Customer Sentiment Analysis and Data Reporting Using NLP

Dr. Latha P, Vikash Vijayan A, Ragul D and Niranjana D

Department of ECE, Vellore Institute of Technology, Chennai, India

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.

Keywords: Customer Sentiment Analysis, Natural Language Processing (NLP), Machine Learning, Opinion Mining, Text Analytics

I. INTRODUCTION

The analysis of customer sentiment proves crucial for business understanding consumer emotions which strengthens both customer experience and strategic business decisions. A company gains a powerful competitive advantage through analysing customer opinions because it helps spot market trends and solve essential problems and provides superior service alignment with consumer needs. The integration of natural language processing (NLP) [1] technology with machine learning algorithms [2] delivers robust sentiment analysis solutions through automated extraction of crucial information from extensive unstructured data resources. Modern technologies enable businesses to gain a more complete picture of how customers perceive products and markets so they can generate effective responses to industry changes.

The implementation of contemporary computational methods turns unprocessed customer feedback into sensible structured business insights. The accurate identification of customer emotions depends on three essential analysis techniques that include sentiment polarity analysis together with aspect-based sentiment analysis and opinion mining. Businesses can make decision-driven operational improvements through dynamic sentiment trend monitoring which becomes possible because of real-time data visualization and reporting tools. Nowadays NLP-based sentiment analysis helps to increase customer happiness through effective customer outreach activities while providing scalable solutions which help companies improve their services and expand their market presence. Sentiment analysis emerges as an essential tool for organizations because it enables the development of customer-focused strategies along with superior brand reputation which leads to industry leadership.

Through sentiment analysis businesses gain more than basic sentiment categorization when they can extract valuable understanding about customer requirements together with their desired attributes and factors causing displeasure. Organization-level sentiment analysis enables businesses to locate distinct points which need enhancement either through product quality or customer service or price strategy or branding perception. The accuracy of sentiment analysis rises when artificial intelligence (AI) integrates with the system because this technology allows the system to recognize sarcastic speech as well as subtle sentiment expressions. Predictive analytics enables businesses to forecast customer reactions while taking ahead-of-time steps that improve user experience before conflicts intensify.

II METHODOLOGY

A. MODULES

Web Scraper for Automated Data Collection

The web scraper facilitates the automatic extraction of reviews from the Skytrax website as part of its responsibility to ensure effective data collection with organized formats. The application accesses web pages using secure HTTPS requests offered by the requests library, then uses BeautifulSoup for pattern matching to obtain critical information, such as review content, date information, and geographical origin of the user. This systematic strategy enables easy progression through various review sections without disruption, continuous access to new data with constant reliability. The data gathered goes through a pipeline of structured processing, as shown in Fig. 1. A model for sentiment analysis is employed to classify customer opinions once the reviews are gathered through the web scraping module. Following the analysis, the data is stored in a cloud data warehouse for ease of access and scalability. Finally, a data visualization module assists in trend analysis and decision-making by reporting insights in graphical form.

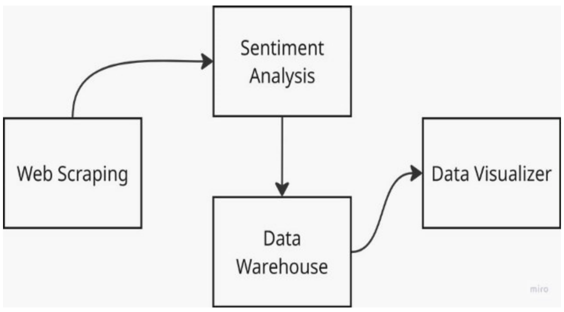

Fig.1: Proposed Model (To Run Locally)

Natural Language Processing (NLP)

The NLP takes review text from customer feedback and applies structural transformations that enable sentiment analysis of the data. This execute critical preprocessing tasks by cleaning text through Regex regular expressions while splitting it into tokens then applying POS tagging for grammatical identification roles. The removal of stop words identifies unimportant terms which do not affect the analysis results and lemmatization converts words back to base forms for normalization purposes. The data preprocessing workflow enables better sentiment classification results with more accuracy.

Sentiment scoring

The sentiment analysis runs on preprocessed data conducted using pre-trained NLP models such as VADER [3] to generate positive and negative and neutral sentiment classifications Directional sentiment scores result from the classifier process thereby allowing airline service providers to understand their customer satisfaction in detail. The established method for sentiment scoring enables organizations to monitor market directions as well as brand expectations while identifying specific service areas that need attention. Weaknesses in customer experience and emerging concerns become more manageable through sentiment analysis which allows airlines to improve their services through immediate customer feedback.

Database Management for Structured Data Storage

The system keeps processed reviews in a systematic format that makes both retrieval and analysis operations efficient and quick. The database management module arranges organized data through CSV files which divide information by airline name and review date and user location. The structured storage system allows easy combination with analytical tools which enables both researchers and airline service providers to extract useful insights from customer feedback. This module creates a structured dataset that enables the airline industry to perform trends identification and sentiment visualization and makes data-based decisions.

B. FUNCTIONALITIES

Automated Review Data Scraping

Through system automation the system extracts airline reviews from Skytrax website in a streamlined manner which leads to efficient data collection. Multiple pages receive automatic review data extraction from the web scraper which collects the review text together with submission date and user country information. The automated process lowers manual workloads while creating uniform data and enables big-scale data collection to secure broad customer feedback quantities for examination purposes.

Text Cleaning and Standardization

Standardization of extracted text becomes possible through regular expressions (Regex) which the system utilizes for cleaning operations. The cleaning process eliminates exceptional characters alongside special symbols along with extra white space to create a uniform input format. Standardization methods in text enhance sentiment analysis processing by improving data accuracy which makes review analysis more effective.

Sentiment Scoring for Deeper Insights

Sentiment analysis techniques [4] run after data preprocessing to generate positive and negative and neutral classification results from customer reviews. Airline service providers can reach better customer satisfaction insights through sentiment scoring since it lets them detect market trends alongside evaluating brand image and locating service improvement areas. Customer feedback analysis becomes effective because of the structured scoring method.

Real-Time Updates and Scheduled Execution

A scheduled execution system with periodic cleaning functions ensures that data remains relevant in the system. The system ensures continuous tracking of customer sentiment through regular updates in the dataset since it includes newly published reviews. Up-to-date reporting enables both business organizations and analysts to observe shifting passenger feedback in real time.

C. Protocols

HTTPS Protocol for Secure Data Transfer

The system maintains web scraping security through its implementation of the HTTPS (Hypertext Transfer Protocol Secure) protocol. By using HTTPS the web scraper and Skytrax website establish encrypted transmission that both protects gathered data from unauthorized access and saves information integrity. The use of HTTPS enables ethical data collection practice while simultaneously ensuring secure data transmission during the web scraping process.

Regex Protocol for Text Standardization

Text preprocessing depends on Regular expressions (Regex) to function as the fundamental protocol for creating organized and clean datasets. A text normalization protocol operates in this step to eliminate unnecessary characters alongside correcting formatting errors as well as removing unsolicited symbols from extracted review contents. Through its standardization of text format the Regex protocol makes NLP operations and sentiment analysis more accurate which produces enhanced insights from customer feedback.

III. DATA COLLECTION APPROACHES

A. Scraping Reviews

A Python script that uses requests and BeautifulSoup libraries executes the extraction task of airline reviews from the Skytrax website. The scraper successfully obtains important review details through its ability to find specific HTML items like review texts in <h3> tags and A specific datetime is defined with the <time> tag. This element's datetime attribute is used to convert the time into a format that is machine-readable and user location data under text_sub_header class elements. The loop-based system of the scraper enables an automatic exploration of numerous pages to properly collect many reviews from diverse airlines. The method leads to the collection of data that both carries structure for processing purposes and ensures reliability for analytics.

B. Data Storage and Management

Structured reviews are stored inside Amazon RDS (Relational Database Service) by using MySQL as the selected database management system. The selection of MySQL for this data storage system occurs because it offers structured dataset [5] processing and optimized performance for indexing and rapid query operations thus making it suitable for text data volumes. The database schema arranges information by using Review_ID, Airline Name, Review Date, User Location, Sentiment Score attributes for precise data retrieval. The preprocessing of received reviews takes place on EC2 (Elastic Compute Cloud) instances which insert preprocessed data into the database. The reliability of RDS is strengthened by automatic backup features alongside replication and redundancy systems which enable data recovery during system failures. The system utilizes AWS CloudWatch to monitor database operations continuously thus enabling early detection of performance problems and extended uptime optimization.

C. Data Processing and Preprocessing

Sentiment analysis only begins after data preprocessing which makes the data more consistent and accurate. The text data goes through Regular Expressions (Regex) to remove special characters together with redundant spaces and formatting inconsistencies. The text becomes more appropriate for sentiment classification when stopword removal and tokenization methods eliminate unimportant words from the analysis process. Normalizing uses lemmatization combined with Part-of-Speech (POS) [6] tagging so the analysis achieves better context understanding of review content. The standardized data structure created through this pipeline increases both accuracy and reliability of sentiment analysis model [7] performance.

D. Scalability and System Optimization

As the number of airline reviews increases the system was created to scale up and perform optimally. The data processing system receives only latest review publications through incremental updates which suppresses duplicate information storage and maintains efficient computational resource utilization. Amazon EC2 load balancing dynamically adjusts resources according to workload needs to keep the system running smoothly under high data collection demand. Real-time trend analysis becomes more efficient through key MySQL field indexing since it enables fast query execution on airline names timestamp combinations. The system implements caching solutions to lower database query demands which results in better system performance. The optimization techniques enhance feedback data processing speed during real-time operations.

E. Ethical Considerations in Data Collection

The system applies strict ethical standards for data collection operations. The web scraping [8] operations follow Skytrax website terms of service to obtain information that remains accessible to the public domain. The scraper contains programmed exclusion features that protect personal information as well as ethical integrity during the data collection process. The system implements HTTPS encryption to protect data security during transmission thus protecting the collected information from unauthorized access and maintaining its integrity. The analyzed review database serves solely for sentiment analysis research without any unauthorized use of personal information to preserve ethical transparency in data operations. The system upholds these principles which both protects legal and ethical standards and generates important findings from airline customer feedback.

IV. DATA ANALYSIS APPROACHES

A. Sentiment Analysis

The analysis of customer sentiment emerges as a critical segment within big data strategy which helps customers retrieve three possible labels to describe their sentiments as positive or negative or neutral. The built algorithm processes lemmatized [9] review data initially to maintain standard classification outputs. A rule-based or machine learning [10] NLP approach provides the system with the capability to apply sentiment scores that enable measuring customer satisfaction in reviews. The classification system generates crucial information about service quality for airlines which enables improved customer perception monitoring throughout time.

B. Keyword Extraction

The tokenized data goes through keyword extraction to find deeper meanings. The evaluation process identifies major issues by analysing commonly used terms which appear throughout customer feedback. The analysis of key terms reveals crucial passenger worries about comfort in their seats as well as the standard of customer service and the way airlines handle passenger belongings and the available in-flight entertainment options. Airlines can determine their most important service areas for customer experience by studying these selected keywords.

C. POS Tagging Insights

The Part-of-Speech [11] tagging process analyses the grammatical structure of reviews through its focus on adjectives together with nouns and verbs. With this method analysts can detect the linguistic descriptors customers use to describe airline services such as comfortable seats and delayed flights and friendly staff. Linguistic pattern classification followed by analysis produces an enhanced customer focus understanding which enables systematic improvements of individual service elements.

D. Frequency Analysis

Frequency analysis reveals the most typical phrases in airliner customer reviews through a statistical perspective of customer feedback. The system counts word appearance in various reviews to display leading themes and customer pain points for each airline. Businesses utilize this analysis to select crucial problem areas according to customer opinion patterns. It helps airlines identify recurring issues to solve them before the problems are experienced again.

E. Data Modelling:

Modeling was done to organize and structure the imported data so that analysis would be easier. This process involves connecting different tables and merging them into a single schema. The fact table incorporates key attributes such as Airline, Airport, Country, Customer Review, Site, and Date to provide a comprehensive dataset. Power BI enhances analysis capabilities and facilitates seamless querying by establishing these connections, enabling aggregated insights across multiple airlines while maintaining data integrity.

V. SYSTEM MODEL:

A. Input Layer

Through Amazon EC2 the system automates data extraction by executing Python code that utilizes Requests and BeautifulSoup to collect necessary information from the Skytrax webpages. The Skytrax website provides airline review details through its web scraping script which selects vital information that includes review text using <h3> tags together with a specific datetime is defined with the <time> tag. This element's datetime attribute is used to convert the time into a format that is machine-readable and user location information within text_sub_header class elements. The data collection runs automatically by scheduling the scraping operation to happen during predetermined times so human involvement becomes unnecessary. The system reliability of EC2 instances and execution monitoring occurs through Amazon CloudWatch tracking which performs instance performance checks while logging errors that produce failure alerts to maintain continuous data extraction.

B. Processing Layer

The processed review data needs preprocessing so that on Amazon EC2 automates the handling of extensive airline review datasets before analysis. The data cleaning procedure eliminates duplicate records along with extraneous symbols and inconsistent field format to maintain data accuracy. The analysis- ready format emerges from the text standardization process which applies Regex- based normalization. Positive and negative statuses are assigned to customer reviews through sentiment classification after defined sentiment analysis models detect sentiment types. Amazon CloudWatch enables operational tracking of real-time monitoring alongside execution time measurement and error detection and resource utilization monitoring for maintaining data processing pipeline reliability and scalability.

C. Storage Layer

The processed data finds storage in Amazon RDS (Relational Database Service) using MySQL due to its strong structured query functionality and its scalable operations and efficient management of extensive text-based datasets. Key data fields in the database schema include review ID together with user location, review text, timestamp, sentiment scores, rating and sentiment label so analysts can execute efficient queries and gain better indexing capabilities. CSV files maintain a complementary role as storage medium between different data applications for analysis through visualization [12] tools. Through this structured storage option it becomes possible to track historical trends and gain effective data retrieval capabilities for analytical purposes.

D. Output Layer

Power BI [13] to analyse the structured dataset which produces interactive dashboards to show sentiment patterns alongside keyword findings and airline services reviews. The collected information provides instant customer satisfaction data which enables businesses to discover maintenance requirements and observe customer emotional changes during time intervals. The automated data processing system provides automatic updates to preserve accuracy and operational efficiency because of minimal human involvement. The real-time monitoring of a data processing workflow through Amazon CloudWatch helps track execution times and detect errors and manage resource utilization to provide reliable system performance.

VI. CLOUD AUTOMATION

A. Automated Data Collection and Processing

Cloud automation [14] executes web scraping [15] along with data processing automatically through a streamlined process which provides both enhanced efficiency and expandable capabilities. Amazon EC2 (Amazon Elastic Compute Cloud) hosts the web scraping process that runs an automated script for periodic review extraction from airline review platforms. Some crucial data points obtainable by this process include review text along with timestamps, user location data and sentiment metrics that reduce the workload for manual data extraction. Computer operations under Amazon CloudWatch receive constant tracking services that monitor system reliability while recording performance metrics and resource utilization and tracking failures.

B. Scheduled Execution for Continuous Data Updates

The sentiment analysis accuracy depends on a programmed execution system which performs automated data retrieval. The predefined timings of the scraping script guarantee the automatic retrieval of new customer reviews immediately. Casual OA management automates passenger sentiment and airline quality tracking so businesses can use real-time information to create quick responses for current market developments. Amazon CloudWatch Alarms monitor the program to find unusual behaviour including long processing times or extraction failures which initiate emergency support procedures.

C. Scalable Cloud Storage, Management and visualization

The processed structured data goes into Amazon RDS (Relational Database Service) through MySQL for storage due to its capabilities for structured queries as well as its scalability and efficient management of large datasets. The review database schema contains essential fields such as review ID as well as user location and content together with timestamp and sentiment scores to optimize querying functions and support historical analysis and indexing capabilities. Data storage in CSV format creates available capacities for analytical tools while providing simple portability of information. Google Looker studio operate as tool for visualizing and analysing the data through interactive dashboards which provide real-time analysis capabilities. The platforms offer stakeholders capabilities to track trends and analyse sentiment distribution and extract keywords and compare airlines which supports assessments of service quality and improvement opportunities. Relational database storage together with advanced visualization tools maintains data structure and accessibility and makes it available for research-oriented and business application insights.

D. Enhancing Research and Scalability Through Cloud Automation

Cloud automation [16] ensures efficient processing of collected data as well as enables extensive sentiment analysis on a large scale. The adoption of cloud-based infrastructure enables the system to increase its capacity automatically as new review collections grow.

The system will operate sustainably over time by performing real-time trend analysis using benchmarks that measure diverse airlines. Fig.2 shows the deployment pipeline architecture. The deployment of cloud automation in research enables more profound understanding of passenger satisfaction which results in better service quality and superior customer experience analysis.

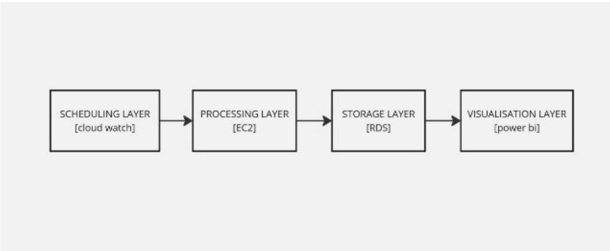

Fig.2: Deployment pipeline for Cloud Automation

E. Reliable Data Backup and Archival Using Amazon S3

Amazon S3 (Simple Storage Service) is a secondary backup layer of the cloud auto-design, providing data stability, fault tolerance, and long-term preservation. S3 buckets store regular backups of processed review datasets, such as sentiment analysis output and CSV exports. Even in case of system failures or accidental loss of data from the main storage (Amazon RDS), this backup process guarantees that previous data remains secure.

Besides robust disaster recovery capabilities, Amazon S3's highly scalable storage and extremely durable architecture enable selective data retrieval for future audits, reprocessing, or advanced analytics. Adding versioning to S3 buckets improves data security by allowing rollback to previous dataset states if necessary. Additionally, lifecycle management policies ensure that less often accessible historical data is stored at a reasonable cost by automating the data transition from frequent access tiers to archive classes like S3 Glacier. This backup plan supports both long-term research goals and real-time operations by improving the system's scalability, security, and dependability.

VII. SENTIMENT ANALYSIS

A. Sentiment Analysis for Airline Customer Reviews

To evaluate customer satisfaction sentiment analysis functions as a vital computational method which analyses written reviews. The research employs sentiment analysis to evaluate airline review sentiments through automated measurement and discover patterns in customer flight experiences. Natural language processing techniques through sentiment analysis group customer feedback into positively or negatively oriented feedback with the option of neutral. Such a method provides researchers with structured procedures to assess airline service quality while delivering insights regarding the elements affecting passenger satisfaction.

B. Implementation of Sentiment Analysis Models

This research utilizes VADER (Valence Aware Dictionary and sEntiment Reasoner) [17] as its sentiment analysis model because it was developed to analyse social media sentiment and review-based content. Due to its strong performance with informal texts VADER [18] proves effective for analysing short reviews which happen to be written in casual speech styles with numerous abbreviations and varied punctuation markings. Each review receives sentiment classification that generates positive or negative or neutral results. The preprocessing step of lemmatization [19] applies to reviews to convert word forms into their fundamental bases which improves the model's identification of word variations.

C. Aggregation and Trend Analysis

The model calculates sentiment scores for each review which allows researchers to aggregate results up to the airline level for broader sentiment trend assessment. The study shows trends in customer satisfaction by analyzing time-based averages of sentiment polarity scores which helps detect patterns found in airline services. Airlines can use sentiment distribution metrics to compare their performance with others in the industry for strategic benchmarking.

D. Validation and Accuracy Assessment

The performance validation of sentiment analysis occurs through manual comparison between automated text classifications against labels that humans have already assigned to reviews. A set of review data receives human-based sentiment labelling for establishing reference standards that help measure sentiment analysis model precision. The model performance evaluation depends on standard validation metrics. Sentiment analysis enables regulatory bodies to monitor airline service quality through data-driven evaluation based on it. It is integrated into airline customer experience research creates valuable insights that help companies make strategic decisions and conduct operational improvements.

VIII. DASHBOARD PREPARATION:

Power BI [20] enables researchers to analyse sentiment trends and review patterns alongside airline comparisons through interactive dashboard visuals. These dashboards track customer perceptions by collecting real-time data from cloud storage systems which helps researchers and stakeholders evaluate airline performance properly. Time-Series analysis in the dashboard reveals the changes in customer satisfaction levels across different airlines where sentiment trends appear. A word cloud along with trend graphs reveal the review patterns and their recurring themes which display passenger concerns and preferences. A complete airline performance analysis is possible through strategic monitoring of essential metrics between overall sentiment distribution and top positive and negative topics and geographical sentiment variations. An airline service comparison permits organizations to define their area of excellence as well as areas that need improvement. The sentiment distribution map of different regions enables airlines to modify their services through localized customization. The continuous updates of Power BI [21] dashboards depend on their linked cloud databases including AWS RDS which automatically refreshes information while maintaining real-time access. These dashboards enable data-based decision-making through organized representations of customer reaction data which leads to strategic choices. Fig.3,4,5 shows the different comparison graphs of the dashboard. Executive teams joined by policymakers and analysts use these insights to enhance their service strategies and develop passenger offerings while conducting performance reviews against airline competitors. Advanced visual analytics through sentiment analysis dashboards delivers essential functions for both airline service quality improvement and customer satisfaction enhancement. Stakeholders benefit from Power BI because they can use various devices to access these insights which promotes flexible real-time monitoring. The aviation industry relies on a flawless fusion of data visualization alongside real-time analytics and sentiment intelligence which allows airlines to both upgrade their service capability and build robust marketplace positions.

Fig.3: Dashboard outputs - gauge and bar chart, bubble chart (map)

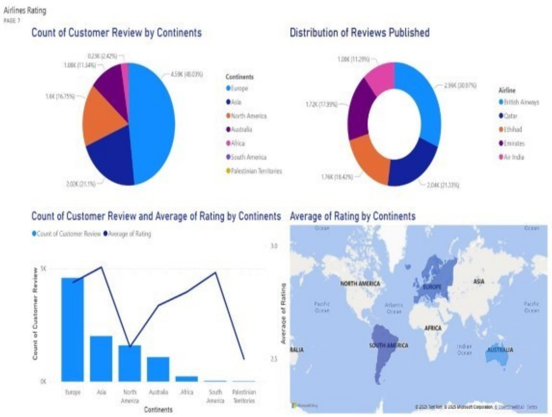

Fig.4: Pie chart, Donut chart, bar-line graph, map - review by continents

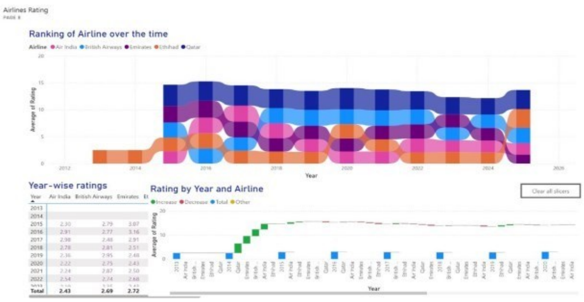

Fig.5: Ribbon chart and waterfall chart - year wise ratings

XI. RESULTS:

Insights gleaned from the analysis of customer sentiment towards the five chosen airlines, along with demographic insights regarding user countries. Fig.6, Airlines Rating comparison of all 5 airlines together.

Airlines' Overall Ratings:

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

Fig.6: Airlines Rating comparison

Qatar Airways emerges as the leader in the rating scoreboard, with the highest sentiment rating among the selected airlines, indicating predominantly positive customer sentiment. Conversely, Air India and Etihad exhibit lower sentiment ratings, suggesting a less favourable perception among customers.

Geographical Distribution of Reviews: The analysis reveals insights into the geographical distribution of reviews, with the following top countries contributing to the analysis***: United Kingdom: 3002 reviews United States: 1167 reviews Australia: 947 reviews India: 439 reviews United Arab Emirates: 324 reviews*** These insights shed light on the global reach and popularity of the airlines, as well as the geographic diversity of their customer base. Furthermore, they provide valuable input for targeted marketing strategies and service enhancements tailored to specific regions. Fig.7 shows the positive, negative and neutral reviews for all the airlines

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.7: Negative, neutral and positive ratings of all airlines.

Comparison with the existing ratings: The sentiment analysis results, which categorized reviews into positive, negative, or neutral sentiments, were converted into star ratings for further analysis. These star ratings were then compared against the ratings provided on the Skytrax website for each airline under study.

Validation Methodology: To validate the accuracy of the sentiment-based ratings, Fig.8,9 demonstrates the ratings from the website, a systematic comparison was conducted between our derived ratings and the ratings displayed on the Skytrax website. This involved extracting the official ratings from the Skytrax website for each airline. Matching these ratings with the sentiment-based ratings derived from our analysis. Assessing the degree of alignment or discrepancy between the two sets of ratings. The comparison revealed a high degree of consistency between the sentiment-based ratings derived from our analysis and the ratings available on the Skytrax website. Across the five airlines under study, the majority of sentiment-based ratings closely matched the official ratings published by Skytrax, affirming the accuracy and reliability of our sentiment analysis approach.

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

Fig.8: Average Ratings of airlines from the sentiment Analysis out of 5

Fig.9: Star Ratings from the website

Power BI serves as the solution for local dashboards because it offers strong visualizations and seamless connection to local data sources. Google Looker Studio Fig.10, serves as the preference for cloud-based automation and reporting since its automation process depends entirely on Google Cloud services. Google Looker Studio delivers reporting through streamlined updates in real time which suits projects based in the cloud. The following figure shows the dashboard created in Google Looker Studio.

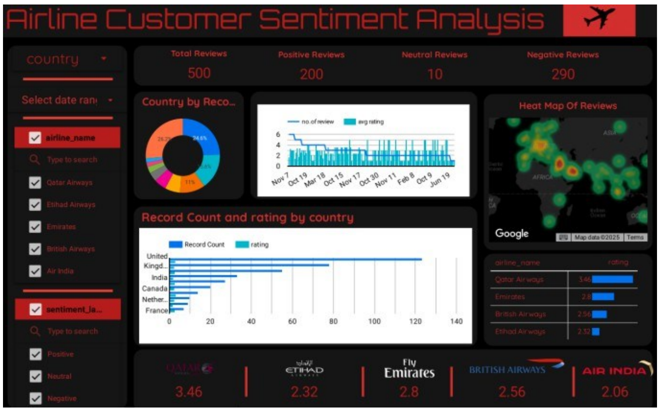

Fig.10: Google looker studio - Dashboard output

X. TESTING:

A proper test of cloud automation workflows serves to validate the smooth and reliable acquisition of airline customer reviews. Predeployment testing of the AWS Cloud-based automated web scraping system should verify its reliability and performance for scheduled review data retrieval and maintenance operations. Multiple script tests must be conducted to verify the extraction of review content and review dates and user locations along with data correctness. The system examines error-management systems to guarantee uninterrupted data-collection despite interruptions such as website structure modifications and temporary server faults and connectivity problems. The script execution is tracked through logs which help find anomalies and enhance performance with the assistance of monitoring tools. The integration process for cloud databases includes testing the accurate storage of gathered data in AWS RDS with proper database organization for future analysis. The testing progress ensures data integrity together with system scalability while promoting the longevity of the automated workflow.

Verifying the sentiment analysis results will ensure the accuracy and consistent sentiment classification of different datasets. The evaluation of the sentiment analysis model VADER utilizes a wide collection of airline reviews to establish proper classification of feelings into positive or negative or neutral categories. The precision and recall along with F1-score measurements require manual human-label dataset comparison to validate the model output. The analysis of differences between automated and manual sentiment assessments helps detect model limitations that include wrongful analysis stemming from sarcasm and ambiguous textual contexts and domain-specific language variations. The reliability of sentiment analysis to mirror real-world perceptions is verified using external sources that include customer satisfaction reports together with airline service rankings. A detailed validation method enables researchers and industry stakeholders to depend on sentiment analysis model insights that support well-informed strategic choices from passenger feedback.

XI. CONCLUSION:

Using Natural Language Processing (NLP) techniques, we successfully applied customer sentiment analysis. We were able to classify sentiments as positive, negative, or neutral by examining [22] and processing customer reviews, providing valuable information about customer satisfaction and opinions. Businesses can use the analysis to improve customer satisfaction, make well-informed decisions, and improve their goods and services. Future improvements might take the shape of employing sophisticated models for increased accuracy and real-time sentiment monitoring.

ACKNOWLEDGEMENT

The authors are grateful to the facility at Electronics and Communication Engineering department of VIT college, Chennai which provided excellent support to finish the work successfully.

REFERENCES

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, 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 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, 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, 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.