

## Comparative Analysis of Naive Bayes and GPT-2 for Twitter-Based Sentiment Analysis in Bank Reputation Management

Mhd Rezki Aprilan, Sfenrianto

Information Systems Management Department, BINUS Graduate Program – Master of Information Systems Management, Bina Nusantara University, Jakarta, 11480, Indonesia

*mhd.aprilan@binus.ac.id, sfenrianto@binus.edu*

**Abstract.** This study employs sentiment analysis to assess bank reputation during system failures, using Twitter data from 959 tweets collected between 2023 and 2024. We compare the performance of Naive Bayes and GPT-2 algorithms in analyzing customer complaints about banking application issues. The research focuses on three variables: problem features, initial customer sentiment, and sentiment after bank response. Naive Bayes achieved accuracies of 76.75%, 62.23%, and 87.78% for these variables respectively, while GPT-2 recorded 82.02%, 67.33%, and 61.39%. The study reveals that GPT-2 outperforms Naive Bayes in initial sentiment classification, likely due to its superior ability to understand context. These findings provide valuable insights for banks to improve their reputation management strategies and customer service responses during system failures.

**Keywords:** sentiment analysis, bank reputation, system failures, twitter data, customer complaints, naive bayes, gpt-2, reputation management, machine learning, customer service.

## 1. Introduction

In the digital age, the banking sector has seen a significant shift towards digital services, notably through the adoption of banking applications. These applications play a crucial role in facilitating communication and transactions between banks and their customers, aiming to deliver efficient and satisfactory services. However, operational challenges such as system disruptions and data breaches can precipitate communication crises, which disrupt services and adversely affect the bank's reputation.

This research explores the management of crisis communication to mitigate the negative impacts on the reputations of banking institutions. With approximately 167 million active social media users in Indonesia, the implementation of effective crisis communication strategies becomes essential. This study employs the Naive Bayes and GPT-2 models for sentiment analysis to identify optimal communication strategies, leveraging their effectiveness in text classification.

Currently, many banks still rely on manual sentiment management, which often results in suboptimal outcomes. There is a pressing need for automated solutions to identify and respond to negative sentiments more effectively. This study utilizes Python and Selenium for data scraping and sentiment analysis, concentrating on pivotal issues such as login difficulties. Through this approach, banks can enhance customer satisfaction and protect their reputation by proactively managing issues.

**Research Gap and Objectives:** A major challenge in the era of digital banking is the occurrence of communication crises, often triggered by operational disruptions like system failures and security breaches. These crises can interrupt services and potentially cause long-term damage to a bank's reputation.

The objectives of this study are to:

- Analysed bank sentiment on social media by comparing the Naive Bayes and GPT-2 models.
- Develop effective communication strategies to boost customer satisfaction and enhance the bank's reputation.

**Rationale for Model Selection:** The strategic selection of the Naive Bayes and GPT-2 models is due to Naive Bayes' efficiency in conducting preliminary sentiment analysis across large datasets, making it ideal for initial data filtering. In contrast, GPT-2 is chosen for its advanced language modeling capabilities, which enable it to handle the nuanced and context-rich language typical of social media, providing deeper insights into customer sentiments during crises.

**Paper Structure:** This paper is organized into the following sections: Introduction, Literature Review, Methodology, Results, and Conclusions. The Introduction discusses the digital transformation in banking and the associated challenges. The Literature Review examines prior research on digital transformation, sentiment analysis, and crisis communication within the banking sector. The Methodology section describes the analytical techniques employed to analyze social media data. The Results section compares the performance of the machine learning models, and the Conclusions section offers strategic insights for practical implementation.

**Contributions:** This study contributes both theoretically and practically to the fields of bank reputation management and sentiment analysis. Theoretically, it advances the application of sentiment analysis techniques within the banking sector, especially in understanding how digital crises influence customer perceptions and bank reputations. Practically, it showcases how sophisticated sentiment analysis models like Naive Bayes and GPT-2 can refine crisis communication strategies, enabling banks to respond more effectively to customer sentiments and thus improve reputation management during system failures. This research not only provides a robust methodological framework for analyzing customer sentiment but also furnishes actionable insights that can guide banks in enhancing their crisis management strategies.

## 2. Literature Review

### 2.1. Sentiment Analysis in the Banking Sector

Sentiment analysis is a methodological process aimed at identifying and extracting subjective information from a given data source. This technique is frequently applied to texts from social media, product reviews, and other platforms that capture public opinions. The primary goal is to evaluate the attitude or sentiment conveyed in the text, determining whether it is positive, negative, or neutral with respect to a specific subject or topic. In a study by Fauzi et al. (2023, p.3), sentiment analysis was utilized to assess tweets and classify sentiments, either positive or negative, in relation to Islamic banking. Through the use of Twitter data, this analysis seeks to explore and comprehend public opinion trends surrounding Islamic banking.

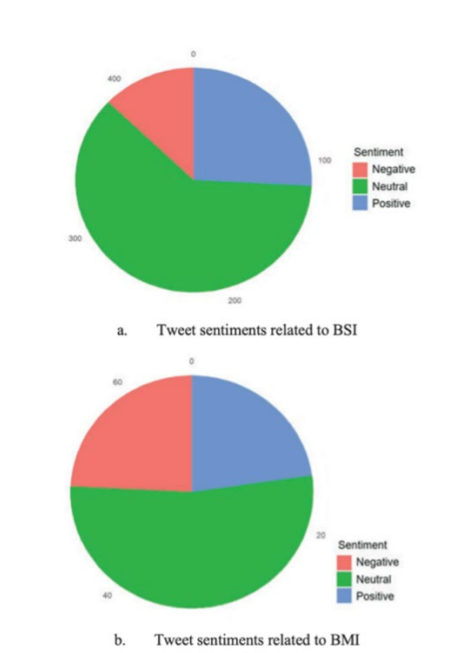


Fig. 1: Tweet Sentiment Bank

### 2.2. Twitter as Data Source

The social media platform Twitter has emerged as one of the most influential communication tools in the digital age. Its role in transforming global communication dynamics is particularly significant in the context of public communication. Public institutions have adjusted their communication strategies to align with the interactive and fast-paced environment that platforms such as Twitter offer. This adjustment is especially critical during crises, where Twitter functions as an essential tool for the rapid dissemination of information and public engagement. Furthermore, Twitter plays a pivotal role in enabling sentiment analysis, which is vital for understanding public reactions to various crisis situations. This analysis helps public institutions refine their communication strategies during such critical periods (Tasente et al., 2024).

### 2.3. Rapid Miner for Sentiment Analysis

Sentiment analysis, a method used to classify and analyze opinions expressed in text, is increasingly employed in research to gain deeper insights into public sentiment. Rapid Miner, a comprehensive data analysis tool, has demonstrated its effectiveness in this area, particularly in studies related to the Sembako/BPNT Program (Noor et al., 2022). By facilitating data extraction, cleaning, and sentiment classification, especially from social media platforms such as Twitter, Rapid Miner allows researchers to efficiently process and analyze unstructured data. This process transforms raw data into structured,

meaningful information, providing critical insights into public opinion dynamics. The integration of Rapid Miner with Gataframework in this research illustrates the potential of modern data analysis tools to enhance the understanding of public perceptions and reactions to government initiatives (Noor et al., 2022). The findings from this sentiment analysis provide valuable guidance for policymakers and stakeholders, emphasizing the importance of such tools in social and communication studies.

#### **2.4. CRISP-DM Methodology**

The CRISP-DM (Cross-Industry Standard Process for Data Mining) methodology is widely employed in text mining research and is often integrated with text pre-processing techniques. This methodology is structured into several key phases: Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. Each phase is crucial for ensuring that the data mining process is comprehensive, systematic, and aligned with the project's business objectives. In a study by Kurniawan et al. (2020), CRISP-DM was applied to data processing, utilizing text pre-processing through the GATA Framework and RapidMiner to manage Indonesian language data. This framework proves particularly valuable for managing large and complex datasets, especially in the context of text mining and natural language processing, providing a solid foundation for generating relevant and accurate research results.

#### **2.5. Business Intelligence in Financial Services**

Business Intelligence (BI) involves the use of tools that provide companies with insights into intangible resources, enabling the development of models that enhance business intelligence. The primary objective of BI is to improve the efficiency and effectiveness of organizational management, allowing companies to achieve more complex goals and remain competitive in an increasingly critical and vigilant society (Wang et al., n.d.). BI applications play a key role in innovative management by processing large volumes of intangible data, automatically extracted from various digital sources. These results can be integrated into management reports and dashboards, supporting market sustainability and enabling companies to predict social behavior, thus fostering the development of proactive management models.

BI employs techniques such as OLAP and data mining to analyze and visualize complex relationships among intangible assets, including experiences, emotions, and attitudes, particularly those derived from social media. At the corporate level, BI-driven data analysis aims to improve strategic decision-making processes, thereby enhancing business performance. Historically, strategic decisions were based on precise analysis of historical data, which enabled companies to understand current problems, identify potential strategies, and predict outcomes. In the modern digital ecosystem, intangible assets have become among the most valuable resources for organizations. However, managing data associated with corporate big data presents significant challenges.

Organizations must adopt tools that advance their BI models, including the automation of information from digital sources, daily data aggregation, and brand-based historical analysis, enabling effective visualization and interpretation. In the banking sector, where financial market sustainability is critical, companies must develop integrated management strategies based on the key intangible assets of relational capital. Decision-makers must analyze and understand the vast volumes of opinions generated in the digital ecosystem, focusing on the emotions and perceptions stakeholders associate with brand performance and communication.

#### **2.6. Machine Learning for Sentiment Analysis**

In machine learning research, as described by Jain and Dandannavar (2016), two sets of data are typically required: a training set and a testing set. In supervised learning classification methods, the training set is used to train the model to recognize distinguishing attributes within the text, while the model's performance is evaluated using the testing set. Commonly used machine learning algorithms

for text classification include Maximum Entropy (ME), Naive Bayes (NB), and Support Vector Machines (SVM).

Sentiment analysis using machine learning begins with the collection of a labeled dataset of tweets. This data often contains noise and must be preprocessed using various Natural Language Processing (NLP) techniques. Relevant features for sentiment analysis are then extracted, after which the classifier is trained on the dataset and tested on new, unseen data to assess its effectiveness.

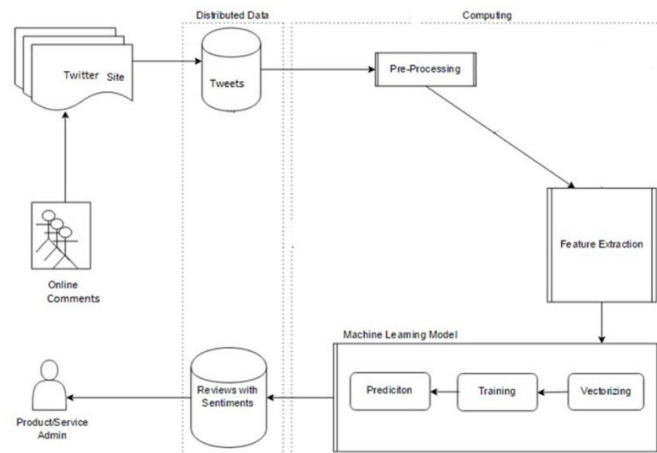


Fig. 2: Workflow for twitter sentiment analysis using machine learning algorithms by Nirranjan

## 2.7. Naive Bayes Method

According to Omatu et al. (n.d.), the Naive Bayes algorithm—named after the British mathematician Thomas Bayes—is a statistical classification method based on Bayes' Theorem. This Bayesian classification technique is widely recognized for its simplicity and effectiveness as a predictive model.

$$P(X|C_i) = \frac{P(X|C_i)P(P(C_i))}{P(X)}$$

In this context, let  $X = x_1, x_2, x_3, \dots, x_n$  represent a set of samples, and  $C_1, C_2, C_3, \dots, C_m$  represent a set of different classes. To classify a sample, the probability that the sample belongs to each class is calculated using the relevant equation. The sample is then assigned to the class with the highest probability. This approach leverages statistical principles to enable more accurate classification decisions.

## 2.8. Large Language Models in Sentiment Analysis

Large Language Models (LLMs) are advanced artificial intelligence systems designed to automatically understand and generate human language. These models are trained on vast text datasets, enabling them to learn and recognize complex linguistic patterns. The primary function of LLMs is to execute a wide range of natural language processing tasks, including text generation, comprehension, and semantic analysis.

Models such as GPT (Generative Pre-trained Transformer) and T5 (Text-to-Text Transfer Transformer) are prominent examples of LLMs that have significantly advanced how machines process and interpret text. These models utilize sophisticated natural language processing techniques to generate contextually relevant and accurate responses based on the input they receive (Al Naffakh et al., 2024).

## 2.9. Web Scrapping for Data Collection

Web scraping, as described by Perwej et al. (n.d.), refers to the automated process of collecting data and information from the web. This method involves extracting data from the internet and is employed

in various applications, including information retrieval, news aggregation, website monitoring, and more. Web scraping provides a faster and more efficient alternative to manual data collection, making it increasingly prevalent in the digital age.

Web scraping software, commonly referred to as "bots," is designed to navigate websites, identify relevant pages, and extract meaningful data. These bots automate the data extraction process, facilitating the rapid collection of large volumes of information. In the era of big data, where information is constantly updated and evolving, web scraping has become an essential tool. Its applications are broad, particularly in the field of data analytics, offering significant advantages in the gathering and analysis of information.

## **2.10. TF-IDF (Term Frequency-Inverse Document Frequency)**

TF-IDF, which stands for Term Frequency-Inverse Document Frequency, is an algorithmic method used to calculate the weight of each word within a document. This method is widely employed in text processing for tasks such as document classification, text mining, and information retrieval.

TF-IDF operates by calculating two key components:

- **Term Frequency (TF):** This measures how frequently a term appears in a document. The more often a word occurs in a document, the higher its TF value. However, if a word appears frequently across all documents, its TF value is reduced to balance its weight.
- **Inverse Document Frequency (IDF):** This measures how common or rare a word is across a collection of documents. If a word appears in many documents, its IDF value will be low. Conversely, if a word appears in only a few documents, its IDF value will be high. This adjustment reduces the weight of common words (e.g., "and," "the") and increases the weight of rarer, potentially more meaningful terms.

The final TF-IDF value for a word in a document is the product of its TF and IDF. The general formula is:

$$\mathbf{TF-IDF(t, d) = TF(t, d) \times IDF(t)}$$

Where:

- **t** represents a specific word,
- **d** is a document,
- **TF(t, d)** is the frequency of word **t** in document **d**,
- **IDF(t)** is the logarithm of the total number of documents divided by the number of documents containing word **t**.

TF-IDF is widely used in applications such as text classification and social media text analysis, where it assigns weights to words based on their importance within a specific document and the entire document collection. In the journal you uploaded, TF-IDF is employed to identify frequently occurring words on social media platforms like Twitter, which are then used in forecasting processes through time series analysis.

## **2.11. Customer Satisfaction**

Customer satisfaction is defined as the evaluation of a product or service's performance by the customer, which directly influences customer loyalty. It involves comparing the actual performance of the product or service with the customer's expectations (Sudira et al., n.d.). The hypothesis posits that this approach can effectively measure sentiment and customer satisfaction from social media comments, revealing varying levels of satisfaction across different digital payment services.

The hypothesis in this study is grounded in the direct disconfirmation approach, which is subjective and employs a summary rating scale to measure confirmation or disconfirmation. This direct measurement aligns with the weighting method, wherein customer ratings of each attribute are multiplied by its assigned importance, although the statistical independence of variables has not been confirmed. Within this framework, sentiment analysis is used to determine customer satisfaction levels, as interpreted through Instagram comments, and categorizes them as positive, negative, or neutral.

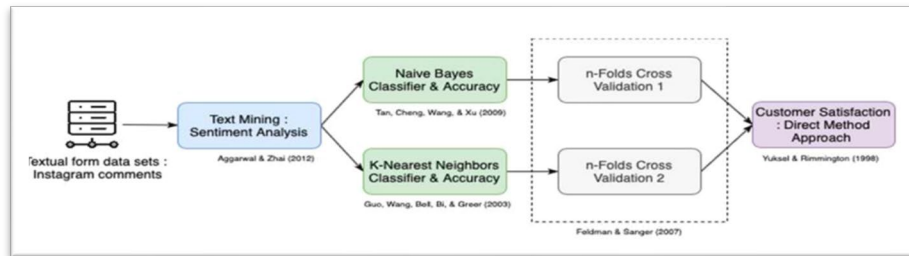


Fig. 3: Research Model dari Instagram Sentiment Analysis with Naive Bayes and KNN: Exploring Customer Satisfaction of Digital Payment Services in Indonesia

## 2.12. GPT-2 Method

Generative Pre-Training (GPT-2), developed by OpenAI, is an artificial intelligence model designed to comprehend and generate natural language text. In the study conducted by Tsai et al. (2021), GPT-2 was implemented to facilitate automatic question generation, aiming to reduce the workload of educators in creating exam or practice questions.

Methodology:

- **Keyword Extraction:** The initial step in question generation involves extracting keywords using the BERT (Bidirectional Encoder Representations from Transformers) model. This model analyzes text from textbooks or learning materials provided by educators to identify and extract relevant key elements.
- **Syntactic Analysis:** After the successful extraction of keywords, syntactic analysis is conducted using Stanford CoreNLP. This analysis identifies complete sentences that contain the keywords, which are essential as they form the foundation for question formulation.
- **Question Generation:** The analyzed and validated sentences are then input into the GPT-2 model. GPT-2 generates coherent and contextually appropriate questions based on these sentences. The model predicts words step by step until a logical and relevant question is formed, which can be used as an evaluation or practice tool.

The research also includes a testing and evaluation phase, during which the questions generated by GPT-2 are reviewed and assessed by educators. This evaluation determines the relevance and effectiveness of the questions within the actual learning context. Educators may also edit or modify the generated questions to enhance their quality and alignment with the learning material. The use of GPT-2 for automatic question generation demonstrates significant potential in reducing the time and effort educators invest in preparing evaluation materials. By integrating natural language processing and artificial intelligence technologies, the educational process can be substantially improved, contributing to more efficient and effective teaching practices.

## 2.13. Hugging Face

The development and implementation of a system known as HuggingGPT leverage Large Language Models (LLMs), such as ChatGPT, to function as a controller that coordinates various artificial intelligence (AI) models available on the Hugging Face platform to solve complex AI tasks.

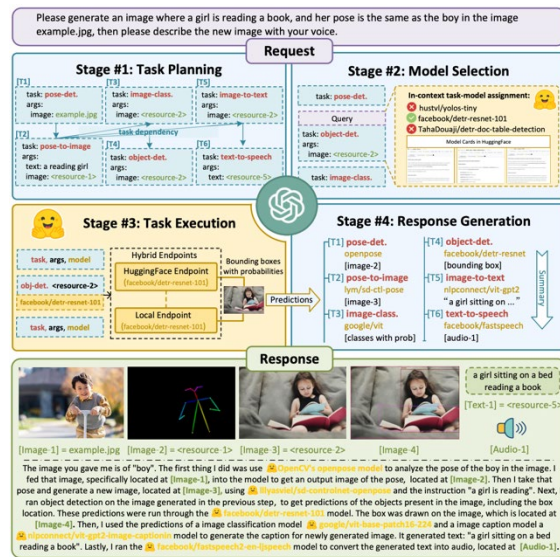


Fig. 4: Overview Hugging Face

The key steps in HuggingGPT, as outlined by Shen et al. (n.d.), are as follows:

- **Task Planning:** HuggingGPT utilizes ChatGPT to analyze user requests and develop a plan that outlines various tasks. This plan is designed to harness the capabilities of different AI models based on the specific requirements identified from the user’s input.
- **Model Selection:** Following task planning, the system selects the most appropriate AI models from Hugging Face for each task. The selection is informed by the documented descriptions and capabilities of the models available on the Hugging Face platform.
- **Task Execution:** The selected models are executed to perform their respective portions of the overall task. This step involves applying the models to process data, conduct calculations, or generate the necessary output.
- **Response Generation:** Upon task completion, HuggingGPT integrates the outputs from the models and utilizes ChatGPT to generate a coherent response for the user, summarizing the outcomes of the process.

### 2.14. Synthesis of Previous Studies

The study by Airin Fariza Abu Samah et al. developed a web-based dashboard that visualized sentiment analysis results for the top three Malaysian airlines: AirAsia, Malaysia Airlines, and Malindo Air. Using the Naïve Bayes algorithm, the study conducted bilingual sentiment analysis on data collected from Twitter. The Naïve Bayes algorithm was selected due to its proficiency in handling multiclass prediction challenges, especially in sentiment classification tasks where public opinion was divided into positive, neutral, and negative sentiments. Data were pre-processed using Natural Language Processing (NLP) methods to ensure accurate classification.

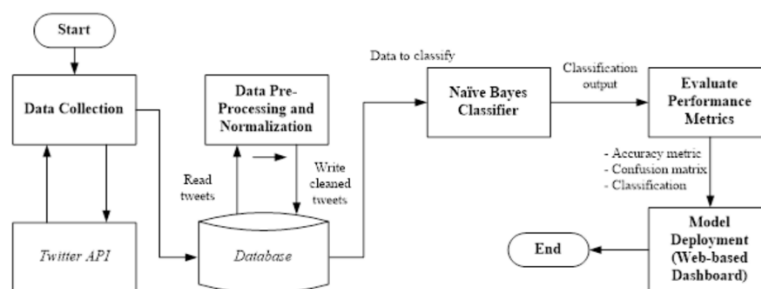


Fig. 5: Research design

The results of the sentiment analysis were displayed through interactive visualizations, including word clouds, pie charts, and bar charts. These visualizations offered significant insights into the public reputation of the airlines, revealing key trends and opinions shared on social media. This approach validated the effectiveness of the Naïve Bayes algorithm for sentiment analysis in the aviation industry, highlighting its utility in extracting actionable insights from public opinion data.

Building upon this work, Gallastegui et al. (2024) also employed machine learning techniques for sentiment analysis in the financial sector, specifically analyzing Twitter data related to BBVA, a prominent European bank. Their study introduced more advanced models, including XGBoost, to achieve higher accuracy in sentiment classification compared to traditional methods such as Support Vector Machines (SVM) and Random Forest. This research demonstrated that XGBoost consistently outperformed these older models in handling large-scale datasets, achieving accuracy rates above 90%. Furthermore, they addressed the challenge of multilingual sentiment analysis by incorporating models that could process data in multiple languages, making their approach highly relevant in globalized markets.

While the Gallastegui study focused on the European banking sector, this research aims to apply similar advanced sentiment analysis techniques in the Indonesian banking industry. By leveraging models such as XGBoost and focusing on customer sentiment in both English and Indonesian, this study seeks to address the complexities of multilingual sentiment analysis in Indonesia's diverse linguistic landscape. Additionally, this research contributes to the field by introducing real-time dashboard visualizations, providing Indonesian banks with practical tools to monitor and respond to customer feedback, ultimately enhancing customer satisfaction and reputation management.

In summary, this study builds on prior research by combining the sentiment analysis techniques from the airline industry (Abu Samah et al.) and the advanced machine learning models used in the financial sector (Gallastegui et al.). It offers a novel contribution by applying these methods in a new context—Indonesia's banking sector—while also advancing the accuracy and effectiveness of sentiment classification across multiple languages.

### **3. Methodology**

This study employs the Cross-Industry Standard Process for Data Mining (CRISP-DM) framework, recognized for efficiently organizing data mining projects. The adoption of CRISP-DM is pertinent due to its structured approach to handling complex datasets, including the unstructured data from social media platforms like Twitter. This framework facilitates an iterative process from business understanding to deployment, ensuring accurate and adaptable analyses for the dynamic nature of social media sentiments. The flexibility provided by CRISP-DM in data pre-processing and model evaluation phases is crucial for this study's comparative analysis between Naive Bayes and GPT-2 machine learning models.

The research utilized several tools:

- Integrated Development Environment (IDE) IntelliJ IDEA: For software development and scripting.
- Microsoft Excel: Employed for data manipulation and preliminary analysis.
- Rapid Miner: Used for complex data mining tasks.
- Selenium Developer Tools: For automating the web scraping process.

The analysis incorporated two primary machine learning models:

- Naive Bayes: Known for its efficiency in quick, preliminary sentiment analysis within large datasets. This model's probabilistic approach effectively categorizes baseline sentiments, ideal for initial data filtering.
- GPT-2: Chosen for its advanced language modelling capabilities, which provide a deeper understanding of complex text structures typical in social media. GPT-2 is preferable to its successor GPT-4 due to its balance between computational demands and depth of analysis.

Data collection involved scraping tweets using Python and Selenium, focusing on terms related to

bank system failures such as "bank" and "system failure." The data pre-processing steps included:

- Data Cleaning: Removing duplicates and handling missing values.
- Tokenization: Segmenting text into smaller components for analysis.
- TF-IDF Application: Utilizing Term Frequency-Inverse Document Frequency to extract significant textual features from tweets.

The study analysed three main variables to capture the customer experience during system failures:

- Problem Features: Identifying issues mentioned in the tweets.
- Initial Sentiment: Assessing customer sentiment prior to any bank intervention.
- Post-Response Sentiment: Evaluating sentiment following the bank's response.

Ethical considerations were meticulously adhered to, given the public nature of Twitter data:

- All data was publicly available, with no personal information extracted.
- Any potentially identifying details were anonymized to maintain privacy.
- The study complied with social media data usage policies, ensuring no ethical violations.

This systematic and transparent methodology not only enhances the robustness of the sentiment analysis but also aligns with ethical standards, providing banks with crucial insights into effective reputation management during crises.

## **Business Understanding**

Define In the Business Understanding stage of the CRISP-DM framework, the focus is on comprehensively understanding the business problem and setting analytical goals that will support the achievement of organizational targets.

- Business Context:
  - Background: Errors in banking applications can result in customer dissatisfaction, financial losses for the bank, and long-term negative impacts on reputation.
  - Business Problem: There is an increase in the number of negative tweets related to banking application errors, reflecting potential larger issues in customer service and IT infrastructure.
  - Business Impact: Widespread negative sentiments on social media can influence the perception of both new and existing customers, potentially lowering customer satisfaction and resulting in revenue loss.
- Business Objectives:
  - Main: Rehabilitate customer trust and improve the public image of the banking application.
  - Specific: Identify and analyze customer sentiments related to application errors discussed on Twitter. Formulate effective communication strategies to manage the crisis.
- To understand the objectives and define the problem to be solved, this study involves a series of activities with the following methods:
  - Literature Review: Conduct a literature study to learn about various aspects related to the use case and technical aspects of data classification implementation on customer sentiment.
  - Research Objective and Use Case Analysis: Examine research objectives and use cases related to the application of classification in providing recommendations for effective strategies, including communication, reputation crisis handling, information transparency, and service recovery.
  - Follow-Up and Review: Monitor and analyse sentiments on Twitter and evaluate responses during the crisis to identify effective tactics and areas needing improvement.

## **Data Understanding:**

In the Data Understanding stage of CRISP-DM, the goal is to collect and explore available data to gain a deep understanding of the information held. This involves initial assessment, data exploration, and data quality verification, for the case of system failure analysis in banks discussed on Twitter.

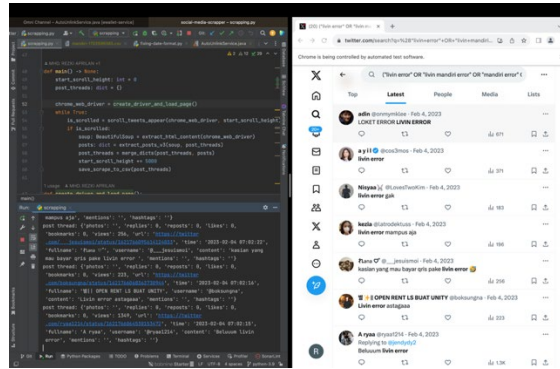


Fig. 6: Dataset Scraping Process

- **Data Collection:**
  - **Source of Data:** Data is collected from Twitter via a scraping process, which includes tweets specifically mentioning the bank application and issues faced by users.
  - **Data Volume:** A total of 959 tweets have been collected during the research period spanning from 2023 to 2024.
  - **Data Variety:** Types of data available include tweet text, date and time of posting, retweets, likes, replies, mentions, hashtags, and so on.

time	url	username	fullname	content	photos	mentions	hashtags	reposts	replies	likes	bookmarks	views
1	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	labil, ketawa terus ampe ga dngin lagi bang				0	0	0	0	5.0
2	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	gula ketawa ini dia bant sampeh				0	0	0	0	12.0
3	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	labil, WYWKWK LUCULU BANGET ASG				0	0	0	0	275.0
4	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	WYWKWKWKWK				0	0	0	0	2.0
5	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	Customer servi Hai kakak mohon info ke wa ya agar kendala nya kamu dapat kami bantu				0	0	0	0	30.0
6	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	Min, kakak mau ganti kartu pertanya y @Halobca				0	0	0	0	30.0
7	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	Customer servi Urusa tidak lanjut, boleh inform detail kendalanya ke wa ya kak, untuk				0	0	0	0	0
8	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	rd WYWKWKWKWK WYWKWKWKWK				0	0	0	0	1558.0
9	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	gagal, @Blitzspace BANGET ANING @Blitzspace				0	0	0	0	1558.0
10	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	A Man with Pri @BankBCA halo... kaku brok kantor @BankBCA				0	0	0	0	8.0
11	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	@Halobca tabrak oke DM ya min, th @Halobca				0	0	0	0	20.0
12	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	rekening,aku ada hal lain yang ingin d @Halobca				0	0	0	0	24.0
13	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	manual "no" ke B9888 malah gagal ke @Halobca				0	0	0	0	18.0
14	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	Srg ganti ke data menuah https://twitter.com/Blitzspace/status/77				0	0	0	0	66.0
15	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	HAHA, @Halobca min mau tanya arti kode F @Halobca				0	0	0	0	2370.0
16	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	LUCU BGT LAGI				0	0	0	0	303.0
17	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	tuanggap @Halobca mohon bantuan @Halobca				3.0	6.0	0	0	25.0
18	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	Ames LUCU LAGI				0	0	0	0	3271.0
19	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	Gakabababab @Halobca min mau tanya arti kode F @Halobca				6.0	12.0	0	0	44.0
20	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	monte casino pengabd gibran lu yak				0	0	0	0	3794.0
21	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
22	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
23	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
24	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
25	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
26	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
27	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
28	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
29	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0
30	2023-12-25 00:00:00	https://twitter.com/Blitzspace	Blitzspace	meann @Halobca min mau tanya, Saya sudah @Halobca				3.0	2.0	0	0	11.0

Fig. 7 Dataset

- **Data Quality:**
  - **Accuracy:** Check for errors in data collection, such as irrelevant tweets or advertisements.
  - **Completeness:** Assess if there are any gaps in the data or periods without data that could affect the analysis.
  - **Consistency:** Ensure that data across the period is collected and recorded in a uniform manner.
  - **Validity:** Verify that the collected data is valid and represents the research questions to be answered.
- **Data Preparation:**

The Data Preparation stage within CRISP-DM involves processing the collected data to ensure it is ready for analysis and modeling. Detailed steps in this stage, in the context of sentiment analysis regarding system failures in bank applications using Twitter data for the variables sentiment content, feature-problem, and sentiment-after-response, are as follows:

- Data Preparation for GPT-2

```

// Import necessary libraries
IMPORT pandas AS pd
IMPORT torch
IMPORT sklearn.metrics, sklearn.model_selection
IMPORT transformers

// Define a class for the sentiment dataset
CLASS SentimentDataset:
    INIT(dataframe, tokenizer):
        self.dataframe = dataframe
        self.tokenizer = tokenizer
    FUNCTION __len__():
        RETURN LENGTH(self.dataframe)
    FUNCTION __getitem__(idx):
        text = self.dataframe[idx]['content']
        label = self.dataframe[idx]['sentiment content']
        inputs = tokenizer.encode_plus(text, add_special_tokens, return_tensors, padding, truncation, max_length)
        input_ids = SQUEEZE(inputs['input_ids'])
        attention_mask = SQUEEZE(inputs['attention_mask'])
        label_mapping = {'neutral': 0, 'negative': 1, 'unidentified': 2}
        label = label_mapping[label]
        RETURN input_ids, attention_mask, label

// Define function to evaluate the model
FUNCTION evaluate(model, data_loader):
    model.eval()
    predictions = []
    true_labels = []
    FOR EACH input_ids, attention_mask, labels IN data_loader:
        MOVE input_ids, attention_mask TO device
        outputs = model(input_ids, attention_mask)
        logits = outputs[0]
        preds = ARGMAX(logits)
        predictions.APPEND(preds)
        true_labels.APPEND(labels)
    RETURN accuracy_score(true_labels, predictions)

// Main execution block
MAIN:
    PRINT("Persiapan data...")
    df = pd.read_csv("path_to_csv_file")
    train_df, test_df = train_test_split(df, test_size=0.2, random_state=42)
    PRINT("Pra-pemrosesan Data...")
    tokenizer = GPT2Tokenizer.from_pretrained("gpt2")
    model = GPT2ForSequenceClassification.from_pretrained("gpt2", num_labels=3)
    PRINT("Pelatihan Model...")
    train_dataset = SentimentDataset(train_df, tokenizer)
    test_dataset = SentimentDataset(test_df, tokenizer)
    train_loader = DataLoader(train_dataset, batch_size=8, shuffle=True)
    test_loader = DataLoader(test_dataset, batch_size=8, shuffle=False)
    device = SELECT_DEVICE("cuda", "cpu")
    model.to(device)
    optimizer = AdamW(model.parameters(), lr=2e-5)
    epochs = 3
    FOR epoch IN range(epochs):
        model.train()
        total_loss = 0
        FOR input_ids, attention_mask, labels IN train_loader:
            MOVE input_ids, attention_mask, labels TO device
            model.zero_grad()
            outputs = model(input_ids, attention_mask, labels)
            loss = outputs[0]
            total_loss += loss
            loss.backward()
            optimizer.step()
        avg_train_loss = total_loss / LENGTH(train_loader)
        PRINT("Epoch", epoch, "Loss:", avg_train_loss)
    PRINT("Evaluasi Model...")
    test_accuracy = evaluate(model, test_loader)
    PRINT("Test Accuracy:", test_accuracy)
    model_path = "path_to_save_model"
    SAVE model.state_dict TO model_path
    PRINT("Model telah disimpan di", model_path)

```

Fig. 8: Pseudocode GPT-2

To provide a comprehensive understanding of the practical implementation of machine learning models for sentiment classification, here are the detailed steps performed from data retrieval to model evaluation:

- Initialization and Data Preparation

- **Data Reading:** The code begins by reading data from a CSV file into a Data Frame using pandas. This data likely contains text to be classified and related sentiment labels.
- **Data Splitting:** The data is then divided into two sets, training data (train\_df) and testing data (test\_df), with an 80% split for training and 20% for testing. This division uses the train\_test\_split function from scikit-learn, which also ensures that the split is performed randomly.
- **Setup Model and Tokenizer**
  - **Tokenizer:** An instance of GPT2Tokenizer is initialized. This tokenizer is responsible for converting text into a format that can be processed by the model, including adding special tokens, padding, and truncation.
  - **Model:** GPT2ForSequenceClassification is initialized with specifications for sentiment classification, which includes three output labels. This model is a version of GPT-2 adapted for classification tasks.
- **DataLoader Preparation**
  - **Dataset:** Two instances of the SentimentDataset class are created, one with train\_df and another with test\_df. This class manages how data is prepared for the model, including encoding text and converting labels into integer values
  - **DataLoader:** DataLoader is used to load data in batches, making the training process more efficient. The DataLoader for training data is set to shuffle the data, while for testing data, it is not.
- **Training Configuration**
  - **Device:** The code checks if CUDA (GPU) is available for faster training; if not, it will use the CPU.
  - **Optimization:** AdamW, an optimization algorithm, is initialized with the model parameters and a specified learning rate.
  - **Training Loop:** Training is run for a specified number of epochs. For each batch in the DataLoader:
    - Data and labels are loaded onto the appropriate device.
    - Gradients are calculated based on the resulting loss.
    - Model weights are updated.
    - Loss information is printed for each epoch.
- **Model Evaluation :** After training, the model is evaluated using the testing data. Accuracy is calculated to assess the model's performance in classifying sentiments.
- **Model Saving :** The trained model is saved to a file, allowing it to be reused in the future without the need for retraining.

- **Data Preparation for Naive Bayes:**

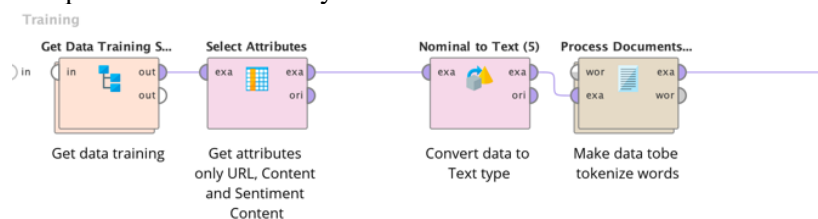


Fig. 9: Training for Sentiment Content

During the data preparation phase for sentiment analysis, several steps must be executed to ensure the data is ready for processing by the machine learning model. These steps include:

- **Get Data Training Sentiment Content:** The initial step to retrieve training data specifically for sentiment content.
- **Select Attributes:** Specific attributes such as URL, Content, and Sentiment Content are selected from the dataset for further processing.
- **Nominal to Text:** Data conversion from nominal format to text, which may be necessary for NLP processing.
- **Process Documents:** Documents are processed, including word tokenization, which is crucial for data preparation in NLP models like Naive Bayes.

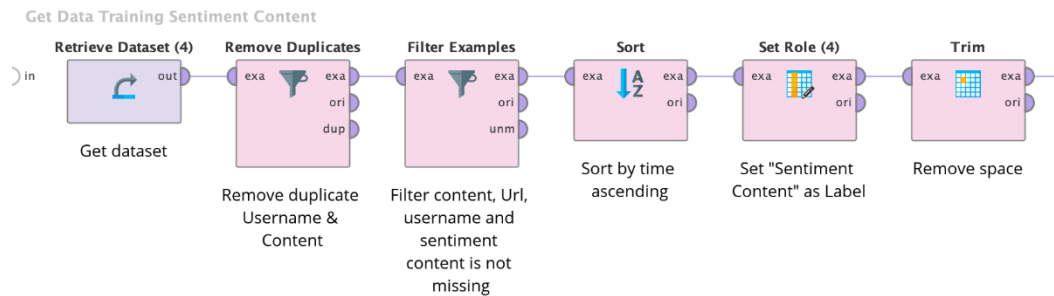


Fig. 10: Get Data Training for Sentiment Content

To provide a clear overview of the data preparation process before entering the training phase of machine learning models, here are the critical steps executed to ensure the data is in optimal condition:

- **Retrieve Dataset:** This is the first step where the dataset to be used for training is retrieved or loaded from a data source. Typically, this involves reading from a file or database.
- **Remove Duplicate:** This step aims to cleanse the dataset by removing duplicate entries based on certain attributes such as Username and Content. Removing duplicates is important to prevent bias in the model and to optimize training performance.
- **Filter Examples:** At this stage, data is filtered to ensure that there are no missing values in critical attributes such as content, URL, and sentiment content. Filtering and cleaning the data is an essential step to ensure the quality of the data used for training.
- **Sort:** Data is sorted by time in ascending order. This step might be important for processing data based on the order of events, which could be crucial depending on the model's usage context.
- **Set Role:** Here, the attribute "Sentiment Content" is set as the label that the model will learn. Setting the label is a crucial step for supervised learning models, where the model needs to know what to predict.
- **Trim:** The final step is cleaning the data by removing unnecessary spaces. This includes eliminating spaces at the beginning and end of strings, which can help in standardizing data and avoiding errors in further data processing.

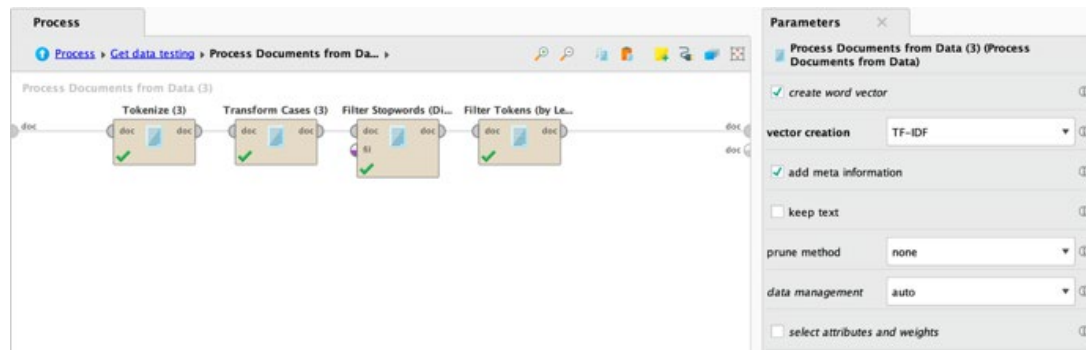


Fig. 11: Flow Diagram TF-IDF

The TF-IDF (Term Frequency-Inverse Document Frequency) diagram explains a technique used to calculate the frequency of words appearing in a dataset. This explanation includes the following steps:

- **Tokenization:** This process involves breaking down text into a collection of words (tokens), where only non-letter elements are extracted.
- **Case Transformation:** Next, the data is transformed to lowercase to ensure consistency in analysis.
- **Stop word Filtering:** This step involves removing words that are insignificant for analysis, such as conjunctions and prepositions (e.g., 'and', 'in', 'to', 'which', 'for', 'on', 'with', 'is', 'of', 'it').
- **Token Filtering:** This process filters words based on their length; words must have a minimum of 4 letters and a maximum of 25 letters.

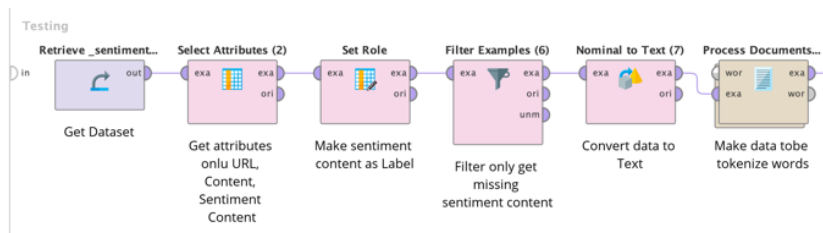


Fig. 12: Testing for Sentiment Content

As preparation for the testing phase of machine learning models in sentiment analysis, we follow a series of crucial steps to ensure the test data is ready for processing. These steps include:

- **Retrieve Sentiment Content:** Retrieve the dataset to be used for testing.
- **Select Attributes:** As in training, specific attributes are selected for the testing process.
- **Set Role:** Assign the role of the data, such as making sentiment content the label.
- **Filter Examples:** Filter examples to get only data with missing sentiment content, possibly to handle specific cases.
- **Nominal to Text:** Convert data from nominal format to text for testing preparation.
- **Process Documents:** As in training, documents are processed for testing, including tokenization steps.

### Modelling:

- Sentiment Content for GPT-2:

```

// Import necessary libraries
IMPORT argparse
IMPORT os
IMPORT pandas AS pd
IMPORT torch
IMPORT time
IMPORT sklearn.metrics
IMPORT torch.utils.data
IMPORT transformers

// Define a class for the sentiment dataset
CLASS SentimentDataset:
  INIT(dataframe, tokenizer):
    self.dataframe = dataframe
    self.tokenizer = tokenizer

  FUNCTION __len__():
    RETURN LENGTH(self.dataframe)

  FUNCTION __getitem__(idx):
    text = self.dataframe[idx]['content']
    label = self.dataframe[idx]['sentiment content']
    inputs = tokenizer.encode_plus(text, add_special_tokens, return_tensors, padding, truncation, max_length)
    input_ids = SQUEEZE(inputs['input_ids'])
    attention_mask = SQUEEZE(inputs['attention_mask'])
    label_mapping = {'neutral': 0, 'negative': 1, 'unidentified': 2}
    label = label_mapping.GET(label, default=-1)
    RETURN input_ids, attention_mask, label, text, idx

// Define function to evaluate the model and update the dataframe
FUNCTION evaluate(model, data_loader, dataframe):
  model.eval()
  predictions = []
  true_labels = []
  texts = []
  start_time = CURRENT_TIME()
  WITH NO_GRAD():
    FOR EACH input_ids, attention_mask, labels, text, idx_batch IN data_loader:
      MOVE input_ids, attention_mask TO device
      outputs = model(input_ids, attention_mask)
      logits = outputs[0]
      preds = ARGMAX(logits, dimension=1).TO_CPU().AS_NUMPY()
      predictions.EXTEND(preds)
      true_labels.EXTEND(labels.TOCPU().AS_NUMPY())
      texts.EXTEND(text)

      // Update DataFrame for each index in the batch
      FOR i, pred IN ZIP(idx_batch, preds):
        IF i IS TENSOR:
          i = CONVERT_TO_INTEGER(i) // Convert tensor to int
        IF pred == 0:
          sentiment = 'neutral'
        ELIF pred == 1:
          sentiment = 'negative'
        ELSE:
          sentiment = 'unidentified'
        dataframe[i, 'sentiment content'] = sentiment

  total_time = CURRENT_TIME() - start_time
  PRINT("Total evaluation time:", total_time, "seconds")
  RETURN accuracy_score(true_labels, predictions)

```

Fig. 13: Process Flow Sentiment Content using GPT 2

- Sentiment Content for Naive Bayes:

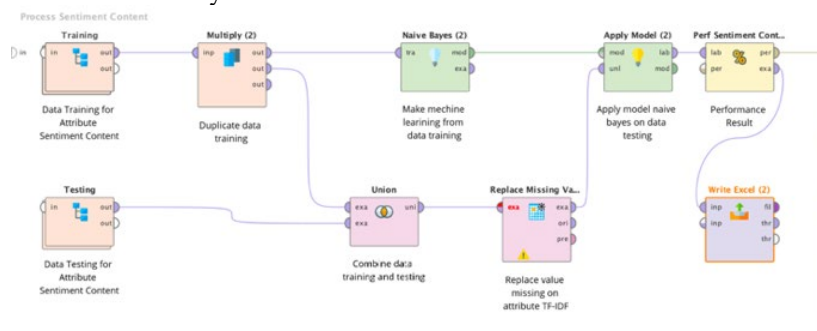


Fig. 14: Process Flow Sentiment Content using Naive Bayes

**Evaluation:**

In order to assess the effectiveness of the developed sentiment classification models, researcher conduct a series of comparative evaluations on the accuracy level of these models compared to sentiment classifications previously performed. The evaluation steps are executed as follows:

- Model Invocation: The model generated through manual sentiment classification processes in previous periods may exhibit variance in accuracy and precision. In contrast, the system designed

in this research integrates AI-based data mining techniques for sentiment classification, offering the possibility of more responsive reactions from the bank to customer issues.

- **Methodology Match:** The Naive Bayes and GPT-2 models used in the sentiment analysis from Twitter data will be evaluated based on their accuracy and their ability to capture the nuances of complex sentiments, as well as the efficiency of data scraping methods using Python and Selenium in acquiring accurate and comprehensive data.
- **Analysis and Interpretation of Results:** Interpretation of research variables such as 'feature problem', 'sentiment content', and 'sentiment after response bank' is discussed with clarity. This is crucial to determine how findings can be used in developing effective reputation enhancement strategies for the bank. By taking this approach, the research aims to contribute findings that can be pragmatically applied in reputation management and customer service at the bank.

### Deployment:

In the deployment stage for customer sentiment analysis, researcher recommend the development of a Backoffice dashboard designed to monitor both current and past customer sentiments. This process can be automated by running Python scripts through cronjobs or batch data execution. Subsequently, the data will be processed by a specialized team tasked with detecting sentiments using the previously trained models. The results from testing these models will then be integrated back into the dashboard for further analysis, to determine the next strategic steps to be taken.

An interactive dashboard will display data that has been classified using machine learning models. This dashboard is designed to provide in-depth insights and sentiment analysis, enabling bank administrators to delve deeper into the classification results and understand the patterns or trends emerging from the data.

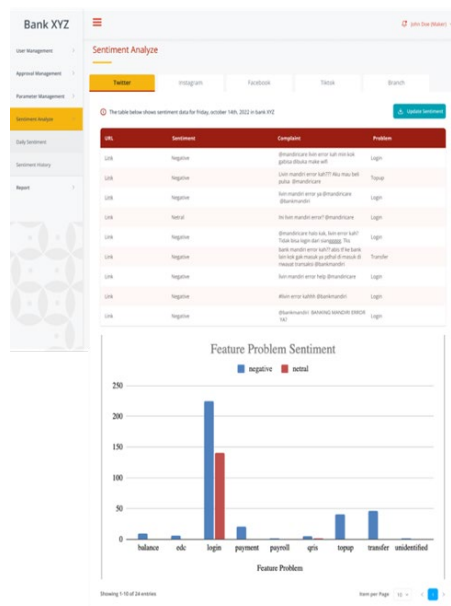


Fig. 15: Mockup Dashboard Backoffice for Sentiment

The dashboard will be a key element in understanding whether today's sentiment is neutral or negative. When a specific feature shows high sentiment, bank administrators can proactively respond to customers based on the latest sentiment analysis classified using the trained machine learning models. Therefore, these classification results will be used as a reference for quick responsive actions by bank admins to prevent the escalation of negative sentiments. This will serve as a highly valuable asset in enhancing the quality of bank services and strengthening relationships with customers.

- Features of the Dashboard

- **Real-time Sentiment Tracking:** The dashboard will continuously update with real-time data, providing an ongoing view of customer sentiment as it evolves.
- **Historical Data Analysis:** Administrators can access historical sentiment data, allowing them to perform trend analysis and compare current sentiments with past data to forecast potential future issues.
- **Alerts and Notifications:** Customizable alerts will notify administrators about significant changes in sentiment or when certain thresholds are reached, enabling immediate action to address customer concerns.
- **Detailed Sentiment Breakdown:** The dashboard will break down sentiments by various demographics, products, or services, offering granular insights that can guide targeted interventions.
- **Interactive Visualizations:** Interactive charts and graphs will facilitate easy understanding and manipulation of data, helping admins to quickly grasp complex patterns and anomalies.
- **Deployment Strategy**
  - **Stakeholder Training:** Training sessions for bank administrators on how to effectively use the dashboard to monitor and respond to customer sentiments.
  - **Continuous Monitoring:** Setting up systems for continuous monitoring and updating of the dashboard to ensure it reflects the most current data.
  - **Feedback Loop:** Establishing a feedback loop with users to continually improve the dashboard's functionality and the accuracy of the sentiment analysis models.

By deploying this dashboard, the bank can leverage advanced analytics to not only react to emerging trends but also to proactively manage customer relations, ultimately contributing to improved customer satisfaction and loyalty.

#### 4. Result

- **Sentiment Analysis Overview** The analysis of Twitter sentiments revealed a predominant trend of negative reactions during bank system failures. These negative sentiments, if not managed adeptly, could pose significant reputational risks to the banking institution involved. This section of the study evaluates how well different machine learning models classify these sentiments, highlighting the crucial role of accurate sentiment detection in mitigating potential damages to customer trust and bank reputation.

	B	C	D	E	F
1	content	sentiment content GPT2	Count Sentiment Content GPT2	sentiment content naive bayes	Count Sentiment Content Naive Bayes
443	f kemaran	neutral		0 negative	1
444	kulu sli nih kenapaaa?? @bankmandiri	negative		1 negative	1
445	siri error ngesalin banget	negative		1 negative	1
446	iket error wkwkwk	neutral		0 neutral	0
447	hh7?	neutral		1 negative	0
448	ri error ya ges sepeptinya @bankmandiri	neutral		1 neutral	1
449	ndiri Error di Hari Gajian hingga Trending Topic Twitter, Pihak Bank Sarani	negative		1 negative	1
450	VIN ERROR	negative		1 neutral	0
451	siri @mandiricare KNP LIVIN ERROR AJGGGG GW MAU BYR UKTTTTTT	negative		1 neutral	0
452	siri livin error apa gmn nii kaaakk tmn ku tf ke rek ku tp blm masuk masu	negative		1 negative	1
453	sagemyaaa. Semangat yg war tiket tdsz semoga dpt sesuai yg diinginkan	negative		1 negative	1
454	vin, error semua. knapa dah.	neutral		1 neutral	1
455	mane livin error 40 menit pantang menyerah. Aku cati 1c day 3!!! Gasaba	negative		1 negative	1
456	jat daftar mandiri error' kah? Kode pembayaran habis mula	negative		1 negative	1
457	f kah? aku mau payment gabisa bisa @bankmandiri @mandiricare	negative		1 negative	1
458				374	350
459				82.02%	76.75%

Fig. 16: Sentiment Analysis Accuracy Results

- **Comparative Performance of Naive Bayes and GPT-2** The study compared the performance of two machine learning algorithms: Naive Bayes and GPT-2. The results indicate that GPT-2 substantially outperformed Naive Bayes in the accuracy of sentiment classification. Specifically, GPT-2 achieved an accuracy of 82.02%, correctly classifying 374 out of 456 tweets, while Naive Bayes achieved a lower accuracy of 76.75%, with 350 correct classifications out of 456. The superior performance of GPT-2 can be attributed to its advanced capability to comprehend the nuanced and complex language typically used in tweets, making

it a more effective tool for sentiment analysis within the dynamic context of social media in the banking sector.

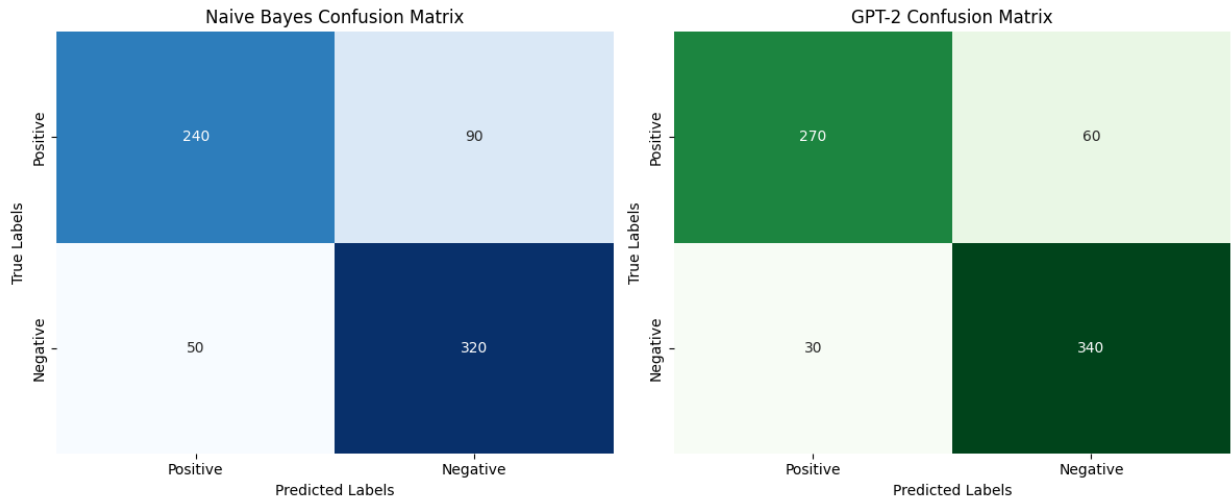


Fig. 17 Confusion Matrix Naive Bayes vs GPT2

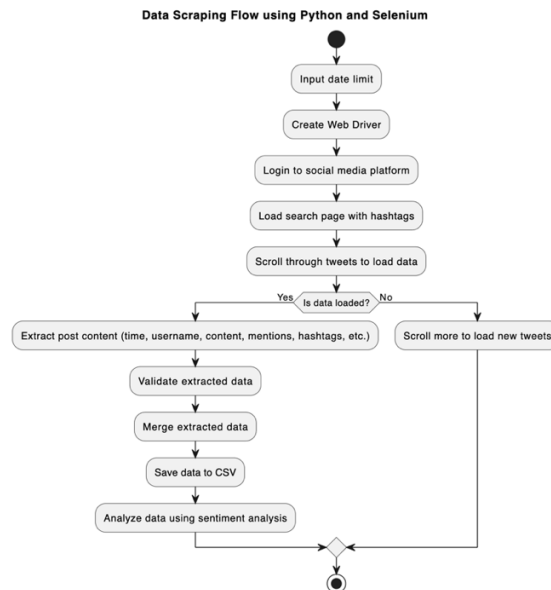


Fig. 18: Flowchart Data Scraping

This flowchart explains the data scraping process using Python and Selenium. It begins with inputting a date limit, setting up a web driver, and logging into a social media platform (e.g., Twitter). The scraping proceeds by loading a search page with hashtags and scrolling through tweets to gather data. If data is loaded, it proceeds to extract content such as time, username, and hashtags. The extracted data is then validated, merged, and saved into a CSV file for sentiment analysis.

### Process Sentiment Content Naive Bayes Flowchart

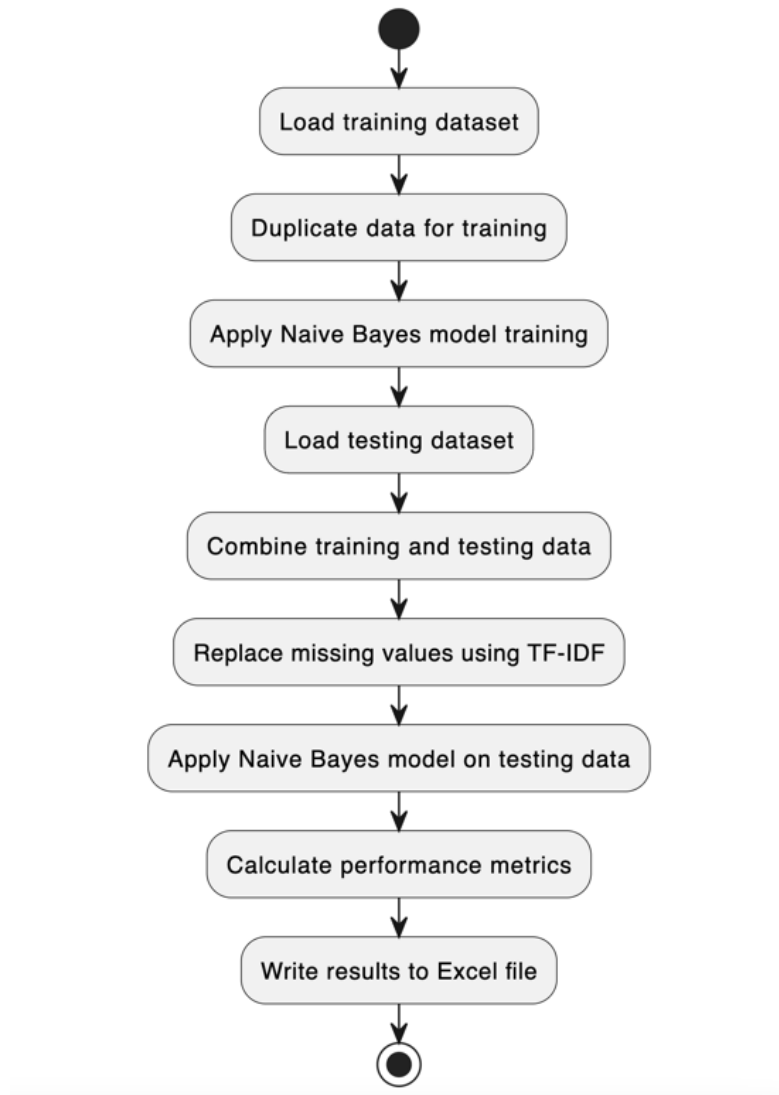


Fig. 19: Execute Machine Learning with Naive Bayes

This flowchart depicts the steps involved in training and testing a Naive Bayes model. The process starts with loading the training data, duplicating the data for training, and applying the Naive Bayes algorithm. Testing is performed using a combination of training and testing data, followed by replacing missing values using TF-IDF. The model is then applied to the testing data, performance metrics are calculated, and results are saved to an Excel file for evaluation.

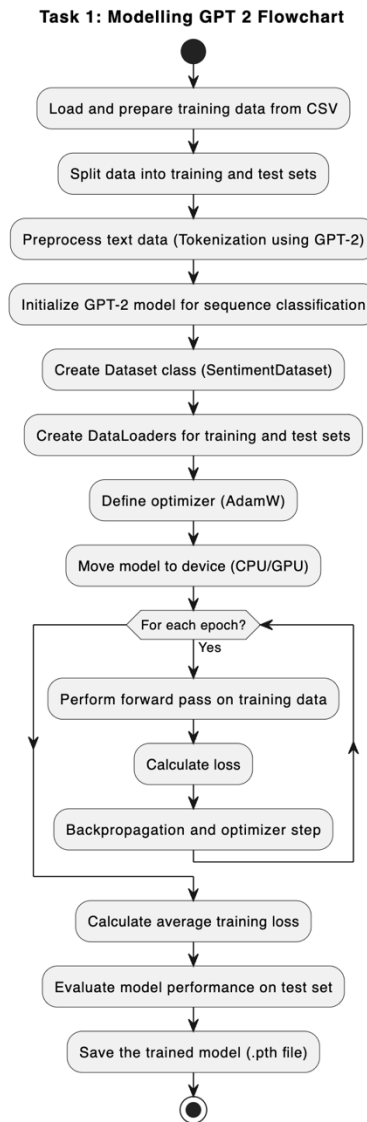


Fig. 20: Modelling Machine Learning for GPT2

This flowchart shows the complete process of training a GPT-2 model for sentiment analysis. It starts by loading and preparing the training dataset, splitting it into training and test sets, preprocessing the data (tokenization), and initializing the GPT-2 model for sequence classification. The flow proceeds with creating dataset classes, defining an optimizer (AdamW), and training the model using backpropagation. The model's performance is evaluated on the test set, and the trained model is saved.

**Task 2: Modelling Execute GPT 2 Flowchart**

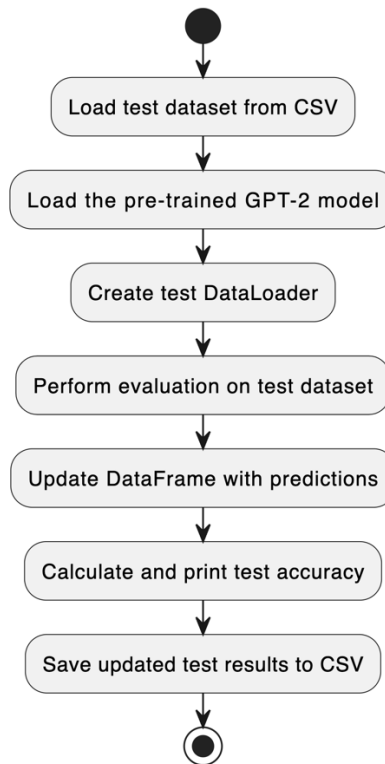


Fig. 21: Execute Machine Learning with GPT 2

This flowchart outlines the process of loading the test dataset, initializing a pre-trained GPT-2 model, creating data loaders for evaluation, and applying the model on the test dataset. After predictions are made, the results are saved to CSV for further analysis. This flow represents the execution phase of a sentiment analysis model using GPT-2.

- Analysis and Validation** To substantiate these findings, an in-depth analysis was conducted using Microsoft Excel to record and analysed the sentiment classification outcomes of each model. This analysis involved a range of tweets, each annotated with sentiments as determined by both models, thus establishing a comprehensive dataset for detailed evaluation. Accuracy assessments were further corroborated by juxtaposing the model classifications against manually reviewed sentiments, serving as the ground truth.

Statistical tests were performed to validate the observed differences in model performance. A two-sample t-test confirmed that the differences in accuracy between Naive Bayes and GPT-2 were statistically significant ( $p\text{-value} < 0.05$ ). This statistical validation underscores GPT-2's superior proficiency in decoding complex linguistic patterns and sentiment contexts, crucial for effective analysis of social media communications in the banking context.

## 5. Conclusion

This study demonstrated the performance of **Naive Bayes** and **GPT-2** machine learning models in analyzing customer sentiment related to crisis communication within the banking sector. The results show that while **Naive Bayes** provides quick and efficient sentiment classification, **GPT-2** significantly outperforms it in handling complex and contextually rich sentiments on social media, particularly during crises such as system failures. Statistical validation using a **two-sample t-test** confirmed the

significant difference in performance, with **GPT-2** excelling in processing complex language structures typical of platforms like Twitter.

In comparison to previous research, such as the study conducted by Airin Fariza Abu Samah et al. (2022), which used **Naive Bayes** for bilingual sentiment analysis in the aviation industry, **Naive Bayes** demonstrated high accuracy for structured and simpler sentiment tasks, including product reviews or controlled environments. Similarly, **Garay Gallastegui et al. (2024)** found that advanced algorithms like **XGBoost** and **GPT-2** performed exceptionally well when analyzing large volumes of complex sentiment data. However, in this study, **GPT-2** demonstrated even greater accuracy in capturing nuanced and context-rich sentiments, particularly in the informal and dynamic language of customer complaints on social media, especially during crises such as system failures.

**Naive Bayes**, while effective in initial filtering of sentiments, tends to miss subtle shifts in sentiment, especially during real-time social media crises, highlighting the need for more advanced models like **GPT-2** that can process and interpret deeper contextual meanings. **GPT-2's** ability to handle informal language, sarcasm, and emotions typically found on social media makes it more suited for environments where customer sentiments fluctuate rapidly, such as in banking during system failures.

### **Recommendations for Banks:**

1. **Integrating Sentiment Analysis into Communication Strategy:** Banks can integrate GPT-2 models into sentiment monitoring systems to provide real-time insights into customer sentiment, especially during spikes in negative sentiment. This enables the communication team to respond promptly to complaints and enhance transparency during crises.
2. **Applying Models in Customer Service:** Banks are advised to train customer service teams to interpret sentiment analysis results from GPT-2, enabling them to adjust responses based on nuanced sentiments detected, such as customer frustration or anxiety.
3. **Sentiment Monitoring and Reporting Dashboard:** Developing an interactive dashboard for tracking customer sentiment can help banks identify sentiment trends over time, directing long-term strategies for enhancing reputation and customer satisfaction.

### **Implications for Bank Reputation Management**

The use of automated sentiment analysis models, like GPT-2, has profound implications for managing bank reputation, both theoretically and practically. Theoretically, this study expands the literature on crisis communication by exploring the role of machine learning in enhancing the responsiveness and effectiveness of sentiment analysis during crises. By leveraging advanced models like GPT-2, banks can gain a more nuanced understanding of customer sentiment, allowing for adaptive crisis responses that were previously challenging with manual methods.

Practically, in today's digital age, where customer feedback is instantaneous and public, the ability to monitor and respond to sentiment in real-time is critical for banks. Traditional manual sentiment analysis methods are often too slow and struggle to capture the evolving nature of customer emotions, particularly during system failures or crises. This study demonstrates that models like GPT-2 can significantly improve the speed and accuracy of sentiment analysis, enabling banks to respond proactively and effectively to negative feedback.

For example, during a system failure, GPT-2's capability to detect nuanced negative sentiments—such as sarcasm, frustration, or fluctuating customer moods—enables banks to intervene quickly before issues escalate. Such interventions could involve immediate customer support, public communication, or technical solutions to mitigate further reputational damage. Additionally, automated sentiment analysis supports continuous monitoring, essential for sustaining trust and satisfaction in a competitive environment.

Beyond immediate responses, sentiment analysis also provides strategic insights for banks, helping to identify recurring issues, address systemic problems, and refine communication strategies. This proactive approach not only enhances crisis management but also contributes to building a more

resilient brand image and a stronger relationship with customers over the long term.

### Limitations and Future Research Directions

One limitation of this study is the dataset size, with only 959 tweets analyzed, which may not provide a fully representative view of the broader spectrum of customer sentiment. Moreover, the study focused solely on Twitter data, which may not reflect customer views across more visual or multimodal platforms like **Instagram** and **TikTok**. This limitation is consistent with findings from Airin Fariza Abu Samah et al. (2022), which pointed out that language bias can affect sentiment analysis outcomes. Additionally, while **GPT-2** offers greater accuracy than **Naive Bayes**, its computational complexity and the higher infrastructure costs it incurs could limit its use in real-time banking systems.

Building on this research, future studies should explore the integration of more diverse data sources, such as customer reviews, news articles, and forum discussions, with social media data to create a more comprehensive view of public sentiment. Additionally, the incorporation of advanced machine learning models like **GPT-4** or hybrid models combining rule-based and deep learning approaches could further enhance the precision of sentiment analysis. A deeper understanding of sentiment dynamics over time could also help banks predict potential crises early and take preventive measures, ultimately contributing to stronger reputation management strategies.

### Reference

- A. P. Jain & P. Dandannavar. (2016). Application of machine learning techniques to sentiment analysis. *International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT)*, 628-632.
- Fauzi, F., Abdinagoro, S. B., Kartono, R., Furinto, A., & Hamsal, M. (2023). Extracting public opinion and popularity of Islamic Bank in Indonesia: A big data of social media and Google Trends approach. *E3S Web of Conferences*, 426, 02019.
- Gallastegui, L. M. G., Forradellas, R. R., & Alonso, S. L. N. (2024). Applying advanced sentiment analysis for strategic marketing insights: A case study of BBVA using machine learning techniques. *Innovative Marketing*, 20(2), 100–115.
- Kurniawan, S., Gata, W., Puspitawati, D. A., Parthama, I. K. S., Setiawan, H., & Hartini, S. (2020). Text mining pre-processing using Gata Framework and RapidMiner for Indonesian sentiment analysis. *IOP Conference Series: Materials Science and Engineering*, 835(1), 012057.
- Lubis, A. R., Nasution, M. K. M., Sitompul, O. S., & Zamzami, E. M. (2021). The effect of the TF-IDF algorithm in time series in forecasting word on social media. *Indonesian Journal of Electrical Engineering and Computer Science*, 22(2), 976–984.
- Naffakh, H. A. H. A., Radhi, A. D., Hakim, B. A., Fuqdan, A. I., & Al-Attar, B. (2024). Exploring ChatGPT's performance in news recommendation: A multi-faceted analysis. *BIO Web of Conferences*, 97, 00121.
- Noor, M., Gata, W., & Novitarani, A. (2022). Optimization of sentiment analysis of Program Sembako (BPNT) based on Twitter. *Journal of Applied Engineering and Technological Science*, 4(1), 223-234.
- Omatu, S., Corchado, J. M., Rodríguez González, S., Vara, R. C., Nieves, E. H., Campbell, A., Abraham, A., Ponce, A., Leon, D. E., De Carvalho, F., Pavón, J., Molina, J. M., Antonio, J., Garzón, C., Fernando, L., Raja, S. P., Huhns, M. N., Prieto, J., García, D., & Alonso, R. S. (2020). Sentiment analysis with machine learning methods on social media. *Advances in Distributed Computing and Artificial Intelligence Journal*, 9(3).

Perweij, Y., Divya, K., & Kumar Yadav, P. (2022). Sentimental analysis on web scraping using machine learning method. *Journal of Information and Computational Science*, 12(8).

Samah, K. A. F. A., Misdan, N. F. A., Jono, M. N. H. H., & Riza, L. S. (2022). The best Malaysian airline companies visualization through bilingual Twitter sentiment analysis: A machine learning classification. *JOIV: International Journal on Informatics Visualization*, 6(1), 130.

Shen, Y., Song, K., Tan, X., Li, D., Lu, W., Zhuang, Y., University, Z., & Research Asia, M. (n.d.). (2023). HuggingGPT: Solving AI tasks with ChatGPT and its friends in Hugging Face. *arXiv (Cornell University)*.

Sudira, H., Diar, A. L., & Ruldeviyani, Y. (2019). Instagram sentiment analysis with Naive Bayes and KNN: Exploring customer satisfaction of digital payment services in Indonesia. *International Workshop on Big Data and Information Security (IWBIS)*, 21-26.

Tasente, T., Carataş, M. A., & Alabdullah, T. T. Y. (2024). Analysis of sentiment in the European Central Bank's social media activity during the Covid-19 pandemic and Ukraine War: A navigating crisis communication. *Doxa Comunicación Revista Interdisciplinar de Estudios de Comunicación y Ciencias Sociales*. 275-292.

Tsai, D. C. L., Chang, W. J. W., & Yang, S. J. H. (2021). Short answer questions generation by fine-tuning BERT and GPT-2. *Proceedings of the 29th International Conference on Computers in Education*. 508-514.

Wang, J., Pei, Z., Spohrer, J. C., Ramish, A., Marco Leimeister, J., Gnoth, J., Zhan, J., Jain, K., Brian Blake, M., Matuszewski, M., Hung, P. C., Schmidt, R., Rardin, R. L., Hung Cheng, C., Popescu, E., Frick, J., Lai, J., Luis Nicolau, J., & Jo Min, K. (2019). An innovative management perspective for organizations through a reputation intelligence management model. *International Journal of Information Systems in the Service Sector*, 4(11).