SENTIMENT ANALYSIS OF THE SWOTIFY PLATFORM OF CENTRAL STUDENT GOVERNMENT IN CAVITE STATE UNIVERSITY- IMUS CAMPUS

Research Project
Submitted to the Faculty of
Department of Computer Studies
Cavite State University
City of Imus, Cavite

In partial fulfillment
of the requirements for the subject of
COSC 200A - UNDERGRADUATE THESIS

SVET JAZMINE DIO
TRISHA GRACE C. GARCIA
VAN ELLIS V. MERCADO
2024

TABLE OF CONTENTS

LIST OF FIGURES

ChavSU: AI-DRIVEN CHAT ASSISTANT FOR CAVITE STATE UNIVERSITY - IMUS CAMPUS

Dio, Svet Jazmine
Garcia, Trisha Grace C.
Mercado, Van Ellis V.

A research project manuscript submitted to the faculty of the Department of Computer Studies, Cavite State University Imus, Cavite in partial fulfillment of the requirements for the degree of Bachelor of Science in Computer Science with Contribution No. _____. Prepared under the supervision of Mrs. Mildred Apostol.

CHAPTER 1 - THE PROBLEM AND ITS BACKGROUND

INTRODUCTION

Sentiment analysis ascertains the text tone to understand viewpoints, opinions, and emotions using natural language processing (NLP). Furthermore, sentiment analysis mainly serves by categorizing writings as positive, negative, or neutral emotions, considering the subjective information, such as personal viewpoints and perspectives, found in the source material. It also recognizes more complex emotions throughout the study, such as joy, rage, or melancholy. Various fields and sectors, such as healthcare, business and finance, education, politics, and others heavily reliant on understanding and interpreting human opinions and emotions, have experienced substantial

impacts from sentiment analysis. According to Mehta and Pandya (2020), getting other people's perspectives is essential when coming up with a decision or producing the result from evaluations that include valuable resources and the experiences of other individuals. In these diverse domains, sentiment analysis has emerged as a transformative tool, enhancing decision-making processes and generating valuable insights from data.

User reviews are commonplace on e-commerce and movie review websites, offering valuable feedback on purchased items and experienced films. Going beyond surface opinions, some users employ sentiment analysis, a technique for analyzing emotions, to gain a deeper understanding of their audience's reactions and inform potential improvements to their product or offering. Through applying diverse analytical tools, several websites gain refined insights into user sentiment. Similarly, Strengths, Weaknesses, Opportunities, and Threats (SWOT) analysis is a common practice among businesses for classifying and prioritizing product or service improvement areas. One of the strategic plans that the Central Student Government Academic Year 2023-2024 sought to foresee for the next five years was the SWOTIFY, inspired by the SWOT Analysis for businesses. It is an ongoing effort and strategy for all Central Student Government officers to review and evaluate themselves and their co-officers inside the organization. It will be put in place afterward for an event or project. It is an online platform where users may examine their strengths, shortcomings, project potential, and activity threats.

Following the principles of SWOTIFY, sentiment analysis emerges as a potent method for extracting valuable insights from unstructured textual data. This technique enables officers to conduct informed assessments and evaluate their performance in various actions or events. By offering a quantifiable measure of subjective information, sentiment analysis facilitates the evaluation of public opinion, user satisfaction, and other factors that influence decision-making processes.

Background of the Problem

The research is centered around examining the emotional tone of the evaluations of the Central Student Government officers, which are conducted through a platform called SWOTIFY. The purpose of SWOTIFY is to regularly evaluate and analyze the performance of the officers within the organization, including their fellow officers. The approach taken by the Central Student Government of Cavite State University-Imus Campus is influenced by the SWOT Analysis method commonly used in the business world. As student leaders, it is their top priority to effectively carry out their duties and responsibilities. Being the sole governing body representing the student body, they must cultivate a sense of duty and responsibility among the students. This initiative aims to empower students to capitalize on educational opportunities for the collective unity and well-being of the CvSU-Imus Campus community. To maximize their time and resources, each group of officers within an academic year at Cavite State University-Imus Campus collaborates closely to plan and execute their activities and projects in a condensed period. As a result, this research aims to utilize sentiment analysis to comprehend the content of their evaluations and determine the emotions conveyed in the assessments.

Statement of the Problem

In general, this study seeks to answer the question, *“To what extent can SWOTIFY’s Sentiment Analysis contribute to the comprehensive evaluation and mutual assessment of the included individuals' sentiments within the Central Student Government, encompassing self-assessment and co-officers evaluation?”*

Specifically, this study attempts to answer the following problems:

To characterize feelings inside the performance assessments on the SWOTIFY platform, this study sets out to find the most accurate model. *“Given the preprocessed performance evaluations from the SWOTIFY platform, what machine or deep learning model achieves the highest accuracy in classifying sentiment (positive, negative, neutral)?”*

The study analyzes how the performance evaluations construct together an overall emotional dominance. *“Using the chosen model with the best classification performance, what are the dominant sentiment trends (positive, negative, neutral) revealed in the performance evaluations as a whole?”*

Are there significant correlations between the sentiments expressed in the evaluations and specific facets of leadership, such as decisiveness, resource utilization effectiveness, and communication style, as identified by the chosen model?

Beyond overall sentiment, what nuanced interpretations can be drawn from the evaluations regarding specific leadership aspects like decision-making clarity, effective resource allocation, and communication style?

Objectives of the study

Generally, the study aims to derive insights from the performance evaluations given by fellow officers utilizing the SWOTIFY platform through sentiment analysis.

Specifically, this study also aims to:

- create a machine or deep learning model using the preprocessed performance evaluations from the SWOTIFY platform;
- choose the model that shows the best performance in terms of the classification metrics set for this study;

- determine and evaluate insights uncovered using the chosen model's overall sentiment score; and
- provide a detailed interpretation of the sentiments using aspects of leadership, including (but not limited to) decision-making, institutional efficacy, and communication skills.

Scope and Limitation

The sentiment analysis emphasizes the functionality of the sentiment analysis and focuses on creating solutions for assessing students' opinions and feedback for the Central State Government at Cavite State University-Imus Campus. In this section, it will discuss the scope and limitations of the application. It will discuss the possible strengths and weaknesses and their impact on the potential features.

Data Ingestion Module. This module involves the ingestion of the textual data from the SWOTIFY platform. This may include a manual interface for uploading the dataset coming from the designated source, or an automated download of the dataset from a database or a repository containing the feedback coming from the SWOTIFY platform.

Preprocessing Module. This module handles the overall cleaning, curating, and preprocessing of the textual dataset to be used in this study. This includes proper identification of the textual dataset, removing stopwords, and implementing necessary NLP techniques. The performance of this SWOTIFY sentiment analysis software application is highly dependent on the quality of the training data, which needs to be carefully cleaned, organized, and labeled.

Machine Learning (ML) Module. This module is mainly focused on identifying sentiment analysis through machine learning algorithms. This

includes the actual training of the dataset to create a series of data modeling tasks implementing multiple algorithms to come up with the best model possible to identify the context and sentence-based sentiment analysis of the feedback captured from SWOTIFY. This will also include the evaluation of the models trained using the metrics appropriate for the machine learning problem, such as accuracy, precision, and recall.

Prediction Module. This module will enable the actual prediction of the unseen data that can be inputted in the future. This helps in automatically recognizing sentiments in a text whenever they occur.

Significance of the Study

Effective communication symbolizes understanding and positive communication between Central Student Government officers, the students, and the university. The sentiment analysis on SWOTIFY will be used to evaluate themselves and their fellow officers within the organization. The Central Student Government and the students at Cavite State University-Imus will both benefit from this study. Sentiment Analysis determines the level of positivity and negativity in a text, as well as the subject and opinion holder. CSG officers can assess the effectiveness of their teamwork, communication, and leadership styles using sentiment analysis, identifying areas of strength and potential areas for improvement. Students benefit from a platform that not only increases the volume of their voices but also ensures that their concerns are heard and addressed as soon as possible. Moreover, this study will benefit the following:

Central Student Government. By identifying the problems or concerns that students have within the student government, this sentiment analysis will assist the Central Student Government (CSG). This will create

and preserve a positive campus culture by fortifying the bonds between the student government and the larger student body.

Cavite State University Students. The result of sentiment analysis will affect the Central Student Government's future performance. The study offers perceptions of the feelings of the CSG, making it easier to comprehend the atmosphere as a whole. The CSG can identify possible problems or concerns early on due to the real-time analysis feature, which enables prompt interventions and solutions.

Future Researchers. The results of this study will allow future researchers to use the methodology created by this study to modify sentiment analysis to fit a particular linguistic context. Future researchers can use the study's findings to guide their work in adapting sentiment analysis models to other specific linguistic communities.

Definition of Terms

The following terms, which are crucial to the study's documentation but may not be familiar to everyone, have been utilized by the proponents.

Central Student Government

Technical definition: also known as CSG, is the organization that holds the title of the highest student governing body in Cavite State University - Imus.
Operational definition: CSG functions as the institutional setting for co-officer evaluations and assessments of SWOTIFY.

Co-officers

Technical definition: refers to the members or individuals who hold positions within the Central Student Government.

Operational definition: the subject and population of utilizing SWOTIFY for sentiment analysis.

Lexicon

Technical definition: used to describe a carefully chosen, domain-specific set of terms or words that have polarity labels or sentiment scores attached to them.

Operational definition: intended to express the sentiment found in textual data in the SWOTIFY.

Machine Learning (ML)

Technical definition: refers to how computational models and algorithms are applied to help automated systems identify and learn from sentiment patterns found in textual data.

Operational definition: is the process of automatically analyzing and classifying the sentiments expressed on the SWOTIFY through the use of statistical models and computational algorithms.

Natural Language Processing (NLP)

Technical definition: a field in Artificial Intelligence (AI) that focuses on the interaction between human language and computers by enabling the development of algorithms and computational models for computers to understand, interpret, and generate contextually relevant and meaningful human language.

Operational definition: integral to SWOTIFY's processing and analysis of textual data regarding co-officer evaluations, enhancing precision and quality in sentiment analysis.

Natural Language Toolkit (NLTK)

Technical definition: refers to a specific library and toolkit for natural language processing tasks within sentiment analysis. An extensive collection of

resources and tools created especially for processing and analyzing data in human languages is offered by NLTK.

Operational definition: this can be used for tasks like tokenization, part-of-speech tagging, and sentiment polarity analysis in the context of sentiment analysis using the initiative of SWOTIFY.

Scoring

Technical definition: a quantitative or qualitative approach for allocating entities, numerical values, or ratings based on predetermined standards.

Operational definition: frequently entails assessing the degree of positivity, negativity, or neutrality expressed in textual data and assigning values corresponding to those sentiments' intensity or polarity.

Sentiment Analysis

Technical definition: it is also known as opinion mining; typically refers to a computational procedure that examines and ascertains the sentiment expressed in textual data by applying lexicon-based techniques, machine learning, and natural language processing.

Operational definition: utilized in the SWOTIFY framework to appraise the viewpoints and assessments of fellow officers in the Central Student Government.

Stop Words

Technical definition: a group of frequently occurring words that are frequently omitted or filtered out during the processing of textual data. These words, which are commonly used in a variety of texts but have little intrinsic sentiment value, include conjunctions, articles, and prepositions.

Operational definition: eliminating stop words to enhance the precision of sentiment analysis on SWOTIFY by prioritizing significant and sentiment-conveying terms.

SWOTIFY

Technical definition: it is a specific evaluation framework and an initiative by the Central Student Government organization. It is derived from the SWOT analysis which utilizes the determination of Strengths, Weaknesses, Opportunities, and Threats related to the appraisal and assessment of fellow officers.

Operational definition: dependent variable of the sentiment analysis; a platform to be utilized for sentiment analysis.

Conceptual Framework

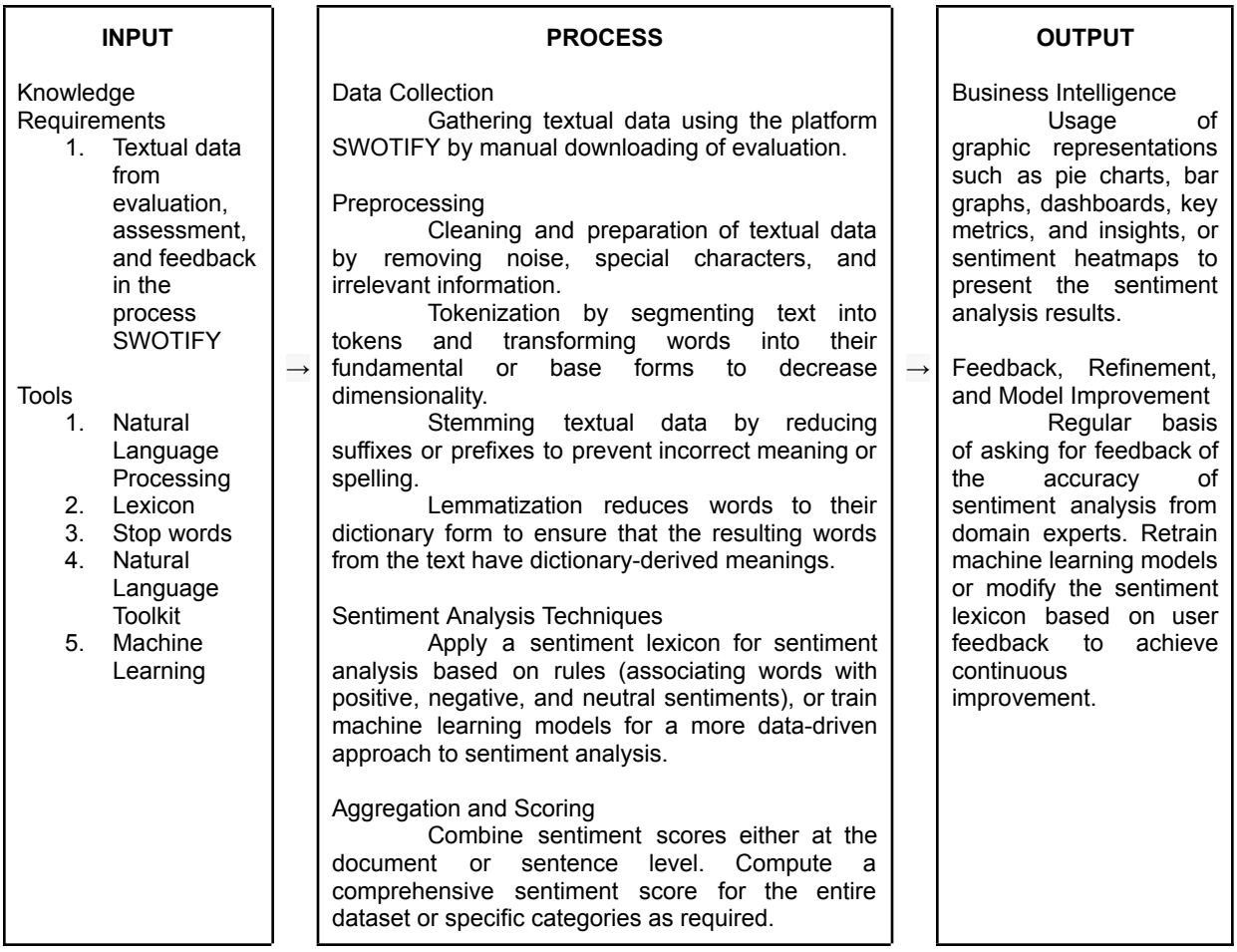

Figure 1: Conceptual Framework of SWOTIFY's Sentiment Analysis

The conceptual framework consists of three main components:

The conceptual framework governing sentiment analysis within the SWOTIFY context adheres to the **Input-Process-Output** (IPO) model, expounding the direction of information and resource flow. The study's proponents gathered all necessary elements carefully. The input phase involves the acquisition of textual data sourced from assessments, evaluations, and feedback, and utilizing sentiment analysis tools such as natural language processing (NLP), natural language toolkit (NLTK), selection of the right lexicon, and machine learning (ML). The process phase encompasses various methodologies, including data collection, preprocessing, sentiment analysis techniques, and aggregation and scoring. The output phase results in the visualization by business intelligence, affording an avenue for model refinement through feedback.

CHAPTER 2 – REVIEW OF RELATED LITERATURE

In this review of related literature and studies, the proponents will explore the existing research and studies in line with implementing sentiment analysis.

Theoretical Background

Implementing sentiment analysis is founded on the theoretical disciplines, forming a narrative of combined linguistics, computer science, data science, and artificial intelligence to create a powerful tool for identifying inherent textual insights.

From a lexicon-based approach, semantic orientation is assumed to be associated with certain numerical values to provide polarity to the emotions. This polarity is either generated via the syntactic patterns of the words in a given corpus and the seed list of words, or the synonyms and antonyms of words that can be searched in dictionaries. Based on the assumption that the semantic orientation of words can be assigned with numerical values, most of the sentiment analysis models relied on these values in creating a statistical calculation that aided in generating lexicons such as WordNet and WordStat.These lexicons aided in enabling sentiment analysis for general implementation.

However, Liu (2010) pointed out that the lexicon-based approach entails disadvantages of identifying opinions from the textual data when having domain-specific polarities. In his example, “quiet” might be deemed negative on the phrase “the speaker is quiet”, while it might be deemed positive on the phrase “the car engine runs quiet”. Another limitation of the lexicon-based approach is the language dependency that is apparent in a lexicon. Since most lexicons are written in English, it is very difficult to cover most of the words that are in the English vocabulary. Also, some words can have different polarities even in the same domain. Thus, finding the domain-dependent polarity becomes tricky and might not be sufficient. While

there have been a lot of possible ways to overcome this limitation, it remains a challenge to most semantic analysis problems since constraint arises on different aspects of natural language processing, such as language and vocabulary evolution, sentiment intensity and consistency, and conflicting conjunction and negation of word polarities.

From a machine learning (ML) approach, semantic analysis is mainly a classification problem. Text classification in ML methods focuses on finding the right features, appropriate weights for the features, selection of methods for each feature, and choosing the right ML model. These algorithms discern patterns and optimize predictions as they learn from the labeled data. As more data feeds into the system, the accuracy of the sentiment detection heightens. One of the most implemented algorithms in sentiment analysis is the Bayesian algorithm, which uses probability theory to predict the sentiment. It offers a simple yet effective semantic approach.

In theory, the use of both approaches will greatly improve the outcome of a sentiment analysis project by involving techniques in both approaches. Referring to the study of Meng in 2012, the proponents will use the lexicon-based approach to identify and label their inputted corpus from SWOTIFY to enable a supervised classification machine learning problem. The researchers may also use machine learning to execute unsupervised learning via clustering to generate the clusters of sentiments that are found in the dataset, with the end goal of converting the clusters to labels, which they can use as the label of each corpi from SWOTIFY.

Related Literatures

A Review of Techniques for Extracting Opinions from Textual Data

Sentiment Analysis is an essential tool for most fields, mainly in the business and education fields. It can help them to know the different points of view based on people's opinions in the form of textual data. Feedback from

people can help them to improve their product or performance. As pointed out by Hung and Alias (2023), challenges for any text-based classification include dealing with text context, sarcasm, and ambiguity, which can affect how a word is interpreted in a given text. Depending on the word pairing or context of the sentence, certain words can have multiple meanings. The proponents of this study aimed to use natural language processing tools to express the sentiment found in textual data in the SWOTIFY. By using such a tool, CSG officers can know each other's best performance in terms of the classification metrics set for this study when an event or project occurs.

Using Machine Learning and Deep Learning to Understand Student Perspectives

The literature conducted by Yang (2021) pointed out that student feedback is typically structured to include open-ended questions. Students are encouraged to share their perspectives on various aspects of teaching. It is critical to know students' opinions and it is a common practice when it comes to decision making. Sentiment analysis and data mining results can be misled into incorrect classifications by satirical statements. By incorporating machine learning and deep learning, proponents will be able to preprocessed performance evaluations from the SWOTIFY platform.

The Impact of Text Pre-Processing on Sentiment Analysis Accuracy

Pre-processing components are required for sentiment analysis in order to structure the text and extract features that can later be exploited by machine learning algorithms and text mining methods. Pre-processing is used to separate a set of characters from a text stream into classes. One of the most important aspects of text pre-processing is tokenizing a stream of characters into a sequence of word-like elements (Palomino and Aider, 2022). Text pre-processing can improve the accuracy of sentiment analysis. It removes noise such as punctuation, stop words (common words such as "the" or "a"), and typos, leaving only the core meaning. This cleaned-up data

makes sentiment analysis algorithms more easily interpretable, resulting in more accurate classifications of positive, negative, or neutral sentiment. Furthermore, pre-processing techniques such as stemming (reducing words to their root form) and lemmatization (removing grammatical endings) can identify sentiment-carrying words even when they appear in multiple forms, improving accuracy even further.

Analyzing Officer Sentiment in Unstructured Feedback on SWOTIFY platform

According to Faizi (2023) students' responses are typically in the form of multiple-choice and closed-ended or open-ended questions. Analyzing this feedback allows instructors to better understand the difficulties and performance of each student. Open-ended feedback requires NLP to analyze because the data collected is qualitative and in textual form. The importance of analyzing students' emotions in learning has recently been discussed in some research to help teachers understand the overall students' feelings during the learning process. The researchers will administer an unstructured open-ended questionnaire to the student government officers via the SWOTIFY platform. Through graphic representation, textual unstructured feedback will assist proponents in identifying textual data from evaluation, assessment, and feedback.

Related Studies

A Review of Sentiment Analysis Methodologies, Practices, and Applications

Numerous approaches and techniques are available for conducting a sentiment analysis. Mehta and Pandya (2020) discussed the two major approaches for conducting sentiment analysis, the (1) lexicon-based approach and the (2) machine learning approach. The lexicon-based

approach involves extracting words from a predefined set of lexicons and calculating their scores. For instance, words with positive, negative, or neutral emotions are assigned respective scores. The machine learning approach takes place by involuntary classification, where trained data and test data, the two sets of documents, are used for classification. Afterward, the text is analyzed for its features and divided into two supervised and unsupervised categories.

Systematic Reviews in Sentiment Analysis: A Tertiary Study

Ligthart and Tekinerdogan (2021) discussed various approaches in sentiment analysis. They categorized approaches into three types: knowledge-based techniques, statistical approaches (including machine learning and deep learning), and hybrid techniques that combine knowledge-based and statistical methods. Among the machine learning approaches, supervised learning is the most popular for sentiment analysis as it allows for training and testing of data. The lexicon-based approach, which scans documents for words expressing positive or negative feelings, is a traditional method that does not require training data. Sentiment analysis is widely used in different domains, with social media being the most popular. The study highlights the need for different techniques in different domains.

Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study

According to Kastrati, et. al. (2021), the main goal of a systematic mapping review (SMR) is to provide an overview of the body of knowledge and research area and to determine the number of publications and the types of research and results available. In contrast to classic systematic literature reviews (SLRs), which focus on identifying best practice based on empirical evidence, SMRs focus on establishing the nature of the evidence. It is also worth mentioning that from a methodological point of view, SLR is characterized by narrow and specific research questions, and studies are

reviewed for these aspects. The SMR, on the other hand, addresses a wide range of research questions, and studies are not evaluated based on qualitative information.

Synthesis

Academic sources illuminate the diverse approaches employed in sentiment analysis, highlighting the crucial role of text preprocessing in achieving accurate results. Moreover, they emphasize how sentiment analysis can foster a deeper understanding of the communication dynamics between Cavite State University - Imus (CvSU-Imus) Central Student Government (CSG) officers and students. For CSG officers, sentiment analysis on the SWOTIFY platform can facilitate self-assessment and evaluation of their colleagues, while also providing valuable insights into student concerns, opinions, and feedback. Existing research and literature consistently demonstrate the efficacy of sentiment analysis in organizational improvement.

Ultimately, sentiment analysis serves as a bridge, strengthening the connection between the CSG and students. The knowledge gleaned from the discussed studies and literature will be instrumental in conducting this research and ensuring the successful implementation of this research. This research, in turn, aims to enhance the overall evaluation and feedback experience for both CSG officers and students.

CHAPTER 3 – METHODOLOGY

This chapter describes the design adopted by this research to achieve the aims and objectives stated in the statement of the problem. The methodologies used in the study are the following: The materials used, research method and design, lists of all the instruments used in the study, the procedure, how the data was analyzed, sampling technique, the details of the participants in the study, its sampling procedure, and system software design.

Materials Used

The researchers will utilize several combinations of essential materials to develop the software application. The main tools to be used are Python 3.7 or 3.11 and JupyterLab. To initiate the development, the researchers will employ Python 3.7 or 3.11 as the primary programming language for building sentiment analysis, along with Jupyter Notebook. This tool will allow users to create and share live code, visualizations, and comments, and it will be used by the researchers to build and present sentiment analysis models.

For sentiment analysis, the researchers will use the Natural Language Toolkit (NLTK), a Python library for natural language processing, which includes various modules such as matplotlib, scikit-learn, and TextBlob. Matplotlib will be utilized to create visualizations such as lines, bars, scatter, pie charts, dashboards, and plots with unstructured coordinates. Scikit-learn will be used to preprocess data and extract features, while TextBlob, a lexicon-based approach, will categorize text as positive, negative, or neutral.

During the training and evaluation of the sentiment analysis model, the software application will utilize the minimum required specifications of the hardware. It will run on a Windows 10 or 11 operating system with at least a 500 GB SSD, 4GB RAM memory, and an 8GB display memory. The application must also be compatible with the latest web browsers, such as Edge, Google Chrome, and Mozilla Firefox. As this web application is

expected to be deployed online, an internet connection will be required for seamless connectivity to the application.

Research Method

The proponents will employ a mixed-method research design to develop sentiment analysis on the SWOTIFY platform for the Central Student Government at Cavite State University - Imus Campus. This approach combines qualitative and experimental research elements to go beyond mere data gathering and analysis.

The qualitative experiment provides a method for assessing CSG officers' performance and this will reveal the officers' different points of view. The experimental approach employs an established setting to statistically quantify the prevalence and intensity of identified sentiments. CSG officers may be required to interact with specific SWOTIFY platform features to provide qualitative feedback by using an unstructured open-ended format. (S. Robinson and A. L. Mendelson, 2012). This method will improve the results' accuracy. This combined strategy will provide the CSG with a better understanding of the perspectives of its members.

Research Instruments

The researchers will collect data for this study using a variety of research instruments. They must prioritize the study's components, such as source codes, documentation, and the software application itself, and organize them accordingly.

The Central Student Government will be the primary respondent of the study, they will evaluate their fellow officers, and in the data collection process, the study's proponents will conduct an unstructured open-ended questionnaire, which will be distributed to the Student Government officers via the SWOTIFY platform itself. Scholz and Zuell (2015) defined open-ended

questions as a research instrument consisting of “*How*,” “*What*,” or “*Why*“ questions and it encourages the officers to express their thoughts in their own words.

Research Design

The researchers will implement a structured design for this study by integrating quantitative and qualitative research designs. The study will attempt to identify subtle details in sentiment expression, thereby enhancing the comprehension of the study variable. As part of the qualitative approach in this study, sentiments, and other metrics aligned with this study will be used to provide a standardized and scalable evaluation. This combined approach will ensure a more accurate and detailed understanding of the sentiments of the variables in this study, capturing the emotional complexities inherent in the textual data.

Participants of the Study

The participant pool for this study encompasses individuals directly involved in the event or project, members of the Central Student Government Executive Board at Cavite State University Imus Campus, student academic ambassadors representing various academic programs alongside their respective advisers, and members serving on relevant committees of the organization.

Sampling Technique

This research will employ purposive sampling to acquire data directly relevant to the study's objectives. Participants will be specifically chosen based on their engagement with SWOTIFY, ensuring that the evaluations analyzed reflect the platform's intended use within the Central Student

Government at Cavite State University - Imus Campus. This focused approach will narrow the sample to officers who have actively submitted and received performance evaluations through SWOTIFY, guaranteeing the data's alignment with the research questions regarding sentiment analysis on the platform. This targeted selection strategy will maximize the study's internal validity and strengthen the connection between findings and the specific context of SWOTIFY within the student government organization.

Methodology Process

Data Collection

Data collection or gathering serves as the initial phase in the methodology process of sentiment analysis. By understanding the importance of gathering important information, researchers meticulously emphasized the collection of textual data. Textual data forms are the primary requirement for the data collection process. The proponents utilized SWOTIFY as a means of gathering a representative dataset. Data collection entails the strategic extraction by manual downloading from the evaluation forms to a spreadsheet or a comma-separated value (CSV) file. This process ensures the inclusion of all types of content in achieving a representative dataset.

Preprocessing

Preprocessing entails improving the quality of the gathered representative dataset harvested from SWOTIFY to enhance the quality and efficiency of sentiment analysis. The preprocessing phase plays a vital role in preparing the dataset for thorough sentiment analysis. The proponents optimized the following preprocessing aspects to ensure that the following algorithms can effectively recognize and interpret the variations of the language used.

The initial process of preprocessing involves cleaning and preparation of the dataset. This process ensures the quality and suitability of the dataset

for later analysis by involving steps in managing the different aspects of data. To preserve data integrity, the researchers will handle any missing or null values that are located through imputation or removal. Outliers are dealt with to avoid bias in later analyses, and deduplication techniques are used to eliminate duplicate entries. Normalization and standardization are used for numerical values, if necessary, to establish a consistent scale.

The next fundamental phase of preprocessing that the proponents will be using in sentiment analysis is tokenization. The main objective of tokenization is to convert raw text into a format that algorithms can easily process and analyze. This segmentation enables a more detailed comprehension of the linguistic structure and patterns present in the text. Tokenization involves dividing a text body, like a sentence or document, into distinct units known as tokens. Essentially, the words or subword units forming the text are tokens.

The stemming process occurs after the proponents used tokenization in the preprocessing stage. This methodology is a text normalization technique used in natural language processing, it reduces affixes or last characters from a word that often leads to incorrect meaning or spelling. The stemming process ensures that the representative dataset is simplified, and it helps the algorithm of the sentiment analysis to be efficient.

The last preprocessing phase that the proponents will be using is lemmatization. It is also a normalization technique but it reduces the words into their dictionary form to safeguard the outcome of normalized words to have derivation from the dictionary. The researchers will extract the core meaning of the word from the dataset to enable a more precise analysis.

Sentiment Analysis Techniques

Sentiment analysis is a set of computational methods that are used to determine the emotion that is perceived in textual data, either positive, negative, or neutral emotion. This methodology's main objective is to utilize sentiment analysis techniques to classify sentiments within preprocessed

textual data. This includes the actual grading of word tokens from the preprocessed textual data to a standardized sentiment scoring.

After aligning the preprocessed data to the standardized sentiment scoring, the proponents opted to test and use selected models and algorithms to capture the inherent emotion and opinion in the textual data by subjecting the textual data set to a series of model training and evaluation. This will involve the initial model training and validation of each of the models, and the tuning of each algorithm in a bid to improve their performance based on the initial findings of the model training.

Aggregation and Scoring

The concluding phase of the methodology process involves aggregation and scoring of the sentiment analysis. This involves the combination of the sentiment scores at either sentence-level or document-level basis for every algorithm implemented in the model training. The computation of the comprehensive sentiment score is done by comparing the actual sentiment score of the testing dataset and the trained model's predicted sentiment score. Overall, the proponents will be deciding on the model with the best score as the main model solution for this study.

REFERENCES

Liu, B. (2010). Sentiment Analysis and Subjectivity, Invited Chapter for the Handbook of Natural Language Processing, Second Edition.

Hung, L. P., & Alias, S. (2023). Beyond Sentiment Analysis: A Review of Recent Trends in Text Based Sentiment Analysis and Emotion Detection. Journal of Advanced Computational Intelligence and Intelligent Informatics, 27(1), 84–95. https://doi.org/10.20965/jaciii.2023.p0084

Kastrati, Z., Dalipi, F., Imran, A. S., Pireva Nuci, K., & Wani, M. A. (2021). Sentiment Analysis of Students' Feedback with NLP and Deep Learning: A Systematic Mapping Study. Applied Sciences, 11(9), 3986. https://doi.org/10.3390/app11093986

Ligthart, A., Catal, C., & Tekinerdogan, B. (2021). Systematic reviews in sentiment analysis: a tertiary study. Artificial Intelligence Review. https://doi.org/10.1007/s10462-021-09973-3

Mehta, P., & Pandya, Dr. S. (2020, February). A review on sentiment analysis methodologies, practices and applications. A Review On Sentiment Analysis Methodologies, Practices And Applications. https://www.researchgate.net/profile/Pooja-Mehta-26/publication/344487215_A_Review_On_Sentiment_Analysis_Methodologies_Practices_And_Applications/links/5f7bfb2992851c14bcb16528/A-Review-On-Sentiment-Analysis-Methodologies-Practices-And-Applications.pdf

Meng, Yanyan, "SENTIMENT ANALYSIS: A STUDY ON PRODUCT FEATURES" (2012). Dissertations, Theses, and Student Research from the College of Business. 28. https://digitalcommons.unl.edu/businessdiss/28

Palomino, M. A., & Aider, F. (2022). Evaluating the Effectiveness of Text Pre-Processing in Sentiment Analysis. Applied Sciences, 12(17), 8765. https://doi.org/10.3390/app12178765

Princeton University "About WordNet." WordNet. Princeton University. 2010. Retrieved from https://wordnet.princeton.edu/

Robinson, S., & Mendelson, A. L. (2012). A Qualitative Experiment. Journal of Mixed Methods Research, 6(4), 332–347. https://doi.org/10.1177/1558689812444789

Yang, R. (2021, February 3). Machine Learning and Deep Learning for Sentiment Analysis Over Students' Reviews: An Overview Study. Preprints.org. https://doi.org/10.20944/preprints202102.0108.v1

Zuell, C., & Scholz, E. (2015). Who is willing to answer open-ended questions on the meaning of left and right? Bulletin of Sociological Methodology/Bulletin de Méthodologie Sociologique, 127(1), 26–42. https://doi.org/10.1177/0759106315582199