

Sentiment Analysis for Enhancing Business Process using Naive Bayes

Emmanuel G. Galupo Jr.; Jeffrey F. Calim; Emmie Faye Marione L. Matabile; Johani D. Basaula;
Luisa M. Mariano; Arnel Balasta; Aerob C. Robles; Antoniette C. Mariano
Our Lady of Fatima University College of Computer Studies Quezon City

Abstract:- In order to grow, businesses nowadays must obtain customer feedback, such as reviews or comments. They are thereby collecting additional information. The process of manually collecting and analyzing data is becoming more and more onerous for the owners of these changes. The purpose of this research is to develop an algorithm-based system that can automatically extract data and support business activities. The technology will reduce the effort of human workers in data analysis because it will automatically examine the entered data. It features a sentiment analysis graph. It also offers a word cloud that makes things easy to comprehend for the business administrator by displaying the most relevant keyword in different sizes according to how frequently the system identified the word from reviews or collected data. The system will forecast which department or sector of the enterprise needs improvement. The researchers will build the system using the Iterative System Design Life Cycle since it is most equipped for handling erratic behavioral shifts and even data science. Using brainstorming approaches, the project concept and approach for this study were explored or written. The instruments for requirements formulation, such as customer interviews and system functionality, usability, and security assessments, must be chosen by the researchers.

Keywords:- Sentiment Analysis, Machine Learning, Feed-backs, Sentiment, Multinomial Naive Bayes.

I. INTRODUCTION

In order to understand how to improve the company services, one of the most important phases in the processes of most businesses and other e-commerce firms is to acquire information from the target customers, such as feedback, comments, and opinions. As time goes on, the amount of data that must be gathered grows faster, necessitating a shift in the methods used to collect and organize it. Furthermore, as this generation is already engaged with the "tech world," more recent methods like sentiment analysis are being employed [2]. Companies may have mountains of customer feedback gathered, but in today's world, organizations are suffering from data fatigue (which doesn't imply better or deeper insights). But for us ordinary people, manually analyzing data without prejudice or inaccuracy is never easy. Due to this problem, the researchers developed a system that enhances business

process reviews through the use of machine learning and sentiment analysis[1]. A well-designed system greatly reduces the possibility of bias when using client input. Sentiment analysis is a rapidly developing field in natural language processing that may be applied in business to quickly understand people's opinions, attitudes, and feelings about a product by looking through a large amount of data. The majority of people, who are usually individuals, voice their opinions about various objects or topics. These topics usually include the products or services of a certain company that have received reviews. Sentiment analysis also aims to identify the emotion conveyed in the text, extract relevant information, and analyze it. Business analysis is a talent that experts in business must have. On the other hand, predictive analysis will examine the model to ascertain the best way to do a specific task. Sentiment analysis can be assessed at three different levels: aspect, phrase, and document levels. Finding out if a document conveys a positive or negative entity is the aim of document-level analysis. Sentence-level analysis is performed to determine whether the sentiment expressed in each sentence is an opinion. This stage seeks to ascertain if the statement offers opinions or accurate information. Aspect-level analysis does a great job of classifying sentiment toward entity aspects. At this stage, identify the entity and its parts. The idea of artificial intelligence's machine learning is to develop a computer program that can learn only from data and without human assistance[9]. There are three types of it, and each one is a difficult topic in the field of information technology: supervised learning involves labeling the training data and defining both the input and the output in order to find patterns and correlations. Unsupervised learning use unlabeled data—which only characterizes the input—to discover patterns and significant relationships. Reinforcement learning employs predetermined rules and entails doing an action that, depending on the outcome, will either result in a positive or negative response from the machine [10]. It solves real-world problems by employing a variety of computer approaches to create models using datasets—collections of data that are examined in order to produce the model. Once built, the model may now perform actions, make decisions, and generate forecasts based on the provided dataset. Despite the rapid advancement of technology, several businesses continue to manually organize their data using Excel and paper. Additionally, some people will never be able to recover from an unplanned data loss that could occur to them at any time using that kind of technique, which means the analysis will be disregarded and lose all

relevance. Running a business requires a lot of labor, particularly with regard to data. Nowadays, sentiment analysis may be used by businesses to rapidly determine the polarity of any sentiment, be it an automated review, public comment, opinion, or customer feedback. Additionally, they can employ sentiment analysis to deliver insights or analytics that help business owners make decisions that will further enhance their company. This is made possible via machine learning and sentiment analysis. Whether a feeling is favorable, negative, or neutral, human analysts who assess sentiments based on research tend to agree on its polarity in 80–85%. Evaluating sentiment analysis models using a Naive Bayes algorithm—more especially, Multinomial Naive Bayes—is the aim of the study. This study uses a variety of sentiment analysis approaches in an effort to improve the business process based on end-user assessments or feedback. The technology analyzes the data using a document-level algorithm after automatically extracting it. Using models, the system predicts the sentiment at the document level. As a result, the entire review of a person is used as the data to create a clean version of it, which is then analyzed by the model and free of words like is and similar keywords. It's challenging to run a corporation or business. Many adjustments, tasks, and endeavors must be completed in order to make things right. Taking into account that the business must meet the needs of its clients. As a result, businesses are accepting more and more consumer input, which could potentially be difficult to manage, particularly if there are a lot of comments. Therefore, having a system that could help a business will make receiving customer feedback easier. a system that keeps track of, evaluates, and filters the issue the business is facing. After analyzing the comments, the system will automatically highlight the terms that customers used most frequently. The dashboard would display all of the data that was analyzed, which might help the company find solutions. The objective of this project is to create a system that may enhance business process assessments of the company and lessen the workload related to data analysis. The system automatically analyzes the data that is submitted, and it shows the information and sentiment analysis graph so that the business may assess the aspects of the reviews for possible modifications in the future. The system also has a word cloud that shows the most essential terms in different sizes based on how frequently the system identified them in the assessments, making it easy for the company administrator to rapidly identify the most important words. Technology will help businesses by helping them identify where modifications need to be made and by anticipating the polarity of a customer's attitude or review. The company will benefit from knowing end users' preferences for certain goods or services. The primary objective of this initiative is customer feedback for the company. The business will use this project's machine learning-based sentiment analysis to automatically produce and analyze customer feedback once it has received data from the firm. All of the data that is analyzed and helps the company discover the problem will be generated by the system. This project will be centered around the input that the firm will get from its

customers. The data that the technology will collect will help the company pinpoint its issues. All of the data that has been examined will be shown in the dashboard following a user's submission of feedback that is automatically processed by the system and uploaded to the database. In order to assist the company in identifying common input—whether positive, negative, or neutral—the system will also indicate words that are used a lot. Additionally, the system will display all of the data that has been researched, emphasizing the features of the good or service that the company needs to improve. As long as the data is collected by the company and approved for use in this project, the system is limited to predicting the polarity of data—positive, negative, or neutral—originating from the dataset about customers' sentiments. This includes opinions, comments, feedback, reviews, and surveys. A word cloud that displays the most relevant term and its text size based on how many times it appears in the dataset will be created by extracting and analyzing all the data, including the recently anticipated data that has been added to the primary data store. This can assist the business create a plan and provide insight into the most common comments made about it. Based on the data the system analyzes, a firm may be able to decide how to improve its offerings by displaying the aspects of a specific product or service that could be improved. Furthermore, the system is limited to analyzing spoken English. The only place to create an account is on the login page. The first account to be created will come with an admin role by default; however, later accounts that are created will not have one and will need to be granted one by the admin in order to be able to access the system. Additionally, the system restricts account access to the following three roles: Admin, Staff, and Guest. While the staff job can only view the dashboard, predict, train, and profile pages, the admin position can access all pages, including the admin page. The dashboard page is the sole page accessible to guests. The file input on the anticipated page can only have one column—the review for each row—and must also have the file type extension.csv. The dataset that needs to be imported must have a file extension of.csv and can only have two columns: polarity in the second column and a review or sentiment in the first for each row. The following things have not been included in this project: gathering information, especially customer reviews of the company; creating suggestions or ideas for new ideas to progress the business (the system can only comprehend English at the moment; this could change based on client data); displaying, modifying, and retrieving the password; and notifying users when a new account is created. the process of making a new account for any role using an admin account.

II. RELATED WORK

The growth of e-commerce has increased the competition between the new players and the traditional offline entrants. As a result, social media platforms are introducing the business world to a more practical means of expressing one's thoughts or emotions regarding a certain service within the sector. Because of this, end consumers

frequently inquire about the thoughts and opinions of others regarding particular products or services. As a result, sentiment analysis is becoming more and more important for organizations to help address the lack of certain services by creating improvement initiatives. Electronic word-of-mouth, or E-WOM, is utilized in marketing to inform decisions [2]. Electronic word-of-mouth has been employed as a method for marketing research, and businesses can use social media to get customer feedback. Sentiment analysis combined with eWOM can help analyze a particular issue and provide solutions for the company. This study's format ought to encompass the application portion as well as the theoretical underpinnings of associated ideas. Electronic word-of-mouth is included in the theoretical basis, and specific studies and data are used in the sentiment analysis. Three-stage designs were utilized in the application part for sentiment analysis. Three distinct scenarios, comprising an algorithm and a practical application, are included in each step. These nine scenarios imply that electronic word-of-mouth comprehension is meant to be used in sentiment evaluation [3]. Businesses can benefit from machine learning (ML) in the management and utilization of user-generated data. In order to create an ontology based on the following phases, businesses need machine learning (ML): (1) describing data problems, (2) building solutions centered on fundamental causes, (3) understanding use cases, and (4) creating ontology organization principles [5][6]. "Its algorithms can still learn to perform tasks without being given clear instructions." [13] As an illustration, it can be used to predict the likelihood that complex production equipment will break down. One method for evaluating a client's assessment from derived tweets is natural language processing. Views are seen as outcomes that need to be quantified. This study looks at tweets from the beginning of the lockdown in the Luzon region to the third week. It automatically classifies mentions of the coronavirus and Covid-19 on social media according to sentiment analysis. Sentiment analysis is used to help specialists make decisions based on their emotions. Sentiment research thus reveals that the excessive party quarantine has multiple effects on the majority of Filipinos. According to popular belief, one's basic needs—namely, access to food and government funding—are in danger. While some are considering the positive aspects of COVID-19, it was mistakenly believed that the virus was to blame rather than the actions of Luzon Twitter users who implemented lockdowns, group quarantines, and social distance. It was anticipated that there would be a surge in negative sentiments among Twitter users, as negative sentiments are known to rise with time [7]. Numerous analytical techniques have long been employed to address business-related inquiries concerning the primary methodology of data warehouse research. Most significantly, data mining and business intelligence help decision makers obtain the information they need. Numerous systems are attempting to adjust to the new regulations that have been modified in the past few years. There is no denying big data's allure. In particular, the problem of assessing large amounts of data derived from social media is the most frequently

utilized one in business analytics, and it highlights the importance of incorporating machine learning into business analytics. This highlights the application areas of popular machine learning algorithms for business analytics and introduces their significance [8]. They employed a random forest classifier method due to its proficiency in predictive analysis. It also provides an attribute that evaluates how crucial it is to train a model, like extracting insights. The benefit of this approach is that it has suitable sample size and dimensionality scalability, yields results that are useful, and requires minimal pre-processing of the data. Based on the baseline random performance value of 0.5, which is equal to 50% of the sentiment analysis model that was produced, the algorithm's performance is not very excellent. Studies that can forecast a company sector and provide an explanation based on financial statements have employed machine learning. Financial statements are now contained in a sizable collection of data that is made accessible through open data sets. Prior to now, the primary goal of dealing with such data was to forecast fraud and bankruptcy. However, we can now produce a predictive version that can help achieve the same datasets with a missing information sector. The proposed methodology treats business sector prediction as a classification task by means of a supervised learning approach based on random data. Upon assessing the attribute deemed crucial for the ultimate classification model, we noticed that certain feature scales are utilized for forecasting significant business sectors. Furthermore, an attribute's existence or absence depended on something other than its inherent value. The effect of the insight can be demonstrated in accounting, where it is investigated how company features and financial statements relate to one another [9]. Sentiment analysis extracts information from huge data using text and natural language. These days, people use social media to share their opinions in a variety of ways, such as by leaving comments on particular articles, reviewing products or services, or occasionally submitting their own content on these platforms. Sentiment analysis of data has demonstrated that it can have a significant influence on the decisions made to enhance the government and specific enterprises. Owners of businesses occasionally use user or customer reviews that have been placed to assess issues and improve their services so that they are suitable for their customers [14]. The majority of web applications are written in java, .net, or other web languages other than Python, which makes it extremely difficult to construct machine learning using the Python ML technique. When a web application written in Java is asked to perform a prediction or train a model, for instance, it must communicate with the Python machine learning algorithm through the Java thread. In order to get around this obstacle, Django is a web-based technology that is entirely written in Python and eliminates the need for cross-language communication because the ML algorithm it uses is also written in Python, which can increase its efficiency [15]. One of the difficulties with sentiment analysis is the language barrier because sentiment can be expressed in a variety of languages, including Urdu, English, and many more. The majority of sentiment analysis research are

written in English, and they use WEKA (Waikato Environment for Knowledge Analysis) to classify text in Roman Urdu in order to construct a sentiment analysis for reviews of cars in that language [16]. A study comprising 2000 reviews—1000 positive and 1000 negative—from several car websites was utilized to test various algorithms. The study's findings indicate that the deep neural network classifier has the highest accuracy result at 82%, followed by the decision tree at 75.75%, bagging at 84.5%, random forests at 78.75%, k-NN at 72%, AdaBoost at 83.75%, SVM at 76.5%, and multinomial Naive Bayes at 89.75%. Machine learning-based sentiment analysis is concerned with classifying reviews as either good or negative, and it is inferred through the extraction of thoughts, emotions, text subjectivity, and other elements [26]. In addition, the machine picks up the ability to appropriately assess emotion on its own without human oversight [11]. Since sentiment analysis may be automated to the point that decisions can be made based on a large quantity of data rather than simply gut instinct, which isn't always correct, it will help determine what the most pressing issues are at this time. Sentiment analysis is a study method that helps researchers fully understand how consumers make decisions by analyzing and identifying the feelings and opinions of users or customers [17]. Businesses are beginning to employ sentiment analysis as a result of recent changes to their business models and methods, which include social developments and processes that can help them adapt to the current world. makes use of Amazon reviews to ascertain the neutrality, positivity, and negativity of the comments. Additionally, this study examined the two machine learning algorithms—the Naive Bayes method and the Support Vector Machine (SVM)—that will underpin the sentiment analysis model as a whole by classifying client input into three groups [19]. The imputed dataset will compute the positive, neutral, and negative categorization scores using Sentiwordnet. The evaluation is contingent upon the presentation and can be ascertained by applying the F-1 Measure, Accuracy, Precision, and Recall to each classification. Based on the exploratory results, Naive Bayes classification outperforms Support Vector Machine (SVM) in terms of accuracy. True positive examples (TP), False positive examples (FP), True negative examples (TN), and False negative examples (FN) were used to finish the figures. A sentiment identification classifier can be made using Deep Learning, sometimes referred to as Deep Neural Network, a machine learning component. A few examples of Deep Learning are CNNs (Convolutional Neural Networks), RNNs (Recurrent Neural Networks), and DBNs (Deep Belief Networks). Researchers have been motivated by deep learning approaches because of their potential performance in comparison to traditional methods like Naive Bayes (NB) and Support Vector Machine (SVM) [20]. According to reports, the Word Cloud is reportedly utilized in a number of contexts to generate diagrams by cleaning text in the phrases that appear most frequently. This is typically accomplished on a regular basis as a plain text outline [21]. Word clouds, often called text clouds or tag clouds, are visual representations of keywords arranged according to word

frequency or other criteria, and may include graphic elements such as font, color, and size [27]. Its value lies in screening, analyzing, and comprehending seemingly jumbled textual data; it also exposes key concepts by the new pattern and provides the text's focus. It highlights the most significant terms in a clear and quantifiable manner. By enabling readers to rapidly assess and examine the text, even across a range of texts, a word cloud can assist readers in understanding the theme and connotation from a number of angles. The suggested model includes comparisons, and the outcome implies that stacking ensembles improves prediction accuracy. Support Vector Machine was determined to be more accurate overall than the other two, while K-Nearest Neighbor is substantially less accurate than Naive Bayes and Support Vector Machine. About 20% more precision has been obtained [22]. Two often used techniques for estimating and optimizing sentiment are Support Vector Machines (SVM) and Naive Bayes classifiers. Hierarchical machine learning techniques do ok in classification problems, however SVM and Multinomial Naive Bayes have proven to be effective in terms of precision and optimization. Sentiment analysis using neural network designs has surfaced in a few sentences [23]. The sentiment prediction methods that use deep convolutional neural networks and recursive neural networks are a little more involved in terms of capturing the semantics of phrases. Convolutional neural networks were able to derive word- or sentence-level qualities like stems and morphological tags, however many neural network architectures found it difficult to extract character-level features and embeddings of abstract words. Few researchers have employed J48, BFTree, and OneR for sentiment estimation. On Twitter, these three classifiers are utilized for text categorization and text emotion detection. Sentiment analysis is becoming more and more popular, particularly in the corporate world. With the right methods, techniques, and tools for sentiment analysis concerning businesses, it will provide a solution for decision-making processes to gain some profits [24]. It also has some benefits in any aspect of business practice, such as review of various product and/or service feedbacks, and so on. Sentiment analysis can be applied to a wide range of business endeavours or establishments, including blogs, reviews of books and movies, comments about restaurants, and assessments of products and services [25]. In this case, the company is able to observe customer opinions about them, which they can capitalize on to enhance their offerings. The three main categories of sentiment analysis approaches currently in use are machine learning, lexicon-based, and hybrid approaches. A hybrid algorithm is the most effective methodology for this type of study since it combines deep learning and machine learning to perform better at sentiment categorization [26]. Additionally, the hybrid approach can comprehend sentiment classification in both bilingual and single-language texts. For the purpose of independent research, models such as word vectors, bag-of-words, Naive Bayes Support Vector Machine (NB-SVM), and recurrent neural networks (RNNs) with long short-term memory (LSTM) are examined in the sentiment classification job. As a consequence, the analysis demonstrates excellent accuracy with an 89%

hybrid method result, demonstrating that it can be independently compared to the other technique. Three (3) different types of issues that machine learning (ML) can address are among the many factors to take into account when explaining ML [4]:

- **Classification:** by recognizing objects and comprehending text or speech. Classification incorporates recommendations and correlations through cluster segmentation.
- **Prediction/Estimation:** the ability to anticipate and project future events.

ML "can produce content in this scenario, from interpolating missing data to generating the next frame in a video sequence," for example.

III. METHODOLOGY

The researchers present the structure of the methodology that has been used in the study. This study contains methodology, algorithm, system design or paradigm, stages or phases under system design or

paradigm, and statistical tools that analyze the collected data.

A. System Design

The Iterative System Design Life Cycle was adopted by the researchers to build the system since it is the most suitable method for handling unanticipated changes in behavior and even in data science. In the planning stage, the researchers began to strategize about determining the project's viability and, concurrently, the potential results once development got underway. In order to finish the project, this process frequently includes gathering materials from both humans and machines. During the analysis phase, the researchers looked at the system's database, flow, and structure. This procedure allowed the system to operate rationally. To provide the project a visual representation, every concept that was reviewed during the design phase has been advanced to the next level. Additionally, this is where the researchers assemble the tools needed to develop a system. Making sure the system functions well and accomplishes its objective is the notion behind the earlier steps used in the implementation phase.

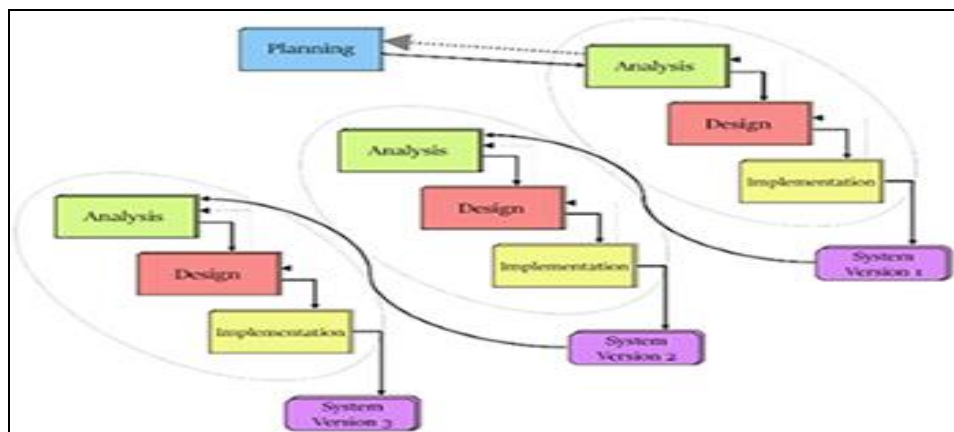


Fig. 1: Iterative System Development Life Cycle

- **Planning:** The planning stage of this study's creation involved the researchers researching current issues with sentiment analysis as well as its problems and prospective remedies. As a result, the researchers developed the topic "Sentiment Analysis using Machine Learning: A System to enhance business process reviews of the Company" and then gathered data from any companies that had feedback or reviews that were pertinent to this study. The main problem with this topic is that humans must manually assess the polarity of sentiment; therefore, the researchers intend to simplify it for people by creating a system. With the assistance of their colleagues and consultants, the researchers created a system that serves as a solution for this analysis. In order to have a success rate in sentiment prediction, the project also employed the most effective and appropriate algorithm.
- **Analysis:** The analysis phase is where the gathering of requirements is done to decide which resources to use in the requirements definition process. In order to build this project, the analysts first interviewed the

client to learn about the existing status of their firm. They next looked at the business flow and processes. To get a better understanding of the situation's facts, the analysts also kept an eye on the business operations. The analysts then examine the problem in order to develop a solution that addresses the issue and a great vision for the improvements the firm would like to see. This is done after monitoring the company's current situation. The researchers will now go on to the length analysis, which necessitates a thorough investigation of how long each step in the creation of this project takes. To produce a clear representation of the data flow in the system, analysts must assess the data procedure.

- **Design:** The system's architectural components were identified by the researchers during the design process in order to build the system's architecture design, which describes the hardware, software, and network environment. The next item that the researchers created after creating the architecture design was a flowchart of the system that shows the flow of algorithms and the sequential order of

actions. In order to assure the interaction between people and computers, the user interface design is a component of the design process in the development of this study. It aids in making this system easier for people to use. In order to highlight development requirements and show how the ultimate structure can function, physical process models are created during the design phase. The functionality of the system in terms of data handling was described by the physical data flow diagram. To create a theoretical framework, the researchers gathered the interconnected study themes. A conceptual framework can help this study understand or visualize how different variables interact with one another. This study was based on sentiment analysis, a technology used in modern enterprises for business analytics [24]. Nowadays, people will actually use an online platform to express their opinions about the goods they have used, the food they have consumed, the movies they have seen, the books they have read, and many other things that can be used as reviews and can help to offer solutions when it comes time to make a decision [25]. Sentiment analysis has a wide range of applications, and companies nowadays utilize it to get client feedback. When processing all of the consumer feedback, machine learning is a business algorithm that can be quite helpful. Machine learning can be used to analyze the typical positives and negatives of a client to create predictions and come to a conclusion. As a result, employing sentiment analysis in company feedback with the use of machine learning as an algorithm can now help generate and evaluate all the same problems at once, identifying their defects and certain things that require improvement. The researchers can create a system that genuinely helps a business owner better their service or product by gathering feedback that aids them in decision-making[9] by utilizing sentiment analysis and machine learning as an algorithm.

Figure 2 illustrates how customer review data is entered into the system and, logically speaking, how the data is cleaned up inside the system using spaCy lemmatization, which reduces words to their simplest forms and eliminates words known as stop words that have no bearing on the meaning of the sentence. The data is preprocessed after cleaning it so that the algorithm can comprehend it. The data was preprocessed and then subjected to sentiment analysis with the Naive Bayes technique to determine the polarity of the data.

- **Implementation:** Programming tasks were carried out at this phase in order to create the concept for the evaluated and planned system structure. To assure machine learning performance, the researchers carried out experiments using

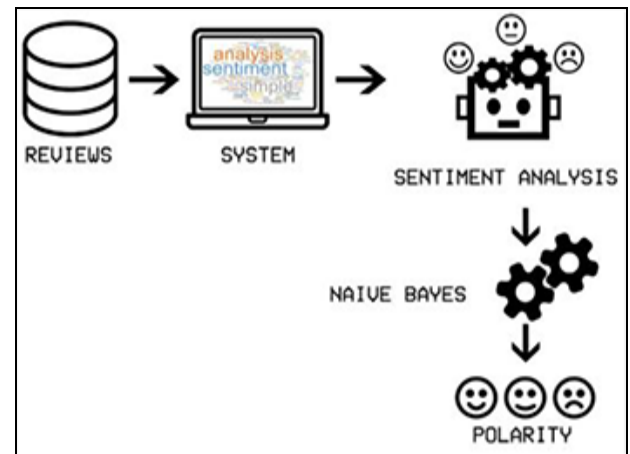


Fig. 2: Conceptual Framework of Sentiment Analysis Review

Jupyter Notebook as the Integrated Development Environment (IDE) and Python 3 as the main programming language. The optimal algorithm for this analysis was selected by the researchers utilizing internet data sets. The researchers created an interactive system design on the Graphical User Interface (GUI) using the Hypertext Markup Language (HTML), Cascading Style Sheets (CSS), Django framework, and Visual Studio Code. The researchers used WampServer and MySQL for the back end. According to what was anticipated and accomplished during the design phase, the system's creation began during this time.

B. Algorithm

In order to improve the system, the machine learning algorithm started to evolve. After some investigation and testing using a dataset of the same kind as the one used in this work that is available online. The stacking classifier from the sci-kit-learn framework was originally going to be used by the researchers since it offers the best accuracy among the algorithms they examined. Additionally, a study reveals that a stacking technique that employs K-Nearest Neighbor, Naive Bayes, and Support Vector Machine as base estimators and Support Vector Machine as the final estimator could boost accuracy by twenty percent (20%). Utilizing stacking algorithms, this research will move forward. The researchers discovered that the system's slow processing was a drawback for the businesses that employed it. As a result, they made the decision to switch to a more effective algorithm called Naive Bayes, which has an accuracy that is comparable to that of the stacking algorithm and faster creation of a model to be used in predicting the review. Additionally, a study comparing various algorithms, including Decision Tree, Bagging, Random Forests, k-NN, AdaBoost, SVM, and Multinomial Naive Bayes, to be utilized in sentiment analysis revealed that Multinomial Naive Bayes produced a superior accuracy than the others. To determine which class a new inputted piece of data belongs to, Naive Bayes calculates the likelihood that each feature in each piece of data in the dataset will occur based on its class.

C. Population and Sampling

Because not every member of the population has an equal opportunity to be a sample or a respondent, the researchers utilized convenience sampling, also known as opportunity sampling, in this study. A form of bias in the selection of the respondents is non-probability sampling. In order to include the greatest number of participants in this study, the researchers use a technique called convenience or incidental sampling. Three (3) IT professionals who met the following criteria made up the respondents for the researchers: male or female; 21–50 years old; IT degree holder or IT professional; Filipino citizen or foreigner; and eager to participate. Additionally, three (3) end users who work in management are chosen based on a variety of characteristics, including gender, age (must be between 21 and 50), nationality (must be either Filipino or foreign), and willingness to participate. There are six (6) respondents total, all of whom must reside in or near the Philippines.

D. Data Gathering Procedure

The researchers employed data-gathering techniques to learn more about the problem they are studying. Through observation and survey questions, the study's data are gathered. A variety of inquiries about the project's effectiveness were included in the survey. Additionally, instructions on how to evaluate the developed system and questionnaire were sent to the responders. The researchers gave the respondents enough time to use the system and complete the survey questionnaires. In this study, the researchers contacted the chosen respondents by email, Facebook Messenger, and other social media platforms to see if they satisfied all the requirements for the population and sampling. If they did, they were immediately eligible. The researcher gives the qualified respondent an information sheet and consent form so they can learn more about the study and decide whether or not to participate on their own volition. The individual can start the trials once they have signed the consent form. The participant in the study examined the system for a maximum of thirty (30) minutes before answering a questionnaire. The trial cycled three (3) times. The procedure was carried out online utilizing Remote Desktop and Google Forms for the qualified participants who signed the consent form. Following the system testing, the participant completes the questionnaire using a Google form. The participant first tests the system via Remote Desktop, which enables the participant to test the researcher's system.

E. Statistical Tool

To collect the study's data, the researchers prepared survey forms or questionnaires. The 4-point Likert Scale was employed by the researchers to evaluate the respondents' answers to their proposed plan. Following the respondents' use of the produced application and completion of the questionnaire, the researchers clarified the responses given by the respondents and evaluated the degree of agreement on functional adequacy, usability, and security. For the questionnaire to be both accurate and effective, the prepared questions must be created precisely using the

weighted mean and worded clearly and easily understood. The standard formula or statistical tool is used to carefully assess and handle the data collected by the researchers. The average of all numbers is calculated by adding all the numbers together and dividing the result by the total number of numbers. It is a statistical measure of the central tendency of a probability distribution.

Table 1: Likert Scale

Ratings	Interpretation
3.51 – 4.00	Strongly Agree
2.51 – 3.50	Agree
1.51 – 2.50	Disagree
1.0 – 1.50	Strongly Disagree

The researchers also used the calculation of percentages with the respondents' surveys. The formula to get the 0% to 100% is:

$$\text{Step value} = 100 / (\text{categories} - 1) \quad \text{Step value} = 100 / (4 - 1)$$

$$\text{Step value} = 33.33\%$$

The calculation of percentage needs to calculate first the maximum possible product, and this is the formula:

$$\text{Maximum possible points} = \text{Total number of responses} \times \text{Maximum percent}$$

$$\text{Maximum possible points} = 9 \times 100\% \quad \text{Maximum possible points} = 900\%$$

The formula to get the percentage of the arithmetic mean is:

$$\text{Percentage} = \text{Sum of the Products} / \text{Maximum Possible}$$

$$\text{Points} \times 100 / 1 \quad \text{Percentage} = 866.67 / 900 \times 100 / 1$$

$$\text{Percentage} = 96.30\%$$

F. Research Ethics

The researchers are looking for IT experts who can fulfill a number of criteria, including being male or female, between the ages of 21 and 50, possessing an IT degree or working as an IT professional, being a Filipino citizen or a foreigner, being able to read and write, and being ready to participate. Additionally, management workers that are end-users have various selection criteria: Male or female; between the ages of 21 and 50; employed by the client's company; of Filipino nationality or another nationality; literate and eager to participate. No payment or reimbursement of any kind is permitted. The researchers used social media posts and other recruitment techniques to urge respondents to participate in the study. Any data that the researchers obtained for this study is kept in a safe folder with a password that only the researchers may access. Only the researchers working on this study have access to your data in order to protect your privacy. Online research investigation came to a conclusion in order to reduce social

danger. During a pandemic, researchers use online platforms to direct information to minimize health risks. The responders are also given instructions and are told that the researchers are happy to offer any clarifications if necessary. To maintain its confidentiality and safeguard your privacy, the data gathered for this study is kept in a password-protected folder on a private cloud that can only be accessed by the researcher. And deletion is the method used to discard this material. To keep it private and safeguard your identity, personal information like your name was substituted with encrypted text, of which only the researchers were aware of the encryption method utilized. As soon as the research is over, all information about you is destroyed by deleting the data from the encrypted folder. By signing the Certificate of Consent, you grant the researchers the right to collect, use, and disseminate the information you have submitted for this project. Information that is necessary for this study to be successful. The community gains from this research by receiving a system that can aid the companies taking part in it in enhancing their community services. It might be a useful resource for academics working on related subjects in the future.

IV. RESULT AND DISCUSSION

The researchers successfully devised a method for predicting the polarity of sentiment by gathering the client’s concerns and studying them with each member to deliver the most efficient method for predicting the polarity of sentiment. After the researchers discussed the solution to

the problem, they began to allocate roles to everyone for them to do the responsibilities assigned to them. The duties are collecting the data to the client in order to train the model or algorithm by the system, creating the user interface of the system to interact with the user, developing the back-end of the system to operate and predict the given data from the client, and documenting the necessary information that is relative to this study. Every business is aware that reviews of their customers are important. Therefore, getting the customer’s review is necessary, and by doing that, they will know if the business needs to improve to meet the satisfaction of their customers. That is why researchers studied and evaluated the client’s data to provide information that may be beneficial for the improvement of the company’s product or service while focusing on how to utilize time management in getting the reviews and processing them to have the analyzed data immediately. By the efficiency of this system, the company can be informed of the results that they gathered and can improve their services if needed.

Table 2 displays the Evaluation Results summary of the IT Expert in three (3) different criteria assisting the potency of the system. The criterion with the highest mean of 3.59 is the Functional Suitability of the system, which is equal to Strongly Agree then comes with the Security criterion with a 3.55 mean also having the same interpretation, and lastly the Usability criterion with 3.29 with the interpretation of Agree.

Table 3: Evaluation Result from Experts

Criteria	Mean	Percentage	Interpretation
Functional Suitability	3.59	86.42%	Strongly Agree
Usability	3.29	76.54%	Agree
Security	3.56	85.19%	Strongly Agree
Total	3.48	82.72%	Agree

Finally, the overall mean of the criteria is three-point-forty-five3.45 which rated the system as Strongly Agree.

Table 3: Evaluation Result from Endusers

Criteria	Mean	Percentage	Interpretation
Functional Suitability	3.15	71.61%	Agree
Usability	3.44	81.48%	Agree
Total	3.30	76.55%	Agree

Table 3 displays the Evaluation Results summary of the End- Users in two (2) different criteria assisting the potency of the system. The criterion with the highest mean of 3.44 is the Usability of the system, which is equal to Agree then comes with the Functional Suitability criterion with a 3.44 mean with also having the same interpretation, as well as the total mean of both with a 3.29 mean which evaluated as Agree.

V. CONCLUSION

This chapter includes a summary of the study "Sentiment Analysis using Machine Learning: A System to enhance business process reviews of the Company." Conclusions and suggestions are also included. While conducting the study, the researchers were aware of the significance of sentiment in society, notably in business. Because they want to know which aspects of their business need to be addressed, the majority of businesses are worried about their reputation. The researchers carried out this experiment to help businesses achieve a more contemporary method of data analysis because some

businesses find it challenging to evaluate customer feedback. By conducting surveys and interviewing people who were required to participate in the study, the researchers collected the data for this project. By analyzing information using the system's machine learning algorithm or model, the researchers' newly built system may forecast the supplied end-user sentiment. To make the system work properly, the end user must train a model using a dataset. The accuracy of the model will be shown if it is complete, allowing the end user to judge its suitability. In order to help the end user evaluate the processed data, the system then displays the analyzed data on the dashboard. The researchers came to the conclusion that the developed system could predict the polarity of the sentiment by having an average percentage of 76.55% with a mean value of 3.30, interpreted as agreed for the end-users functional, suitability, and usability, while on the other hand, the average percentage given by the IT expert is 82.72% with a mean value of 3.48. As a result, the system is currently usable and capable of assisting the client in data analysis. Future studies should improve the system by automatically converting Tagalog or any dialect to English. This will spare the system's user the time-consuming task of manually deciphering client input. Additionally, as mentioned by one of the end-users during the study, the system should offer the option to display ratings rather than polarity, according to the researchers. In order to save some time while retaining part of their jobs, the researchers will also suggest this technique to businesses that currently manually examine customer reviews. In light of this, the researchers propose that the developed system be a software program that can be accessed from any device, particularly a mobile one. Additionally, the researchers recommend that the system

ACKNOWLEDGMENT

The researchers would like to extend their sincere gratitude and appreciation to everyone who took part in this study. Their tireless efforts, knowledge, and assistance were crucial in enabling this investigation. Their continuous support, contributions, and cooperation have greatly improved this research and helped us accomplish our objectives. Finally, we would like to extend our sincere gratitude to all of the volunteers and study participants without whom the data collection and analysis would not have been possible. We really appreciate your excellent contributions to this study.

REFERENCES

- [1]. Dumbleton, R. (2021). Sentiment analysis: The complete guide to sentiment analysis. Retrieved April 01, 2021, from <https://getthematic.com/insights/sentiment-analysis/>
- [2]. Seo, S., Kim, C., Kim, H., Mo, K., Kang, P. (2020). Comparative study of deep learning-based sentiment classification. *IEEE Access*, vol. 8, pp.6861-6875, 2020, doi: 10.1109/ACCESS.2019.2963426.
- [3]. Canbolat, Z. and Pinarbasi, F. (2020). Using Sentiment Analysis for Evaluating eWOM: A Data Mining Approach for Marketing Decision Making. 10.4018/978-1-52258575-6.ch007.
- [4]. Henke, N., Bughin, J., Chui, M., Manyika, J., Saleh, T., Wiseman, B., & Guru, S. (2016). *The Age of Analytics: Competing In A Data-Driven World*. Retrieved from <https://www.mckinsey.com/me-dia/mckinsey/business-functions/mckinsey-analytics/our-insights/the-age-of-analytics-competing-in-a-data-driven-world/mgithe-age-of-analyticsfull-report.ashx>
- [5]. Pons-Muñoz de Morales, S. (2020). Big data and sentiment analysis considering reviews from e-commerce platforms to predict consumer behavior. *Treballs Finals del Màster de Recerca en Empresa, Facultat d'Economia i Empresa, Universitat de Barcelona, Curs: 2019-2020*. Retrieved from <http://diposit.ub.edu/dspace/handle/2445/173181>
- [6]. Earley, S. and Bernoff, J. (2020). Is Your Data Infrastructure Ready for AI? *Harvard Business Review*, 1-5. Retrieved from <https://hbr.org/2020/04/is-your-datainfrastructure-ready-for-ai>
- [7]. Pastor, C. (2020). Sentiment Analysis of Filipinos and Effects of Extreme Community Quarantine due to Coronavirus (COVID-19) Pan- demic. *Poblacion, Lingayen, Pangasinan, Philippines*, 7(7), pp. 91-95.
- [8]. Okatan, Kağan. (2020). Machine Learning for Business Analytics. 10.4018/978-17998-2566-1.ch013.
- [9]. Angenent, Mitch & Pereira Barata, António & Takes, Frank. (2020). Large-Scale Machine Learning for Business Sector Prediction. 10.1145/3341105.3374084.c
- [10]. Tyagi, N. (2020). 6 Major Branches of Artificial Intelligence (AI). *Analytics Steps*. Retrieved from <https://www.analyticssteps.com/blogs/6-major-branches-artificialintelligence-ai>.
- [11]. Bhatt, N. and Swarndeep, S. (2020). Sentiment Analysis using Machine Learning Technique: A Literature Survey. *International Research Journal of Engineering and Technology (IRJET)*, 7(12), pp. 798-802.
- [12]. Barba, P. (2019). Sentiment Accuracy: Explaining the Baseline and How to Test It. Retrieved from <https://www.lexalytics.com/lexablog/sentiment-accuracy-baseline-testing>.
- [13]. Dubovikov, K. (2019). *Managing Data Science*. Packt Publishing Ltd.
- [14]. Rokade, P. and Kumari, D.A. (2019). Business intelligence analytics using sentiment analysis-a survey. *International Journal of Electrical and Computer Engineering (IJECE)*. 9. 613. 10.11591/ijece.v9i1.pp613-620
- [15]. Ranjith, P. R. (2019). ML Model Management using Python Django Web Framework. In *IEEE India Info Vol. 14 No. 3* (pp. 140-144).
- [15]. Khan M., Malik K. (2019). Sentiment Classification of Customer's Reviews About

- Automobiles in Roman Urdu. In: Arai K., Kapoor S., Bhatia R. (eds) *Advances in Information and Communication Networks*. FICC 2018. *Advances in Intelligent Systems and Computing*, vol 887. Springer, Cham. https://doi.org/10.1007/978-3-030-03405-4_44
- [16]. Saura, J.R., Reyes-Menendez, A., Alvarez-Alonso, C. (2018). Do online comments affect environmental management? Identifying factors related to environmental management and sustainability of hotels. *Sustainability*, 10, 3016.
- [17]. Kopera, S., Wszendybył-Skulska, E., Cebulak, J., Grabowski, S. (2018). Interdisciplinarity in Tech Startups Development—Case Study of ‘Unistartapp’ Project. *Found. Manag.*, 10, 1–10.
- [18]. Vanaja, S. and Belwal, M. (2018). Aspect-Level Sentiment Analysis on ECommerce Data. 2018 International Conference on Inventive Research in Computing Applications (ICIRCA), 1275- 1279.
- [19]. Behdenna, S., Barigou, F., Belalem, G. (2018). Document Level Sentiment Analysis: A survey. *EAI Endorsed Transactions on Context-aware Systems and Applications*. 4.154339. 10.4108/eai.14-3-2018.154339.
- [20]. Kabir, A., Karim, R., and Newaz, S. (2018). The Power of Social Media Analytics: Text Analytics Based on Sentiment Analysis and Word Clouds on R. *Informatica Economică*, p. 25.
- [21]. Tribhuvan, Padmapani & Bhirud, Sunil & Deshmukh, Ratnadeep. (2018). Stacking Ensemble Model for Polarity Classification in Feature Based Opinion Mining. *Indian Journal of Computer Science and Engineering*. Vol. 9. 91-95. 10.21817/indjse/2018/v9i3/180903004.
- [22]. Singh, J., Singh, G., and Singh, R. (2017). Optimization of sentiment analysis using machine learning classifiers. *Hum. Cent. Comput. Inf. Sci.* 7, 32. <https://doi.org/10.1186/s13673-017-0116-3>
- [23]. Ziora, L. (2016). The sentiment analysis as a tool of business analytics in contemporary organizations. *Uniwersytetu Ekonomicznego w Katowicach, Zeszyty Naukowe*, p. 240.
- [24]. Hamdan, H., Bellot, P., Bechet, F. (2016), *Sentiment Analysis in Scholarly Book Reviews*, <http://arxiv.org/abs/1603.01595v1> (access: 4.03.2016).
- [25]. Liu, G., Xu, X., Deng, B., Chen, S., Li, L. (2016). A hybrid method for bilingual text sentiment classification based on deep learning. *Proc. 17th IEEE/ACIS Int. Conf. Softw. Eng. Artif. Intell. Netw. Parallel/Distrib. Comput. (SNPD)*, pp. 93-98.
- [26]. Huang, Y., Wang, Y., & Ye, F. (2019). A Study of the application of word cloud visualization in college english teaching. *International Journal of Information and Education Technology*, 9(2), 119-122.