optimize workflow.

The output layer employs Amazon CloudWatch to track the execution process of data processing. The combination of CloudWatch with its execution time monitoring and error detection capabilities enables resource utilization management to make sure the system runs efficiently at rising data volumes. The system requires real-time monitoring abilities because they contribute to keeping the data pipeline from extraction to visualization both safe and efficient.

The system develops an automated airline review analysis platform through data extraction with Amazon EC2 together with Amazon RDS structured storage alongside Power BI interactive reporting capabilities. The complete solution enables airline companies to discover essential customer satisfaction intelligence which helps them optimize service quality and develop better passenger experiences by making decisions supported by data analysis.

3.7 Cloud automation

3.7.1 Automated Data Collection and Processing

Cloud automation stands as the core implementation that speeds up internet data extraction and processing workflows while keeping operations flexible and mostly human-independent. Amazon Elastic Compute Cloud enables automatic execution of web scraping scripts using Amazon EC2 which periodically collects necessary airline review data from online sources. Automation of this data collection process reduces both time expenditure and manual labor needed for businesses to analyze rather than conduct retrieval operations. The gathered dataset consists of basic airline review information together with timestamps, user location data, sentiment indicators and rating metrics which form fundamental elements for quality evaluation and sentiment analysis.

The performance monitoring of EC2 instances by Amazon CloudWatch operates smoothly through continuous tracking which observes important measurements including CPU usage, memory consumption, and execution time. The monitoring solution in Amazon CloudWatch detects anomalies that could point toward system errors or problems with extraction performance. Failure alerts from the system trigger necessary troubleshooting measures to solve any encountered issues before restarting service operation. Implemented cloud automation capabilities enable the system to operate with three main strengths: high availability and reliability and efficiency when handling large airline review volumes.

3.7.2 Scheduled Execution for Continuous Data Updates

The quality and speed with which customer review information arrives determine the performance level of sentiment analysis systems. The real-time acquisition of airline reviews occurs through scheduled system execution that runs web scraping scripts according to predetermined time frames. The automated retrieval schedules running in the system operate without human involvement to gather new customer reviews for processing.

Automated scheduled execution serves as a vital business benefit since it maintains organizations with up-to-date passenger sentiments. The instant arrival of new review listings at airline review platforms enables airlines to obtain immediate perspectives regarding customer satisfaction because the system performs immediate data analysis. The system enables businesses to detect emerging market trends fast and detect service issues immediately which allows them to modify operations promptly for superior customer service.

Amazon CloudWatch Alarms functions as a monitoring system to detect unusual performance indicators including abnormally long duration of processes and extraction failures and irregular data metrics. Suspicious system activity automatically activates emergency response protocols through the system to resolve every disruption quickly. The synergy between automatic data collection and continuous observation makes the sentiment analysis pipeline more reliable for businesses to base their decisions on concrete data.

3.7.3 Scalable Cloud Storage, Management, and Visualization

Following data processing of airline review information it gets stored within a scalable structured database system that allows quick search capabilities and historical analysis access. The processed data obtains storage within Amazon RDS using MySQL while benefiting from its structured query capability and its scalability features and its efficient data management methods. The review database holds review ID along with user location data and review content and timestamp data and sentiment scores and ratings in essential fields to enable quick analyst data retrieval.

The accessibility of data storage is enhanced through CSV files which serve as supplementary storage capabilities. The files work as a translator to create connections between analytical tools so they function with multiple visualization systems. Data scientists along with analysts can efficiently import structured review data to their preferred exploration tools through CSV format.

Data visualization together with interactive data analysis is powered by Google Looker Studio as a robust business intelligence tool. Real-time analysis in Looker Studio occurs through dynamic dashboards that exhibit trends in sentiment along with keyword distributions as well as airline service reviews and customer feedback patterns. The dashboard presents data in a user-friendly manner to enable business executives along with airline management and customer experience teams to conduct airline comparisons and spot improvement possibilities and enhance service quality.

The system produces valuable data for business applications and research through its storage of relational database data while using modern visualization tools that maintain organized and accessible information. The collected information serves stakeholders for following market trends and passenger satisfaction measurement which produces strategic decisions that improve customer satisfaction.

3.7.4 Enhancing Research and Scalability Through Cloud Automation

Large-scale sentiment analysis finds essential support from cloud automation because it ensures dynamic scalability and maintains system performance when data volumes expand. Dynamic resources adjustments through cloud-based infrastructure scaling creates an automatic response that modernizes resource allocation power as reviews enter the system. Such design ensures continuous operational sustainability because it adequately handles bigger data volumes while preserving system performance.

The main benefit of cloud automation consists of its ability to execute real-time trends analysis amongst multiple airline organizations. Businesses can develop more profound insights about customer satisfaction development while maintaining service quality benchmarks through continuous benchmarking across multiple airlines in real-time and historical sentiment data. The automatic capacity adjustment according to data processing requirements allows business operations to reach maximum efficiency and minimize costs simultaneously.

The deployment framework shown in Fig. 3 (Deployment Pipeline Architecture) improves research functionality through superior analysis of passenger emotional trends. Airline businesses gain better service quality alongside improved customer experience through strategic decisions which stem from the deployment of this system.

The system architecture adopts scalability which allows it to adjust to changing data requirements so researchers and businesses obtain a dependable tool for sustained sentiment analysis.

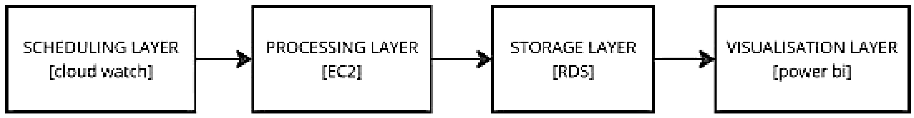

Fig.3: Deployment pipeline for Cloud Automation

3.7.5 Reliable Data Backup and Archival Using Amazon S3

The Amazon S3 (Simple Storage Service) functions as a backup layer of secondary storage in the cloud-based system for the purposes of long-term data preservation and disaster recovery and fault tolerance. The S3 framework included in the architecture preserves both data safety and stability when system failures occur as well as when users delete data or damage databases.

The system implements regular backup procedures to safely store processed review datasets together with sentiment analysis outputs and CSV exports in Amazon S3 buckets. The system incorporates data redundancy which enables users to obtain access to previous datasets after Amazon RDS primary storage fails thus preventing complete data loss.

Amazon S3 serves as a dependable disaster recovery platform that enables business storage capabilities to maintain historical data for extensive analysis purposes. The extended storage of passenger sentiment data through Amazon S3 storage facilities allows businesses and researchers to execute time-based studies and improve sentiment analysis algorithms by making use of historical assessment data.

The data security of Amazon S3 improves through the implementation of Versioning which enables businesses to retrieve earlier dataset states in case of need. This beneficial function allows users to retrieve previous dataset versions to protect their data from corruption and accidental modifications and processing errors.

The Amazon S3 Lifecycle Management Policies offer an economical way to retain important historical data. Advanced S3 Glacier archiving (low-cost archival storage) automatically assumes responsibility for less actively used data after its initial presence in standard storage classes. The data storage costs are minimized alongside access to important historical datasets through this technique.

Systems that use Amazon S3 provide businesses with better scalability and enhanced security and reliability which makes sentiment analysis work continuously through data-driven approaches that serve present-time decisions and future strategic research.

3.8 Sentiment analysis

3.8.1 Sentiment Analysis for Airline Customer Reviews

The analysis of sentiment helps to measure customer satisfaction through computational study of airline review content. Research teams together with airline service providers benefit from this method to extract passenger experience information while detecting specific satisfaction and dissatisfaction aspects. Sentiment analysis techniques help classify customer feedback into positive sentiments as well as negative ones and neutral feedback categories. Overall customer sentiment receives structure through this classification system while airline companies use it to understand the factors which affect their passenger satisfaction.

Numerous passengers post descriptive text reviews on the airline review platforms that focus on flight comfort together with in-flight services staff behavior baggage handling ticketing and general travel satisfaction. Extensive textual data cannot be properly analyzed by manual methods that is why automated sentiment analysis techniques become essential. Original text: Efficient processing of large datasets occurs thanks to natural language processing methods in sentiment analysis algorithms which extract important insights from customer reviews. Specific insights from customer reviews enable airline businesses to develop strategies which improve service quality while allowing them to monitor service level performance to identify areas needing enhancement for improved overall passenger satisfaction.

The opinions of airline customers create immediate effects on both their brand recognition and their customer loyalty levels. Positive customer feedback stems from positive experiences of friendly staff and smooth check-in procedures together with exceptional in-flight services but negative feelings emerge because of delayed flights and poor service and seating discomforts. Neutral sentiments in reviews require additional textual assessment to understand their meaning because these reviews lack clarity. Sentiment tracking throughout time allows airlines to base their decisions on data which enhances service quality through better retention of customers while improving their market posture.

3.8.2 Implementation of Sentiment Analysis Models

This research adopts VADER (Valence Aware Dictionary and sEntiment Reasoner) sentiment analysis model to achieve accurate and efficient sentiment classification. VADER stands as a rule-based model which specializes in social media content sentiment analytics thus delivering exceptional customer review analysis. The technique excels at handling ordinary writing patterns since these reviews typically combine standard abbreviations together with emoticons as well as unconventional capitalization and punctuation marks found in user-created feedback.

Data preprocessing stands as the initial step in sentiment analysis to make text information ready for efficient sentiment classification. The first essential preprocessing function requires lemmatization which converts words into their basic dictionary forms. During this step the analysis identifies "delayed" and "delaying" and "delays" as related terms with their base meaning of "delay." Standardization of linguistic structures in customer feedback leads to superior sentiment classification precision.

Once preprocessing is completed, the VADER sentiment analysis model assigns sentiment polarity scores to each review. VADER uses a lexicon-based approach, where words are assigned predefined sentiment intensities. It considers factors such as word negations, intensity modifiers, and punctuation emphasis to produce more nuanced sentiment classifications. Based on computed polarity scores, reviews are categorized as:

- **Positive sentiment** (high positive score) – Indicates favorable customer experiences.
- **Negative sentiment** (high negative score) – Reflects dissatisfaction with airline services.
- **Neutral sentiment** (score near zero) – Represents mixed opinions or unclear sentiments.

The implementation of sentiment analysis using VADER provides airlines with a scalable solution to evaluate customer feedback efficiently. By automating the sentiment classification process, airlines can systematically analyze thousands of reviews without requiring manual interpretation.

3.8.3 Aggregation and Trend Analysis

The next process following sentiment classification focuses on combining sentiment scores from multiple reviews to generate broader assessment insights. Research teams aggregate individual customer reviews but combine them into time-based average sentiment scores and distribution patterns for discovering satisfaction trends.

The overall customer sentiment measurement for each airline becomes possible through the airline level sentiment aggregate process which serves for airline performance comparison. The study of weekly or yearly sentiments expressed by customers in feedback reveals long-term service quality developments for airlines. Successful service enhancements can be detected by an expanding trend in positive sentiment scores while decreasing scores indicate customer dissatisfaction is starting to develop.

Trend analysis enables airlines to conduct seasonal sentiment evaluation through which they measure customer sentiment shifts across travel seasons including peak holiday periods and operational peaks. Airlines utilize this understanding to select effective service approaches along with forecasting larger customer loads and solving customer service-related issues.

The analysis of customer sentiment helps companies perform competitor benchmarking initiatives. Companies can determine their competitive advantages and weak points through sentiment score comparison with their aviation competitors. The comparative analysis enables airlines to foster strategic choices that lead them to focus on resolving specific customer problems to create distinctive market positions.

3.8.4 Validation and Accuracy Assessment

To ensure the reliability of sentiment analysis results, rigorous validation and accuracy assessments are performed. Automated sentiment classification must be compared against human-labeled data to evaluate its correctness and precision.

A sample dataset of airline reviews is manually annotated by human evaluators, who assign sentiment labels based on their interpretations of customer feedback. These manually labeled reviews serve as a **ground truth dataset**, against which the automated sentiment classification results are compared. The validation process involves assessing standard performance metrics such as:

- **Accuracy** – The percentage of correctly classified reviews.
- **Precision** – The proportion of correctly predicted positive/negative reviews out of all predicted cases.
- **Recall** – The ability of the model to correctly identify positive or negative reviews.
- **F1-score** – A balanced measure combining precision and recall.

Research evaluation of these metrics ensures that the sentiment analysis model operates efficiently with high reliability thus decreasing misclassification possibilities. Processing alterations together with sentiment dictionary and model parameter adjustments are common steps needed to correct assessment discrepancies.

Sentiment analysis provides essential benefits to regulatory monitoring purposes throughout the airline sector. The monitoring bodies responsible for airline service standards should exploit sentiment analysis data to discover large-scale service problems while monitoring general service patterns for better passenger protection. Airline organizations which incorporate sentiment analysis in their experience research obtain important data for enhancing services while planning their strategy and staying ahead of competitors.

The analytics system for sentiment provides airlines with essential tools to optimize their customer service management process. The analysis of customer feedback results in data-based decisions which enables airlines to foresee upcoming passenger expectations while continuously delivering better services. Airlines benefit from automated sentiment evaluation to extract insights from customer feedback which provides strategic direction needed for sustained industrial success within the aviation sector.

3.9 Dashboard preparation:

Power BI delivers a tool that research teams use to perform sentiment trend assessments and review pattern analyses while conducting airline comparison assessments using interactive visual dashboards. Fig.4 demonstrates the Power Bi flow. Dashboards operate through real-time data collection from cloud storage to provide stakeholders and researchers with appropriate airline performance evaluation. The Time-Series functionality within the dashboard shows how customer satisfaction modifies across various airlines alongside their sentiment development patterns. Through word clouds combined with trend charts the review patterns become visible to show what passengers think about and prefer during their travel. Complete airline performance assessments become achievable by inspecting key indices which link total sentiment trends with major positive and negative content themes and geographical sentiment patterns. Organizations utilize airline service comparisons to understand their best features together with necessary improvement zones. Ariana can customize its services according to the sentiment distribution patterns in different regions. Power BI dashboards get continuous updates through their cloud-database relationship with AWS RDS which provides automatic refreshes of information with live access capabilities. When presenting customer reaction data in an organized manner these dashboards allow strategic decisions to emerge from data-based decision making. Three different comparison graphs that represent the dashboard appear in Fig.5,6,7. Executive teams collaborate with policymakers plus analytics experts to use this input for improving service delivery and developing passenger products through assessments of competition performance. Sentiment analysis dashboards powered by advanced visual analytics technology functions as a solution which enhances airline quality services and improves passenger satisfaction results. Users benefit through Power BI because diverse devices enable them to access these insights to conduct real-time monitoring operations in flexible ways. The aviation industry requires perfect integration between data visualization and real-time analytics and sentiment intelligence to enhance service capabilities and establish strong marketplace positions.

Fig.4: Power BI flow

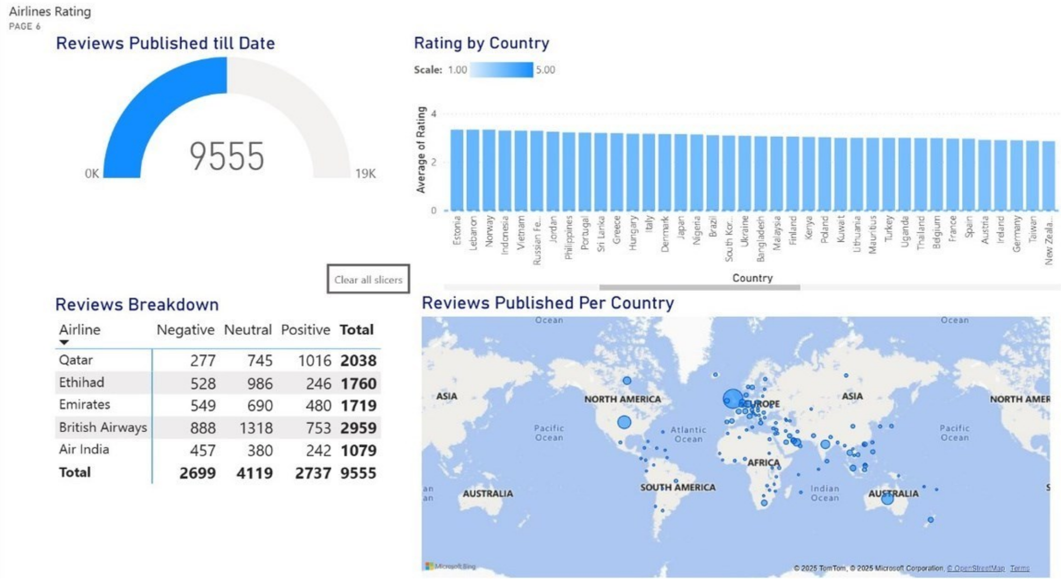

Fig.5: Dashboard outputs - gauge and bar chart, bubble chart (map)

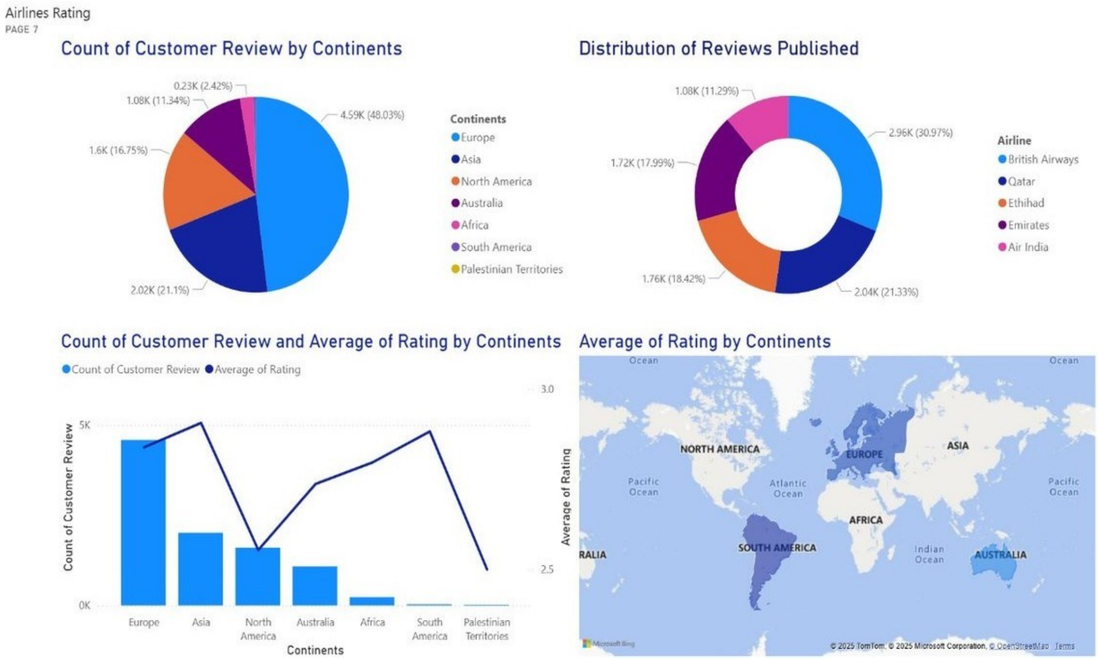

Fig.6: Pie chart, Donut chart, bar-line graph, map - review by continents

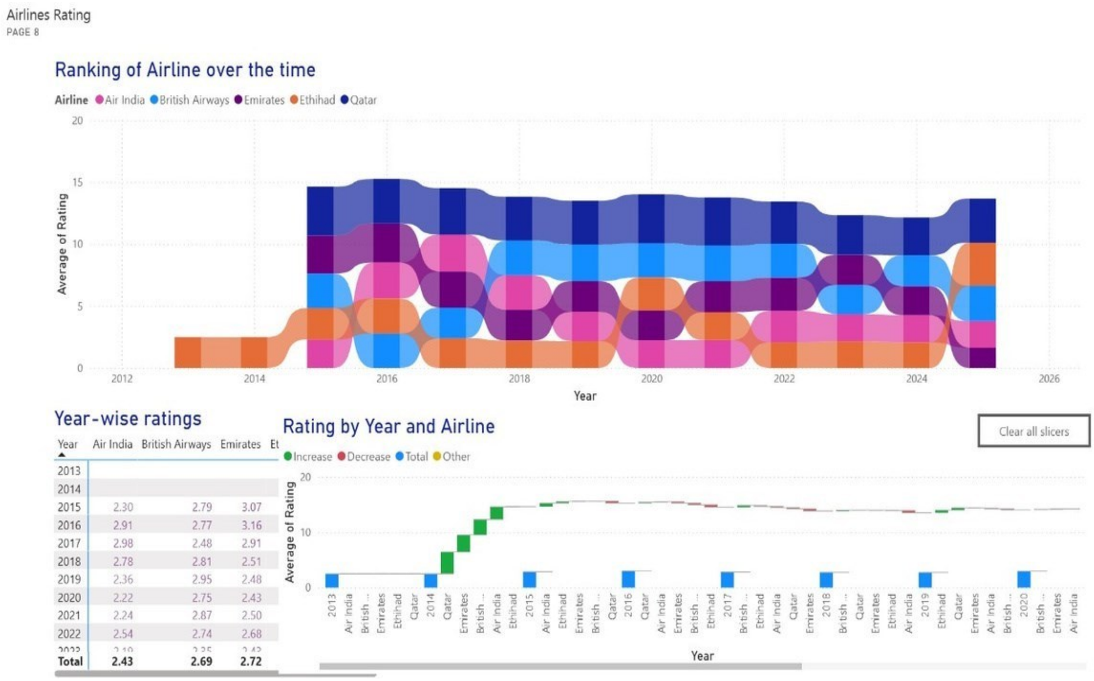

Fig.7: Ribbon chart and waterfall chart - year wise ratings

Chapter 4

Results and Discussions

Objective

The output layer of the data pipeline utilizes Power BI to analyse processed data that has been stored. With Power BI users can develop interactive dashboards and reports which present complete information about sentiment patterns combined with keyword trends and airline service performance metrics. Business receive fast access to critical insights through visualized customer sentiment formats which allow them to measure satisfaction and detect service quality weaknesses.Power BI dashboards present fundamental metrics which show both airline sentiment distributions and keyword frequency patterns and total review changes throughout time. The analysis enables airline management to base their choices on factual data so they can resolve frequent passenger issues while improving journey quality. The continuous tracking of sentiment trends enables airlines to respond immediately to developing customer problems and execute quick corrective actions upon detecting sentiment spikes..

The output layer employs Amazon CloudWatch to track the execution process of data processing. The combination of CloudWatch with its execution time monitoring and error detection capabilities enables resource utilization management to make sure the system runs efficiently at rising data volumes. The system requires real-time monitoring abilities because they contribute to keeping the data pipeline from extraction to visualization both safe and efficient.

4.1 Results and Comparison

The analysis provides insights into customer sentiment towards five selected airlines, along with geographical distribution of reviews. Additionally, the comparison includes sentiment-based ratings and official ratings from Skytrax to validate the accuracy of sentiment analysis. Fig. 8 shows the airline rating comparison.

Airlines' Overall Ratings

The sentiment analysis results indicate varying customer perceptions of the five airlines. The ratings derived from customer sentiment analysis are as follows:

- Air India: 2.46
- British Airways: 2.69
- Qatar Airways: 3.59
- Emirates: 2.74
- Etihad: 2.42

Fig.8: Airlines Rating comparison

Qatar Airways holds the highest sentiment rating, reflecting a predominantly positive customer experience. Evaluation results show that the traveling public has positive impressions of the services they use. The customer satisfaction levels for these airlines are lower according to ratings which indicate that Air India and Etihad register the most negative sentiment scores. The sentiment grading system functions as a vital tool to locate areas in service that need improvement as well as service gaps..

4.1.1 Geographical Distribution of Reviews

The analysis also provides insights into the global reach of these airlines by examining the geographical origins of customer reviews. The top five contributing countries for customer reviews are:

United Kingdom: 3,002 reviews

United States: 1,167 reviews

Australia: 947 reviews

India: 439 reviews

United Arab Emirates: 324 reviews

Fig.9 shows the positive, negative and neutral reviews for all the airlines. The research demonstrates passengers from the UK, US and Australia produce most of the airline reviews online. The review distribution information enables airlines to identify their main customer markets thereby informing improvements to regional services and marketing plan development. The airline receives feedback from UAE and Indian passengers which indicates expanding business prospects in these territories.

Airline	Negative	Neutral	Positive	**Total**
Qatar	242	664	956	**1862**
Ethihad	491	1033	236	**1760**
Emirates	516	654	466	**1636**
British Airways	834	1272	720	**2826**
Air India	401	341	225	**967**
Total	**2484**	**3964**	**2603**	**9051**

Fig.9: Negative, neutral and positive ratings of all airlines.

4.1.2 Sentiment Distribution Across Airlines

The general public perception was analyzed through customer review classifications of positive sentiments alongside negative and neutral sentiments. The classification system makes it possible to detect typical passenger experiences. .SplitContainer analyses the proportions of positive and negative customer feedback to determine how much passengers like their experience and reveals potential service quality and customer support and flight experience weaknesses.

Airline management teams can detect growing passenger concerns through sentiment-tracking which allows them to develop fast solutions for improving customer satisfaction. Extraction of regular key phrases from negative reviews works as a simplistic method for tracking concerns like baggage management and flight disruptions as well as onboard staff offerings. Fig.10,11 demonstrates the ratings from the website, a systematic comparison was conducted between our derived ratings and the ratings displayed on the Skytrax website.

Airline	Rating out of 5
Ethihad	2.42
Air India	2.46
British Airways	2.69
Emirates	2.74
Qatar	3.59

Fig.10: Average Ratings of airlines from the sentiment Analysis out of 5

Fig.11: Star Ratings from the website

4.1.3 Comparison with Existing Ratings

The reliability assessment of sentiment-based ratings involved conducting a comparison with Skytrax ratings. The verification procedure required collecting Skytrax ratings from airlines then comparing them against results obtained through customer sentiment measurements. The main goal was to assess the rating agreement between evaluative methods while verifying the correctness of our sentiment analysis system.

The reliability of sentiment-based ratings matched the official Skytrax ratings which demonstrates sentiment analysis to be an effective airline performance evaluation tool. The validation process allows increased confidence in sentiment analysis as a performance evaluation tool while providing a different option to traditional rating systems.

4.1.4 Dashboard Visualization for Sentiment Analysis

To present these findings in an interactive and accessible manner, Power BI dashboards were developed. These dashboards provide real-time visualization of sentiment trends, keyword frequencies, and airline performance metrics. Some of the key features include:

- Sentiment Overview: Breakdown of positive, negative, and neutral customer reviews
- Trends Over Time: Monitoring shifts in customer sentiment across different periods
- Keyword Analysis: Identifying commonly mentioned terms associated with positive or negative reviews

For cloud-based reporting and automation, Google Looker Studio was implemented. This tool enables seamless integration with Google Cloud services, ensuring that reports remain automatically updated with the latest customer feedback. Google Looker Studio is particularly useful for projects that require cloud-based automation, making it an ideal solution for real-time sentiment tracking and airline performance monitoring.

4.1.5 Cloud-Based Automation and Monitoring

Amazon CloudWatch functioned as a built-in component of the system to manage efficiently the analysis of large customer feedback data sets. The monitoring features of CloudWatch in the system involve real-time execution monitoring, error notifications, and resource utilization tracking. These features make sure that sentiment analysis processes are not disrupted, scalable, and performance-enhanced. CloudWatch's comprehensive logging and performance metrics enable developers and operations teams to monitor every step of the data processing pipeline, minimizing the time taken for troubleshooting and increasing operational transparency.

Executive monitoring of sentiment analysis operations using CloudWatch allows the system to identify operational bottlenecks and optimize efficiency. Sentiment analysis results are aided by in-built error detection systems that maintain both accuracy and reliability of data throughout the analysis process. Integration offers automatic delivery of insights to airline management, who require minimal operational intervention to obtain real-time data. Automated notifications and personalized dashboards keep stakeholders up to date without requiring technical knowledge, facilitating a faster decision-making process.

Structural automation facilitates effective monitoring of customer opinions which enables airlines to identify upcoming trends and facilitate data-driven decisions that enhance overall passenger satisfaction. By identifying problems such as flight delays, customer service issues, or comments on amenities early on, airlines can act proactively to solve issues before they become major problems. The system not only collects and categorizes sentiment information, but also ties it to flight routes, time periods, and even individual service personnel, providing a detailed snapshot of performance and customer opinion.

Power BI is the answer to local dashboards due to its capability for robust visualization and easy integration with local data sources. As it can link data from Excel, on-premises SQL databases, and other enterprise applications, Power BI enables offline analysis and offers dynamic filtering features. Analysts are able to create personalized reports centered on route performance, customer loyalty measures, or patterns of issue recurrence—presenting strategic findings in boardroom presentations or operations reviews.

Google Looker Studio (Fig. 12) is the choice for cloud-based reporting and automation because its automation process relies solely on Google Cloud services. Google Looker Studio provides reporting in the form of efficient updates in real time, which is appropriate for cloud-based projects. The tool supports collaborative report construction and effortless sharing with stakeholders worldwide, making it suitable for global airline teams working with multiple markets. When combined with cloud-based sentiment analytics platforms, it facilitates automated visual narratives of customer trends, heightening executives' and operational teams' clarity. This combination of real-time tracking and dynamic reporting makes sure that airlines are responsive, customer-centric, and continuously evolving as per market feedback.

In addition, the system is designed to incorporate feedback from multiple channels such as social media, customer service chat logs, mobile app reviews, and in-flight survey responses. Through this omni-channel integration, sentiment analysis offers a complete view of passenger satisfaction at every touchpoint within the passenger experience. Airline companies can associate passenger emotions with particular stages of travel—booking, check-in, boarding, in-flight, and post-arrival—enabling targeted enhancements. This holistic view enables more personalized services, such as tailored loyalty offers or route-specific service adjustments.

A second transformative advantage is predictive modeling and long-term planning. By integrating real-time sentiment with historical patterns, airlines can predict potential service problems and plan proactive solutions. For instance, if negative sentiment regarding baggage delays increases steadily on specific routes, the system can identify this trend and initiate operational interventions. In the long run, this predictive function leads to continuous service improvements and enhanced brand reputation, ultimately driving customer retention and loyalty performance.

Moreover, the combined application of Power BI and Google Looker Studio enables layered delivery of insight—local and strategic. While Power BI enables ground-level operational teams with real-time visibility into service metrics and issue resolution timelines, Looker Studio enables leadership with high-level overviews and cloud-integrated strategic dashboards. This two-reporting framework increases alignment between departments and guarantees that every level of the organization can take action on the same unified data source, closing the gap between operational implementation and executive strategy.

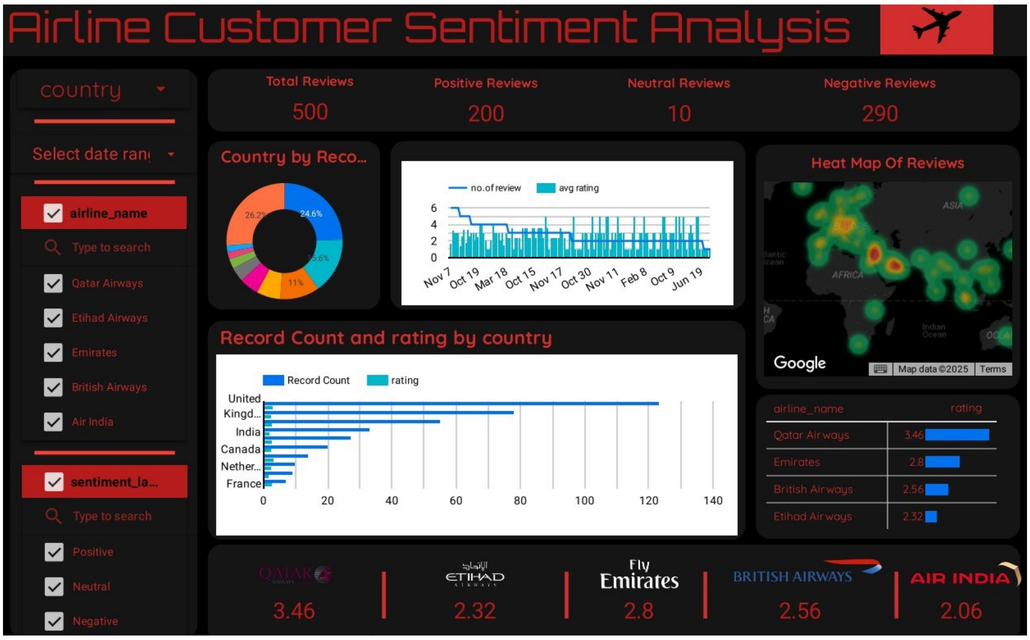

Fig.12: Google looker studio - Dashboard output

Chapter 5

Conclusion and Future Scope

Conclusion and Future Scope

A cloud-based automated sentiment analysis system deploys pioneering innovations that augment the customer review extraction and storage feature as well as analysis processing. The system employs NLP methods to determine sentiments which lead to categorized results of positive, negative, and neutral that inform business comprehension of customer feedback. Data retrieval effectiveness gets a boost through AWS-based automation which lessens manual effort and enhances workflow reliability. Error-handling methods ensure continuous data collection irrespective of site design or server issues that may lead to interruptions. Sentiment classification systems employ precision and recall and F1-score measurement metrics to confirm correctness while mirroring genuine customers' sentiments. The derived business intelligence offers data-based operational decisions that construct improved customer services and functions as proactive feedback responses.

For large-scale deployment support, system design is made scalable and resilient. It can seamlessly integrate with customer relationship management (CRM) platforms and big data pipelines so that sentiment insights are always up to date and readily available. Business managers can easily visualize sentiment trends in real-time using built-in dashboard tools and identify important areas of feedback at a glance. This visibility ensures that issues can be caught before they become problematic and enables more strategic decision-making. Additionally, the modular nature of the platform guarantees that it can be tailored to a particular industry, for example, e-commerce, healthcare, or finance, each of which can have its own sentiment indicators and customer expectations.

In the future, the accuracy of sentiment can be improved through further development by incorporating deep learning models along with real-time monitoring features. The analysis platform can gain deeper customer understanding by adding the capability to process several forms of data that contain both review audio and images. Multimodal sentiment analysis—examining visual indicators, voice, and text context at the same time—can give a more comprehensive perception of customer sentiment. In addition, adding multilingual support will enable companies to send their sentiment analytics globally, leveraging input from various bases of customers. Continuous framework enhancement will enhance its level of reliability so that companies can proactively handle customer issues to be competitive.

Data protection and security are key concerns when designing this sentiment analysis platform. As customer comments tend to involve sensitive data, the platform makes use of end-to-end encryption and complies with data privacy regulations like GDPR and CCPA. Role-based access controls along with audit logging are used in order to keep a check on accountability and protection against unauthorized data access. Besides, anonymization methods strip personally identifiable information (PII) from data sets prior to analysis, which maintains user privacy while still facilitating in-depth insight.

The platform also provides doors for predictive analysis by correlating sentiment trends to business performance measures like sales, churn rates, or customer lifetime value. With machine learning models trained on past sentiment and outcome data, companies can predict customer behavior and act ahead of time. For instance, a sudden spike in negative sentiment surrounding a particular product feature might set off automated warnings and trigger product team reviews or customer service contact. By converting raw customer opinion into strategic insight, the system not only responds to opinion, but assists in shaping the future of customer experience efforts.

Chapter 6

Appendix

```
import requests
from bs4 import BeautifulSoup
import pandas as pd
import mysql.connector
import re
import time
import nltk
from nltk.sentiment.vader import SentimentIntensityAnalyzer

nltk.download('vader_lexicon')

# Amazon RDS Database Configuration
rds_host = 'database-2.c8xg22su41px.us-east-1.rds.amazonaws.com'
rds_user = 'admin'
rds_password = 'admin123'
rds_database = 'airline_reviews'

# Establish RDS connection
def rds_connection():
    return mysql.connector.connect(
        host=rds_host,
        user=rds_user,
        password=rds_password,
        database=rds_database
    )

# Airline URLs
airlines = {
    'Air India': 'https://www.airlinequality.com/airline-reviews/air-india/',
    'British Airways': 'https://www.airlinequality.com/airline-reviews/british-airways/',
    'Qatar Airways': 'https://www.airlinequality.com/airline-reviews/qatar-airways/',
    'Emirates': 'https://www.airlinequality.com/airline-reviews/emirates/',
    'Etihad Airways': 'https://www.airlinequality.com/airline-reviews/etihad-airways/'
}
```

```
# Scrape reviews function

def scrape_airline_reviews(airline, url, pages=3):
    reviews_list = []
    headers = {
        "User-Agent": "Mozilla/5.0 (Windows NT 10.0; Win64; x64) "

"AppleWebKit/537.36 (KHTML, like Gecko) "

"Chrome/122.0.0.0 Safari/537.36"
    }

    for page in range(1, pages + 1):
        print(f"Scraping {airline} - Page {page}")
        response = requests.get(f"{url}page/{page}/", headers=headers)
        print(f"Status Code: {response.status_code}")
        if response.status_code != 200:
            print(f"+ Failed to fetch {airline} - Page {page}")
            continue

        soup = BeautifulSoup(response.text, "html5lib")
        review_divs = soup.find_all("div", class_="text_content")
        print(f" Found {len(review_divs)} reviews on page {page}")

        h3_dates = soup.find_all('time')
        h3_country = soup.find_all(class_="text_sub_header userStatusWrapper")

        country_list = []
        for tag in h3_country:
            text = tag.get_text()
            country_match = re.search(r'\((.*?)\)', text)

country_list.append(country_match.group(1) if country_match else "Unknown")

        dates_list = [h3.text.strip() for h3 in h3_dates]

        for idx, div in enumerate(review_divs):
            review_text = div.get_text(strip=True)
            country_text = country_list[idx] if idx < len(country_list) else "Unknown"
            date_text = dates_list[idx] if idx < len(dates_list) else "Unknown"

            reviews_list.append({
                "Airline": airline,
```

```
                "Review_Date": date_text,
                "Review_Text": review_text,
                "Country": country_text
            })
        time.sleep(1)
    return reviews_list


# Perform Sentiment Analysis + Add Rating
def analyze_sentiment(df):
    sid = SentimentIntensityAnalyzer()
    sentiments = []
    labels = []
    ratings = []

    for text in df['Review_Text']:
        sentiment_score = sid.polarity_scores(text)['compound']
        sentiments.append(sentiment_score)

        if sentiment_score > 0.05:
            label = 'Positive'
            rating = 5
        elif sentiment_score < -0.05:
            label = 'Negative'
            rating = 1
        else:
            label = 'Neutral'
            rating = 3

        labels.append(label)
        ratings.append(rating)

    df['Sentiment_Score'] = sentiments
    df['Sentiment_Label'] = labels
    df['Rating'] = ratings
    return df


# Insert Data into RDS MySQL if NOT EXISTS
def insert_to_rds(df):
    conn = rds_connection()
    cursor = conn.cursor()
    cursor.execute('''
        CREATE TABLE IF NOT EXISTS airline_reviews (
            id INT AUTO_INCREMENT PRIMARY KEY,
            airline_name VARCHAR(255),
            review_date VARCHAR(255),
            review_text TEXT,
            country VARCHAR(255),
            sentiment_score FLOAT,
            sentiment_label VARCHAR(50),
```

```
        rating INT
    )
''')

inserted_count = 0

for _, row in df.iterrows():
    # 🟩 Check if this review already exists
    cursor.execute('''
        SELECT COUNT(*) FROM airline_reviews
        WHERE airline_name = %s AND review_date = %s AND review_text = %s
    ''', (row['Airline'], row['Review_Date'], row['Review_Text']))
    result = cursor.fetchone()

    if result[0] == 0:
        # 🟩 Insert new review with rating
        cursor.execute('''
            INSERT INTO airline_reviews
            (airline_name, review_date, review_text, country, sentiment_score, sentiment_label, rating)
            VALUES (%s, %s, %s, %s, %s, %s, %s)
        ''', (row['Airline'], row['Review_Date'], row['Review_Text'], row['Country'],
row['Sentiment_Score'], row['Sentiment_Label'], row['Rating']))
            inserted_count += 1
        else:
            print(f"⚠ Duplicate found. Skipping review for {row['Airline']} on {row['Review_Date']}")

    conn.commit()
    print(f"🟩 {inserted_count} new reviews inserted successfully.")
    cursor.close()
    conn.close()

if _name_ == "_main_":
    all_reviews = []
    for airline, url in airlines.items():
        airline_reviews = scrape_airline_reviews(airline, url, pages=10)
        all_reviews.extend(airline_reviews)

    # 🟩 Convert to DataFrame
    df_reviews = pd.DataFrame(all_reviews)
    print("🟩 Reviews scraped. Performing sentiment analysis...")

    if df_reviews.empty:
        print("+ No reviews scraped. Exiting...")
```

```
exit()

# Perform Sentiment Analysis + Add Rating
df_reviews = analyze_sentiment(df_reviews)
print(df_reviews.head())

# Insert into RDS only if not duplicate
insert_to_rds(df_reviews)
print(" Process completed. Data inserted without duplicates!")
```

Bibliography

[1] Jamin Rahman Jim, Apon Riaz Talukder, Partha Malakar, Mohsin Kabir, Kamruddin Nur, and M.F. Mridha, "Recent advancements and challenges of NLP-based sentiment analysis: A state-of-the-art review," *Natural Language Processing Journal*, vol. 6, 2022.

[2] X. Li, J. Zhang, Y. Du, J. Zhu, Y. Fan, and X. Chen, "A novel deep learning-based sentiment analysis method enhanced with emojis in microblog social networks," *Enterprise Information Systems*, vol. 17, no. 5, pp. 2037160, 2023.

[3] A. Quazi and M. K. Srivastava, "Twitter sentiment analysis using machine learning," in *VLSI Microwave and Wireless Technologies*, Singapore: Springer Nature Singapore, pp. 379-389, 2023.

[4] Emily Ho, Michelle Schneider, Sanjay Somanath, Yinan Yu, and Liane Thuvander, "Sentiment and semantic analysis: Urban quality inference using machine learning algorithms," *ResearchGate*, 2024.

[5] J. Ferdoshi and J. Ferdoshi, "Dataset for Twitter sentiment analysis using RoBERTa and VADER," *Mendeley Data*, 2023.

[6] F. Gulu-zada, F. Fataliyev, G. Huseynova, and S. Rustamov, "A hybrid approach to Azerbaijani lemmatization: Integrating rule-based methods with ML PoS tagging," *2024 IEEE 18th International Conference on Application of Information and Communication Technologies (AICT)*, Turin, Italy, 2024.

[7] Bandar F. Alkhalil, Yu Zhuang, Khalid T. Mursi, and Ahmad O. Aseeri, "Enhancing trust factor identification in e-commerce: The role of text segmentation and factor extraction with transformer models," *2025 IEEE 4th International Conference on AI in Cybersecurity (ICAIC)*, pp. 1-8, 2025.

[8] C. Bhatt, Gaitri, D. Kumar, R. Chauhan, A. Vishvakarma, and T. Singh, "Web scraping: Huge data collection from the web," *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, Ghaziabad, India, 2023, pp. 375-378.

[9] A. F. Aghayev and S. A. Molodyakov, "Lemmatization of nouns in the Azerbaijani language using embeddings," *2024 XXVII International Conference on Soft Computing and Measurements (SCM)*, Saint Petersburg, Russian Federation, 2024, pp. 220-223.

[10] Ananto Tri Sasongko, Muhamad Ekhsan, Wahyu Hadikristanto, and Agung Nugroho, "A comprehensive evaluation of machine learning models for sentiment analysis in employee reviews," *2024 International Conference on Electrical Engineering and Computer Science (ICECOS)*, pp. 424-429, 2024.

[11] Vani K. and D. Gupta, “Investigating the impact of combined similarity metrics and POS tagging in extrinsic text plagiarism detection system,” *2015 International Conference on Advances in Computing, Communications and Informatics (ICACCI)*, Kochi, India, 2015.

[12] P. Thongnim, V. Yuvanatemiya, P. Srinil, and T. Phukseng, “An exploration of emission data visualization in Southeast Asian countries,” *2023 10th International Conference on Information Technology, Computer, and Electrical Engineering (ICITACEE)*, Semarang, Indonesia, 2023, pp. 160-165.

[13] V. Deshmukh, J. Prithviraj, A. Rautkar, R. Agrawal, C. Dhule, and N. Chavhan, “Interactive data visualization platform to present effective teacher performance with Power BI,” *2023 International Conference on Advances in Computation, Communication and Information Technology (ICAICCIT)*, Faridabad, India, 2023.

[14] K. R. Singh, A. Arora, D. Goud, K. S. B. Ambika, S. V., and P. More, “The future of smart factories: Real-time automation and big data analytics on the cloud,” *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, 2024, pp. 1-5.

[15] K. S. Dogra, N. Nirwan, and R. Chauhan, “Unlocking the market insight potential of data extraction using Python-based web scraping on Flipkart,” *2023 International Conference on Sustainable Emerging Innovations in Engineering and Technology (ICSEIET)*, Ghaziabad, India, 2023, pp. 453-457.

[16] M. A. Gorodnichev, E. D. Nalepova, E. A. Merkulova, P. D. Rudych, and A. N. Savostyanov, “Automation of EEG data processing with HPC community cloud,” *2023 IEEE 24th International Conference of Young Professionals in Electron Devices and Materials (EDM)*, Novosibirsk, Russian Federation, 2023, pp. 1320-1323.

[17] V. Chauhan, A. Bansal, and A. Goel, “Twitter sentiment analysis using VADER,” *International Journal of Advance Research Ideas and Innovations in Technology*, vol. 4, pp. 485-489, 2018.

[18] S. Jawale, “Twitter sentiment analysis,” *International Journal of Scientific Research in Engineering and Management*, vol. 07, no. 04, 2023.

[19] S. F. Chaerul Haviana, S. Mulyono, and Badie’Ah, “The effects of stopwords, stemming, and lemmatization on pre-trained language models for text classification: A technical study,” *2023 10th International Conference on Electrical Engineering, Computer Science and Informatics (EECSI)*, Palembang, Indonesia, 2023.

[20]S. Sanap and K. Syed, “Real-time data visualization in PWAs using Power BI Embedded,” *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, 2024.

[21]S. Kitey and D. M. Kotambkar, “Visual intelligence in supply chain management with Power BI,” *2024 15th International Conference on Computing Communication and Networking Technologies (ICCCNT)*, Kamand, India, 2024.

[22]Gasparetto, Matteo Marcuzzo, A. Zangari, and A. Albarelli, “A survey on text classification algorithms: From text to predictions,” *ResearchGate*, 2024.

Biodata

Name: Vikash Vijayan A
Mobile No.: +918903673004
E-mail: vikashvijayan.a2021@vitstudent.ac.in
Permanent Address: No.4 Kumaran St kamaraj nagar, Puducherry-605011

Name: Ragul D
Mobile No.: +916369522562
E-mail: ragul.d2021@vitstudent.ac.in
Permanent Address: No 76, Audco avenue, Kanchipuram 631502

Name: Niranjana D
Mobile No.: +917550245117
E-mail: niranjana.d2021@vitstudent.ac.in
Permanent Address: No. 73A, Pammal main road, krishna nagar, Chennai 600075