International Journal of Information Technology and Computer Science Applications (IJITCSA)
p-ISSN: 2964-3139 e-ISSN: 2985-5330
Vol. 03, No. 03, page 111 - 117
Submitted 30/9/2025; Accepted 17/12/2025; Published 18/12/2025

A Comprehensive Business Intelligence Framework for Social Media–Driven Sentiment Analysis: A Case Study of United Airlines

Marvin Imbin De Castro
College of Information and Computing Sciences, University of Santo Tomas, Manila, 1008, PHILIPPINES
e-mail : mbindecastro01@ust.edu.ph

Corresponding Autor: Marvin Imbin De Castro

Abstract
Understanding customer sentiment in real time has become increasingly critical for service-oriented industries, particularly airlines operating within highly competitive and socially sensitive environments. This study proposes an integrated Business Intelligence (BI) framework for sentiment analysis of United Airlines using social media data sourced from Twitter. The framework aims to transform large volumes of unstructured, high-velocity text data into actionable insights that support informed decision-making, customer experience enhancement, and brand reputation management. The proposed architecture incorporates sequential analytical components including data ingestion, preprocessing, natural language processing, machine learning–based sentiment classification, and BI-driven visualization. Modern text analytics techniques such as tokenization, lemmatization, and vectorization are applied to prepare textual content for polarity detection, while supervised learning algorithms are evaluated to classify sentiment into positive, negative, and neutral categories. The study outlines the rationale for adopting a scalable, cloud-compatible architecture that supports both batch and stream processing to accommodate the dynamic nature of social media data. Key implementation challenges—such as handling noisy and ambiguous text, managing evolving linguistic patterns, overcoming API rate limitations, and ensuring data quality—are examined. The paper further discusses best practices to mitigate these challenges, including robust data-cleaning pipelines, periodic model retraining, careful feature engineering, and the incorporation of governance principles for ethical data use. The results demonstrate that integrating sentiment analytics within a BI context enables organizations such as United Airlines to monitor customer perceptions more effectively and respond proactively to emerging issues. The framework provides a practical foundation for organizations seeking to operationalize social media analytics for strategic and operational decision support.

Keywords— Sentiment Analysis, Business Intelligence, Social Media Analytics, Twitter Data, Machine Learning, United Airlines

1 Introduction

The company in focus is the American passenger carrier United Airlines, whose business goal is to provide convenient and reliable flights to its customers while keeping prices competitive. In 2017, a widely circulated video on Twitter showed misconduct by United Airlines personnel toward a passenger, resulting in significant negative traction across social media platforms [1]. In today's digital environment, social media has become an essential component of daily communication and public interaction [2]. Beyond social networking, social media platforms facilitate rapid information creation, sharing, and dissemination across diverse user communities [3]. Twitter, in particular, allows users to express opinions in short text messages known as "tweets," originally limited to 140 characters [4]. Although Twitter provides a vast amount of publicly accessible content suitable for analysis, the platform's data is inherently unstructured and presents several analytical challenges [5]. As people often adopt casual writing styles on Twitter, the resulting content tends to be noisy due to grammatical inconsistencies, informal language, and punctuation errors, making it challenging to extract meaningful value from the raw text [6]. Text mining addresses these issues by applying computational techniques to process and interpret unstructured data, with natural language processing (NLP) serving as a key component of this workflow [7]. NLP facilitates the examination of large bodies of text and enables the extraction of useful patterns and linguistic features [8]. Building upon these techniques, sentiment analysis can be performed to assess emotional expressions within tweets, allowing analysts to determine the polarity, relevance, and objectivity embedded in the textual content [4].

2 Business Issues

2.1 *Corporate Culture affecting customer satisfaction*

For United Airlines, the primary business objective is to provide reliable and convenient flights to customers while maintaining competitive pricing structures. However, focusing predominantly on operational efficiency and low-cost travel may lead the airline to overlook other key performance indicators such as customer satisfaction and service quality [9]. In a widely shared tweet, Corn (2020) described a negative flight experience in which he claimed that United Airlines staff failed to enforce passenger safety protocols, particularly COVID-19 standard operating procedures during the height of the pandemic [10]. Although United Airlines' official Twitter account responded to his post, the reply was perceived as insufficient and subsequently attracted additional negative sentiment from other users, reinforcing concerns about low employee engagement and revealing deeper issues related to organizational culture [10 - 12]. Furthermore, Zhang (2020) reported a consumer ranking of major U.S. airlines, in which United Airlines received a score of 67 out of 100, indicating moderate customer satisfaction but highlighting room for improvement across various dimensions such as comfort, check-in processes, and customer service [13].

2.2 *Strained relationship between executive and low-level employees*

Behind United Airlines' operational success lie deeper organizational issues rooted in top-down management practices. According to Matousek (2018), United Airlines CEO Oscar Munoz acknowledged the company's long history of strained relationships between senior leaders and frontline employees, noting that despite recent improvements, considerable animosity persisted within the workforce [14]. In 2018, the airline replaced its quarterly performance-based bonuses with a lottery system that offered a smaller number of employees substantially larger rewards [15]. Although this initiative was presented internally as a positive cultural shift, it was met with strong criticism from employees and the media, illustrating a disconnect between leadership intentions and employee perceptions [16]. Matousek further reported statements from employee representative Hobica, who expressed a lack of trust toward upper management following the bonus restructuring and suggested that unresolved tensions continued to damage workplace morale. She argued that when employees feel mistreated, they are more likely to extend that mistreatment to customers, thereby reinforcing the broader cultural and public-relations challenges previously observed within United Airlines [17].

3 Proposed Solutions

Discussion of the results of research and testing obtained is presented in the form of theoretical descriptions, both qualitatively and quantitatively. Experimental results should be displayed in the form of graphs or tables. For graphs, follow the format for charts and drawings.

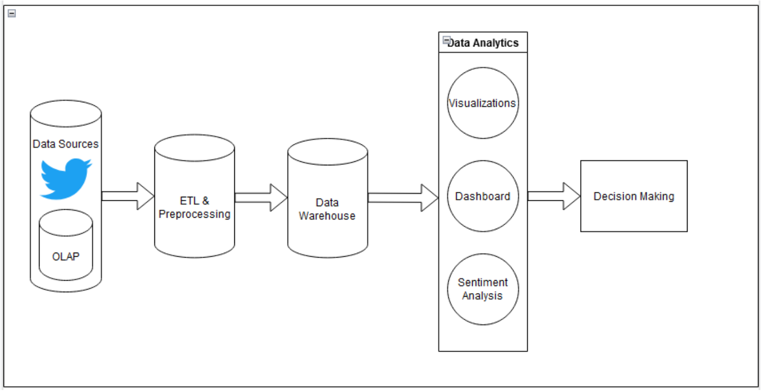

Figure 1. Proposed Technical Architecture

With the proposed technical architecture in Figure 1, United Airlines will be collecting feedback data from Twitter as well as any other OLAP database that relates to customer feedback services. The data that is collected will then be pre-processed to remove duplicated data, empty data, unstructured data, and various other factors that will affect data analytics. The data will also be transformed into the proper formatting such as proper dates or variables to be loaded into the storage. The storage method that will be used is the data warehouse that will then be used for data analytics. With data analytic tools, sentiment analysis can be created to gauge customer reactions towards United Airlines, which will be used in visualizations and dashboards. With this, United Airlines can use the information gained to form better decisions based on customer or even employee feedback.

As stated earlier, the Corporate Culture affects customer satisfaction. Using a proper Business Intelligence framework, United Airlines will be able to see how much their brand is being viewed in a negative light and how their customers view the overall service of their business. United Airlines can adjust their work culture appropriately to meet their own goals based on how satisfied the customers are, how much they are willing to change, and what kind of service to provide for their customers. The proposed solution can also cover the turbulent relationship among higher-ups and lower-level employees. By looking for tweets of employees or collecting the data from employee feedback forms instead, United Airlines will be able to gauge how their employees actually feel towards the company and the higher ups. This can provide United Airlines with the opportunity to provide more incentives to boost morale, reduce the tension between management and other employees, or regain trust among the employees that are lower down the corporate ladder. This could possibly help to alleviate issues even in their Corporate Culture, which in turn could boost customer service, as happier employees can perform their jobs better.

4 Supporting Factors

4.1 *Data Warehousing for large volumes of incoming data to be analyzed*

Data warehouses play a critical role in enabling Business Intelligence systems, as they provide centralized storage for large volumes of historical data used across various analytical processes [8]. With access to such historical datasets, analytical techniques such as text mining can be applied to extract patterns, themes, and relationships embedded within unstructured textual content [3]. In the context of United Airlines, tweets and customer feedback can be subjected to text mining to identify recurring phrases or issues associated with negative or neutral sentiment, allowing the organization to target areas requiring improvement. For example, a high frequency of terms related to “customer service” within negative tweets may indicate widespread dissatisfaction, signaling the need for operational or cultural adjustments [4].

4.2 *Stream analytics for fast paced data analytics*

Stream analytics can be integrated into a Business Intelligence framework to enable continuous data processing, allowing information to move through the system in real time for immediate analysis [8]. This capability is particularly valuable for United Airlines, as it provides rapid insight into customer satisfaction trends, allowing the organization to respond quickly to emerging issues or controversies. Such timeliness is essential because the primary data source in the proposed architecture is Twitter—a real-time microblogging platform where user-generated content evolves rapidly, often within minutes [4]. Given the high velocity and dynamic nature of Twitter data, traditional batch-processing approaches may not be sufficient, making stream analytics a more suitable solution for capturing, analyzing, and acting upon fast-changing sentiment signals [5].

4.3 *Stream Data Analytics with Dashboards and Visualizations*

Due to the rapid speed at which data may arrive from platforms such as Twitter, a Business Intelligence framework can be designed to continuously stream incoming information directly into automated visualization and dashboarding tools [5]. This approach eliminates the need for data analysts to repeatedly generate ad hoc reports from discrete data batches, replacing them with dashboards that update in near real time. Such continuously refreshed visualizations allow United Airlines to monitor emerging trends, whether positive, neutral, or controversial, and incorporate these insights into complex decision-making processes as quickly as possible [3].

5 Required Technologies and Techniques

5.1 *Oracle Cloud Infrastructure and Data Warehouse solutions*

Because of the sheer volume of tweets generated daily—particularly for a major airline—an extensive and fully integrated data storage solution is necessary to retain historical records for analysis [5]. In this context, Oracle offers suitable data warehouse architecture through its Oracle Cloud Infrastructure (OCI) Streaming service, which can fulfill United Airlines' requirements for scalable organizational data management [18]. OCI Streaming supports the ingestion of high-velocity data streams into cloud-based storage, enabling the processing of large quantities of continuous data that can subsequently be utilized for real-time analytics and decision support.

5.2 *Azure Machine Learning Libraries and Studio*

Using machine learning algorithms available through Azure Machine Learning libraries, data can be preprocessed, and sentiment analysis can be performed using the Azure Machine Learning Studio [8]. Text preprocessing involves applying several procedures to each tweet, including the removal of stop words, numbers, URLs, and email addresses, as well as expanding verb contractions, eliminating duplicate characters, and removing emojis [4]. With regard to sentiment analysis, Azure's workflow consists of multiple stages required to produce a final sentiment score. These stages typically include training a model to identify meaningful patterns within the dataset and applying algorithms such as logistic regression to develop a classification model capable of distinguishing sentiment categories. Ultimately, a score model is generated that provides predicted values along with the probability of each tweet being classified as negative, neutral, or positive [3].

5.3 *Tableau for visualizations*

Tableau is an interactive data visualization platform used to generate insights through the creation of visualizations and dashboards that support organizational decision making [1]. As discussed previously, such dashboards facilitate rapid interpretation of key analytical outputs. Furthermore, Tableau supports data streaming capabilities, enabling the production of near real-time visualizations within the application [3]. This allows United Airlines to access continuously updated dashboards at any time, providing the most recent sentiment trends and tweets visualized for immediate operational awareness.

5.4 *Lexicon-based Approach in Sentiment Analysis*

The first sentiment-analysis method proposed is a lexicon-based approach, which calculates sentiment scores from sets of sentences or entire bodies of text [3]. In this method, words appearing in the text are labeled as positive, negative, or neutral, and their polarity is determined using a valence-aware dictionary such as VADER. VADER is a sentiment model optimized for social media content, designed to capture polarity, emotional tone, and linguistic intensity within short, informal text expressions [4]. It utilizes a lexicon of lexical features that are mapped to sentiment values, with each score representing the emotional intensity associated with the word. For example, terms such as hate, despise, and detest are categorized as strongly negative. The model can also interpret contextual cues, such as negation, allowing it to classify "I do not like" correctly as a negative expression. Once individual texts are labeled, the overall sentiment can be computed by aggregating the total counts and weighted scores of positive and negative classifications, resulting in a combined sentiment score for the dataset [1 - 5]. The following are steps for the dictionary technique:

1. Clean the data by removing any duplicated tweets
2. Convert the tweets from the dataset into documents so that we can utilize our lexicon-based analyzer. Filter out the columns but leave documents data.
3. Use a valence dictionary to label the polarity of words, positive or negative. Each word from each document will be compared against the dictionary and will be assigned a sentiment tag.
4. Generate sentiment scores for each tweet and determine the number of positive and negative words in a tweet from the filtered list of tagged words.
5. Introduce an extremity score, where StSc (sentiment score) > 0 is positive sentiment, StSc < 0 is negative score, and StSc = 0 is neutral score. This will be used to predict the sentiment of the target column (document data)

6 Results and Discussion

The dataset consists of tweets containing sentiment towards 9 of the largest passenger airlines in the U.S. For this research we will be focusing on United Airlines. The goal of this analysis is to observe the polarity of sentiments by United Airline customers on Twitter towards United Airlines, then draw conclusions from the results to aid with decision-making. The software used to visualize the charts that make up the dashboard in Figure 2 is Tableau Public.

After cleaning the dataset and filtering the results related to United Airlines, the dataset contained 3,820 tweets related to United Airlines as seen in Figure 2, which made up 26% of the entire U.S. airline dataset. From this, we were able to segregate the polarity of sentiment through a pie chart where nearly half of the tweets 1,877 harbored negative sentiment, followed by 1,383 positive tweets and 560 tweets were neutral towards United Airlines.

In Figure 2, a bar chart was created to find out the reasons for negative sentiment towards United Airlines. The results were filtered to show only negative sentiments. For this context, the sentiment score refers to the strength of negative sentiment where a higher value has stronger negative sentiment and vice versa. According to the bar chart, the highest frequency tweets 336 for negative sentiment falls under customer service issues. This can be linked to the corporate culture issues discussed previously were poor customer service stems from unhappy low-level employees due to strained relationships between top-level employees. Moreover, this topic has the highest sentiment score of 65, suggesting customer service incurring the strongest negative remarks.

As for the locations for highest frequency of tweets, the word cloud in Figure 2 shows the state New York has the highest count of negative tweets 118 towards United Airlines, followed by San Francisco with 77 negative tweets, then Chicago with 76 negative tweets respectively. Coincidentally, the highest frequency of negative tweets is from states with the large populations, where New York has the largest population, followed by San Francisco and Chicago respectively. This can also be seen in the world map seen in Figure 2, where the size of the circle reflects the number of tweets made to United Airlines for each city. The size of each circle is in conjunction to the size for each word in the word cloud.

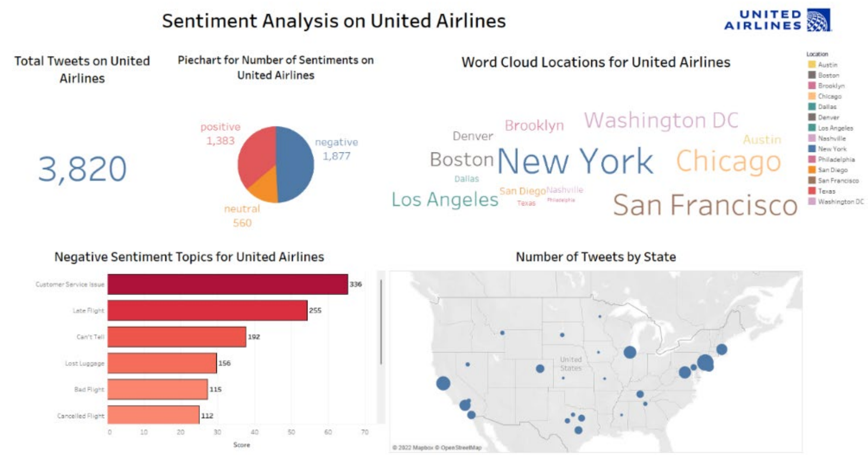

Figure 2. Dashboard for Sentiment Analysis

7 Challenges in Implementation

7.1 *Inconsistent Data Values*

The dataset being analyzed appears to have inconsistent values in some of its columns, particularly the location column. The dataset is values pulled from Twitter and the platform allows its users to put any values in their location. Because of this, the dataset analyzed for this research does not reflect total accuracy for location as the values are inconsistent. Creating a visualization such as creating a heat map for sentiment polarity in a country may seem complete; however, the inconsistent geographic values may tell a different story. Furthermore, tweets replied to certain conversations may not be related to the topic discussed, creating noise in the data.

7.2 *Topic Manipulation*

Twitter is also famous for its conversations between non-humans or bots. One of the issues from both conversations is their intention to skew a conversation in a particular direction. This is because bots tend to converse using keywords and are fixated towards a specific element of a conversation. Thus, both are creating additional noise in the data compared to human conversations.

In addition to this, users can intentionally manipulate algorithms governing trending topics for purposes unrelated to their original intent. According to Alizadeh (2021), for example, K-pop fans in January 2021 appropriated the trending political hashtag "#ImpeachBidenNow" to share images and videos of their favorite K-pop idols, effectively replacing the political discourse with unrelated fan content [5]. Such behavior presents challenges when studying hashtags associated with United Airlines, as trending topics can be easily distorted by user-driven interventions, and no reliable solutions currently exist to fully mitigate this type of manipulation within social media analytics.

7.3 *Data acquisition*

Data pulled from tweets despite using certain keywords or hashtags may not be completely related to the target topic which in this case is tweets related to United Airlines. For instance, using the keyword 'United' on its own could introduce systematic bias or biased samples of more popular topics. Instead of retrieving tweets related to United Airlines the search query could return unrelated tweets i.e. return tweets related to football club 'Manchester United'. This bias can be avoided or greatly reduced when using as many feasible search queries as possible to increase search relevance to the target topic.

7.4 *Problems of Sentiment Analysis on Tweets with inappropriate English*

Previously seen in the introduction, a 'casual approach' of tweets refers to the slang and expressions of spoken language, such as the terms 'no' and 'nah'. While both words are understood in a community from regular social interactions, not all frameworks can reach a conclusion from such casual language.

Emojis portray human expression in the form of a picture without any verbal communication. While it serves as a reflection of human emotions, its meaning can vary depending on how it is used in a sentence. For instance, the smiling emoji ':)' evokes happy emotions. However, its meaning can be changed depending on how it is used. As of now, frameworks do not have the necessary information to derive accurate sentiments from emojis. Not being able to breakdown, the meaning of emojis limits the use of association rules for sentiment analysis.

8 Best Practices

8.1 *Pre-process/Data Wrangle*

To avoid many of the challenges in implementation, especially with inappropriate english, inconsistent data values, or problems arising from topic manipulation especially with duplicate data showing the high likelihood of bot data to be mitigated but not solvable by any stretch of the means since bots are getting more advanced with each passing day. Cleaning the data is a key step to having a usable data set that does not have unnecessary data. Finding time to either stage the data manually or set up the data to automatically dispose of said data will be impartial towards storing the data in the warehouse without issues and a smooth analytics process. Preprocessing the data automatically is for the best as the high influx of tweets due to data streaming, so analysts do not have to waste their time constantly cleaning data manually and focus on analyzing the changes of attitudes based on twitter sentiment.

8.2 *Obtaining access to Twitter API and additional access*

To retrieve tweets from Twitter, developers must first register for a developer account in order to obtain API keys and authentication tokens, which allow access to approximately 18,000 tweets per 15-minute window for data extending up to the previous seven days [17]. Additional tiers of access must be requested through the developer portal due to monthly limitations—such as the cap of 500,000 tweets per month—and other restrictions imposed by Twitter's API policies. These constraints can be significantly limiting for a company like United Airlines, which may receive large volumes of tweets daily, far exceeding standard access thresholds. Consequently, developers would need to apply for elevated API permissions to support robust analytics. However, because United Airlines owns its internal OLTP systems, the company maintains full control over internal data operations, allowing analysts to access and process organizational data without such external limitations.

9 Conclusion

In conclusion, based on the business issues previously mentioned and confirmed by Figure 2, United Airlines is plagued by service issues that could be the fault of corporate culture and bad coordination with upper management and the lower-level employees, which is known due to the highest negative sentiments which all relate to said customer service issues. With the proposed business intelligence framework, these business issues can be mitigated with proper understanding of why their customers are not satisfied and in which region. Overall, it is up to United Airlines to make the call on how to restructure the company or make changes based on the data.

BIBLIOGRAPHY

[1]. A. Ohlheiser, “The full timeline of how social media turned United into the biggest story in the country,” The Washington Post, Apr. 11, 2017. [Online]. Available: https://www.washingtonpost.com/news/the-intersect/wp/2017/04/11/the-full-timeline-of-how-social-media-turned-united-into-the-biggest-story-in-the-country/

[2]. A. Sharma and U. Ghose, “Sentimental analysis of Twitter data with respect to general elections in India,” Procedia Computer Science, vol. 173, pp. 325–334, 2020, doi: 10.1016/j.procs.2020.06.038.

[3]. A. Sarlan, C. Nadam, and S. Basri, “Twitter sentiment analysis,” in Proc. Int. Conf. Information Technology and Multimedia (ICIMU), 2014.

[4]. V. A. Kharde and S. S. Sonawane, “Sentiment analysis of Twitter data: A survey of techniques,” International Journal of Computer Applications, vol. 139, no. 11, 2016.

[5]. K. Alizade, “Limitations of Twitter Data,” Towards Data Science, 2021. [Online]. Available: https://towardsdatascience.com/limitations-of-twitter-data-94954850cacf

[6]. O. Harfoushi, D. Hasan, and R. Obiedat, “Sentiment analysis algorithms through Azure Machine Learning: Analysis and comparison,” Modern Applied Science, vol. 12, no. 7, p. 49, 2018, doi: 10.5539/mas.v12n7p49.

[7]. D. Corn, “Hey @ United, a family member just flew across the country on your airline,” Twitter, Dec. 9, 2020. [Online]. Available: https://twitter.com/DavidCornDC/status/1336361619107049472

[8]. B. Zhang, “The 11 best and worst airlines in America,” Business Insider, 2018. [Online]. Available: https://www.businessinsider.com/best-worst-airlines-america-consumer-reports-2018-3

[9]. M. Matousek, “United Airlines CEO Oscar Munoz admits strained employee relations,” Business Insider, Mar. 2018. [Online]. Available: https://www.businessinsider.com

[10]. M. Matousek, “United Airlines replaces employee bonuses with lottery system,” Business Insider, Mar. 2018. [Online]. Available: https://www.businessinsider.com

[11]. M. Matousek, “United’s new bonus lottery sparks backlash from employees,” Business Insider, 2018. [Online]. Available: https://www.businessinsider.com

[12]. M. Matousek, “Employees express distrust after United bonus controversy,” Business Insider, 2018. [Online]. Available: https://www.businessinsider.com

[13]. Oracle Corporation, “Oracle Cloud Infrastructure Streaming—Overview,” Oracle Cloud Documentation, 2023. [Online]. Available: https://docs.oracle.com/en/cloud/

[14]. Tableau Software, “Tableau Platform Overview,” Tableau Documentation, 2023. [Online]. Available: https://www.tableau.com/

[15]. Tableau Software, “Tableau Streaming and Real-Time Analytics,” Tableau Help, 2023. [Online]. Available: https://help.tableau.com/

[16]. C. J. Hutto and E. Gilbert, “VADER: A Parsimonious Rule-based Model for Sentiment Analysis of Social Media Text,” in Proc. Eighth Int. AAAI Conf. Weblogs and Social Media (ICWSM), 2014. [Online]. Available: https://ojs.aaai.org/index.php/ICWSM/article/view/14550

[17]. P. S. Beri, “Sentiment Analysis using VADER,” Medium, 2020. [Online]. Available: https://medium.com/

[18]. M. Beck, “How to scrape tweets from Twitter,” Medium, 2021. [Online]. Available: https://towardsdatascience.com/how-to-scrape-tweets-from-twitter-59287e20f0f1