Plagiarism Checker X - Report

Originality Assessment

25%

Overall Similarity

Date: Aug 8, 2024
Matches: 1293 / 5129 words
Sources: 9

Remarks: High similarity detected, please make the necessary changes to improve the writing.

Verify Report:
Scan this QR Code

International Journal of Information Technology and Computer Science Applications (IJITCSA) p-ISSN: 2964-3139 e-ISSN: 2985-5330 Vol. 2, No. 1, page 12 - 22 Submitted 24/11/2023; Accepted 31/12/2023; Published 10/01/2024 ©2024 Pratama et al. Jejaring Penelitian dan Pengabdian Masyarakat (JPPM) Multinomial Naive Bayes Algorithm for Indonesian language Sentiment Classification Related to Jakarta International Stadium (JIS) Daffa Rizki Surya Pratama1, Tb Ai Munandar2 and Khairunnisa Fadhilla Ramdhania3 Informatics Dept. Faculty of Computer Science, Universitas Bhayangkara Jakarta Raya1,2,3 Email : daffa.rizki.surya.pratama19@mhs.ubharajaya.ac.id, tb.aimunandar@dsn.ubharajaya.ac.id,khairunnisa.fadhilla@dsn.ubharajaya.ac.id Corresponding Autor: daffa.rizki.surya.pratama19@mhs.ubharajaya.ac.id 1 Abstract The research focuses on analysing public evaluations, particularly those on Google Maps, about the Jakarta International Stadium (JIS). The study aims to employ the multinomial Naive Bayes algorithm to ascertain the sentiment expressed in these reviews. The objective of this study was to employ the multinomial Naive Bayes method to analyse the reviews on Google Maps pertaining to the Jakarta International Stadium (JIS). The utilised data consists of 2971 public reviews on Google Maps specifically pertaining to Jakarta International Stadium (JIS). These reviews were acquired through web scraping using a data miner. The acquired data is next processed in the text preparation phase to generate a prepared dataset suitable for analysis. This preprocessing stage includes operations such as casefolding, stopword removal, tokenizing, and stemming. The study yielded an accuracy of 0.83, or 83%, when tested on 733 data points. Out of these, 292 positive data points were correctly anticipated, while 59 positive data points were incorrectly forecast. Additionally, 317 negative data points were correctly predicted, while 65 negative data points were incorrectly predicted. The conducted modelling is subsequently categorised using a novel dataset of 161 review data points, with the objective of discerning the sentiment expressed within the dataset. The analysis of the new dataset yielded 101 reviews with positive sentiment and 50 reviews with negative sentiment.

 Keywords— 1 review, sentiment analysis, Jakarta International Stadium (JIS), Multinomial Naive Bayes, Classification. 1. Introduction The Jakarta International Stadium (JIS), formerly known as the Bersih Human Wibawa Stadium (BMW), is situated in Papanggo Village, Tanjung Priok District, North Jakarta. It serves as a football stadium. 5 The inception of the JIS stadium was initiated in 2008; however, because of conflicts over land ownership, the construction of the stadium was delayed until 2019 and ultimately finished in 2022. 7 The JIS boasts opulent amenities and underwent extensive renovations, resulting in a spectator capacity of 82 thousand [1]. Since the inception of the construction of the Jakata International Stadium, the community has provided evaluations concerning its construction, thereby generating public sentiment regarding the adequacy of the current facilities and infrastructure, including parking lots, public transportation, substandard access roads, and limited entrance access [2]. Public sentiment disseminates via diverse electronic platforms, including Google Maps. Google Maps reviews have an impact on public perceptions. Due to the stadium's adherence to international standards, this emotion has a particular impact on the presence of JIS. 1 The objective of this study was to ascertain the sentiment of each review sentence that individuals from the general public, specifically Google Maps users, had submitted on the review page. In addition, it can serve as a basis for management to address the deficiencies present at that specific location. If this view is appropriately

 analysed, it can provide evaluative information for decision-making by JIS managers. Prior academics employed the Naive Bayes approach to conduct sentiment analysis in many scenarios, including the evaluation of sentiment towards the Indonesian Xing Fu Tang Beverage Store. By applying multinomial Naive Bayes to the Google Review data, we achieved an accuracy rate of 78% [3]. The researach conducted by [4] has the same approach with Bali tourism destinations in Google Maps. In his research, 8 the Naive Bayes algorithm was employed to analyse five tourist attractions in Bali and determine th e top five favourites. By employing the Naïve Bayes technique, we achieved an impressive accuracy rate of 94.64% for the Nusa Penida destination, establishing it as a highly recommended tourism destination.This research will focus on analysing public attitudes towards JIS by examining 1 reviews on Google Maps. How can multinomial Naive Bayes be applied to JIS sentiment analysis usin g Google Maps reviews?The objective of this project is to perform sentiment analysis on the review data pertaining to Jakarta International Stadium, which is stored in the comments/reviews section on Google Maps. 1 The multinomial Naive Bayes method will be employed to categorise the review data related to Jakarta International Stadium on Google Maps. 2. Related Works Multiple researchers have conducted 2 studies on sentiment analysis, including those conducted by [3], [5]–[7]. The researchers have effectively conducted sentiment analysis research employing diverse research methodologies. To determine individuals' perspectives on a particular item, sentiment analysis studies often rely on individuals' viewpoints regarding a well-known tourist attraction or a subject of discussion, as seen in the research conducted by [8]–[10]. Various methodologies can be employed to conduct sentiment analysis research, including machine learning, which enables the categorization of public opinion towards a certain entity. The sentiment classification technique employs machine learning algorithms to construct 2 models for sentiment analysis. The Naive Bayes method is frequently employed for sentiment classification due to its reputation for being straightforward and efficient in doing sentiment analysis, as evidenced by multiple researchers who have utilised it [11]–[16]. 3. Research Methodology This study employs a multi-stage approach to perform sentiment analysis on evaluations pertaining to Jakarta International Stadium (JIS) on Google Maps, as illustrated in Figure 1. Figure 1 illustrates the many stages involved in sentiment analysis. 1) Data collection Data

Collection refers to the process of gathering and recording information or data from various sources. Data collection is the preliminary phase conducted to acquire review data from the Google Maps review feature specifically related to the Jakarta International Stadium (JIS) for the purpose of doing sentiment analysis. The data

collection in this research employs web scraping methodologies, which involve the systematic gathering of data from many sources on the internet, including the utilisation of web browser APIs or applications. Web scraping often involves using an automated software to send queries or requests to a web server and then retrieving or parsing the desired data based on specific information [17]. The research use the Data Miner web browser plugin for web scraping. 2) Data preprocessing Data preprocessing refers to the steps used to clean and transform raw data into a format that is suitable for analysis and modelling. Data preprocessing is a technique used to prepare data collections, transforming them from raw and unstructured formats into organised formats that are more easily manageable [18]. Data preparation activities have the objective of constructing the ultimate data set that will be utilised for modelling. The phases involved in data preprocessing include casefolding, stemming, tokenizing, and removing stopwords. 3) Term Frequency-Inverse Document Frequency (TF -IDF) TF-IDF is a method used to assign weights to words based on their importance in a document collection. There are various methods for word weighting, one of which is TF-IDF (Term Frequency-Inverse Document Frequency). TF-IDF is a technique used to convert textual data into numerical data. TF-IDF is a statistical metric employed to assess the significance of a word inside a document. Term frequency (TF) refers to the frequency of a word's occurrence in a document and its level of significance. The term "DF" refers to the frequency of documents that contain a specific word, indicating the word's level of commonality. The IDF (Inverse Document Frequency) is the reciprocal of the DF (Document Frequency) value. The TF-IDF word weighting computation is obtained by multiplying the term frequency (TF) with the inverse document frequency (IDF). The equation employed is analogous to equation (1) [19]. (1) The equation (1) can be expressed as $w(i,j) = tf(i,j) \times \log(N/df(i))$, where $w(i,j)$ is the weight of term i in document j, $tf(i,j)$ is the term frequency of term i in document j, N. The variable $w(i,j)$ represents the weight of the word-i in the j-th text, $tf(i,j)$: The count of i-words in the j-th document, N represents the total number of documents, while $df_{((i))}$ represents the number of documents that contain the ith word. 4) Partitioning of Test Data with Training Data The goal of partitioning test data and training data is to segregate data sets for the purpose of modelling. The data is partitioned into two segments, with a proportion of 20% allocated for testing and 80% for training. The purpose of partitioning the data set in a 20:80 ratio is to enhance the fitness of the classification model by providing a substantial amount of training data during the modelling process. 5) Modelling of multinomial classification using Naive Bayes Multinomial Naive Bayes is a probabilistic learning model derived from Bayes theory. It is widely employed in Natural Language Processing (NLP) and operates on the principle of term frequency, which quantifies the occurrence of words in a document. This model provides an explanation for two aspects: the presence or absence of a word in a text, and the frequency of the word's occurrence in the document [20]. Multinomial Naive Bayes is a text classification model that utilises supervised learning. Therefore, it requires pre-labeled data before training can commence. The formulation of Multinomial Naive Bayes can be represented by equation (2) [21]. (2) The probability of class C appearing in document D, denoted as $P(C|D)$, is derived from equation (3), which calculates the total number of words in the document, represented as n. (3) Equation (4) calculates the probability of class C, denoted as $P(C)$, by dividing the total number of documents by the number of class C documents, represented as N_c.

 $P(w_i|c)$ represents the conditional probability of the ith word given in class C. count $(w_i|C)$ denotes the frequency of the ith word in class C.

count (C) refers to the total number of words in class C, whereas |V| represents the total number of unique words across all classes. (4) 6) Evaluation of the model This model evaluation is conducted to assess the efficacy of the utilised model. This evaluation involves multiple computation phases utilised in the assessment procedure, including accuracy, precision, recall, and Fscore. This model evaluation uses the confusion matrix technique to display the outcomes of the classification that has been conducted [22]. 7) Updated Test Data Testing new data with classification modelling is conducted in order to determine sentiment classification. This process is referred to as the new test data stage. 8) Application of Classification Models to Unseen Test Data The employed model is a sentiment analysis classifier utilising the multinomial Naive Bayes technique.VisualisationThe sentiment study conducted at the Jakarta International Stadium (JIS) will yield both positive and negative outcomes. These results will be presented visually through a bar chart, depicting the respective quantities of positive and negative sentiments. The wordcloud tool will also be used for data visualisation in order to identify frequently employed words. 4. Results and Discussion 1) Research data. The dataset utilised in this study consisted of 2971 reviews pertaining to the Jakarta International Stadium (JIS) extracted from Google Maps reviews. The process of gathering data was conducted through the utilisation of online scraping methods, facilitated by a web browser extension known as Data Miner. The acquired data consisted of user names and reviews. The data collection was conducted between April 2023 and May 2023. The data presented in Table 1 was collected from reviews on Google Maps. Table 1 displays the outcomes of the data collection process. user_id review in Bahasa _damarsty nungguin jis rampung lama banget, kasian stadion bola naru pernah dipake fun footballðŸ™ƒ tar kalo udah 100% semoga kaga ada drama pager roboh lagiðŸ˜€ Abdul Rohim Sangat bagus dan megah stadion nya Abdul Tatang (Tatang Doel) Stadion yg sangat bagus dan nyaman, serta bersih. Dan tertata rapi. Adam Juliawan stadion sekeren ini engga standar fifa masa iya? adityaariwibawaa Dilarang ngerokok tp yg nge band ngerokok di backstage wkwk ðŸ˜‚ maulanasari Bagus Banget & instagramable Menil tata sayangnya belum standar PSSI mim125 Luar biasa, megah dan bagus sekali, alhamdulillah saya beserta keluarga berkesempatan utk melihat pertandingan Barcelona U 18 VS Atletico Madrid U18, keren banget....terima kasih bapak Anies Baswedan. novitaardnt____ Sumpah keren bangeettt ðŸ˜ The acquired data will thereafter be manually labelled. Below is a selection of the data that was categorised in Table 2. Table 2 displays the data regarding the results of labelling user_id sentiment class review in Bahasa

 _damarsty Negatif nungguin jis rampung lama banget, kasian stadion bola naru pernah dipake fun footballðŸ™ƒ tar kalo udah 100% semoga kaga ada drama pager roboh lagiðŸ˜€ Abdul Rohim Positif Sangat bagus dan megah stadion nya Abdul Tatang (Tatang Doel) Positif Stadion yg sangat bagus dan nyaman, serta bersih. Dan tertata rapi. Adam Juliawan Negatif stadion sekeren ini engga standar fifa masa iya? adityaariwibawaa Negatif Dilarang ngerokok tp yg nge band ngerokok di backstage wkwk ðŸ˜, maulanasari Positif Bagus Banget & instagramable Menil tata Negatif sayangnya belum standar PSSI mim125 Positif Luar biasa, megah dan bagus sekali, alhamdulillah saya beserta keluarga berkesempatan utk melihat pertandingan Barcelona U 18 VS Atletico Madrid U18, keren banget....terima kasih bapak Anies Baswedan. novitaardnt____ Positif Sumpah keren bangeettt ðŸ˜ Upon completion of data collection and tagging, the data will undergo the data preparation stage. The data preprocessing stage consists of various steps, including case folding, 1 stopword removal, tokenizing, and stemming. The first step performed is casefolding, which involves converting all uppercase letters to lowercase and eliminating punctuation marks and emoticons from sentences. Table 3 displays the outcomes of casefolding. Table 3. Casefolding Results Data Review Before Case Folding Review After Case Folding udah ga penasaran udah pernah masukðŸ˜, udah ga penasaran udah pernah masuk AKSESNYA SUSAH! SAMPAH aksesnya susah sampah wadidaw panjang banget ges 4 jam wadidaw panjang banget ges   jam nungguin jis rampung lama banget, kasian stadion bola naru pernah dipake fun footballðŸ™ƒ tar kalo udah 100% semoga kaga ada drama pager roboh lagiðŸ˜€ nungguin jis rampung lama banget kasian stadion bola naru pernah dipake fun football tar kalo udah   semoga kaga ada drama pager roboh lagi coba kalau gak ada trek larinya keren itu coba kalau gak ada trek larinya keren itu The subsequent step following case folding is stopword removal, 1 with the objective of deleting words that lack significant significance in the phrase. Table 4 displays the outcomes of the stopword phase. Table 4 displays the data obtained from the analysis of stopwords. Review Before Stopword Review After Stopword udah ga penasaran udah pernah masuk udah ga penasaran udah pernah masuk aksesnya susah sampah aksesnya susah sampah wadidaw panjang banget ges   jam panjang banget jam nungguin jis rampung lama banget kasian stadion bola naru pernah dipake fun football tar kalo udah semoga kaga ada drama pager roboh lagi nungguin jis rampung lama banget kasian stadion bola naru pernah dipake fun football tar kalo udah semoga kaga drama pager roboh coba kalau gak ada trek larinya keren itu coba kalau gak trek larinya keren persija auto support pak anis periode persija auto support pak anis periode Following the completion of the stopwords step, the tokenizing stage is executed to divide the sentence into individual words, represented as tokens. The outcomes of tokenizing are displayed in Table 5. Table 5. Tokenizing Result Data Review Before Tokenizing Review After Tokenizing udah ga penasaran udah pernah masuk ['udah', 'tidak', 'penasaran', 'udah', 'pernah', 'masuk'] aksesnya susah sampah ['akses', 'susah', 'sampah'] wadidaw panjang banget ges   jam ['panjang', 'banget', 'jam'] nungguin jis rampung lama banget kasian stadion bola naru pernah dipake fun football tar kalo udah semoga kaga ada drama pager roboh lagi ['nungguin', 'jis', 'rampung', 'lama', 'banget', 'kasi', 'stadion', 'bola', 'naru', 'pernah', 'dipake', 'fun', 'football', 'tar', 'kalo', 'udah', 'moga', 'kaga', 'drama', 'pager', 'roboh'] coba kalau gak ada trek larinya keren itu ['coba', 'kalau', 'tidak', 'trek', 'lari', 'keren'] persija auto support pak anis periode ['persija', 'auto', 'support', 'pak', 'anis', 'periode']

1 MULTINOMIAL NAIVE BAYES ALGORITHM FOR INDONESIAN LANGUAGE

SENTIMENT CLASSIFICATION 17 Following the tokenizing stage, the data will undergo a stemming process to normalise words by removing all affixes. The outcomes of the stemming phase are displayed in Table 6. Table 6. Stemming Results Data Review Before Stemming Review After Stemming ['udah', 'tidak', 'penasaran', 'udah', 'pernah', 'masuk'] udah tidak penasaran udah pernah masuk ['akses', 'susah', 'sampah'] akses susah sampah ['panjang', 'banget', 'jam'] panjang banget jam ['nungguin', 'jis', 'rampung', 'lama', 'banget', 'kasi', 'stadion', 'bola', 'naru', 'pernah', 'dipake', 'fun', 'football', 'tar', 'kalo', 'udah', 'moga', 'kaga', 'drama', 'pager', 'roboh'] nungguin jis rampung lama banget kasi stadion bola naru pernah dipake fun football tar kalo udah moga kaga drama pager roboh ['coba', 'kalau', 'tidak', 'trek', 'lari', 'keren'] coba kalau gak trek lari keren ['persija', 'auto', 'support', 'anis', 'periode'] persija auto support anis periode 2) TF-IDF Word Weighting 4 Term Frequency-Inverse Document Frequency (TF-IDF) Word weighting is a procedure that converts textual data into numerical data by assigning weights to words. TF-IDF is a statistical metric employed to assess the significance of a word in a document following the preprocessing phase. Table 7 displays the outcomes of computing several data points for TF-IDF weighting. Table 7. TF-IDF calculation results Term TF - IDF D1 D2 D3 D4 D5 D6 D7 Udah 1,69018 3 0 0 0 0 0 0 Penasaran 0,84509 0 0 0 0 0 0 Pernah 0,84509 0 0 0 0 0 0 Masuk 0,84509 0 0 0 0 0 0 Akses 0 0,84509 0 0 0 0 0 Susah 0 0,84509 0 0 0 0 0 Sampah 0 0,84509 0 0 0 0 0 Panjang 0 0 0,84509 0 0 0 0 Banget 0 0 0,84509 0 0 0 0 Jam 0 0 0,84509 0 0 0 0 Keren 0 0 0 0,84509 0 0 0 Stadion 0 0 0 0,84509 0 0 0 Jis 0 0 0 0,54406 0 0 0,54406 Trek 0 0 0 0 0,84509 0 0 Lari 0 0 0 0 0,84509 0 0 Persija 0 0 0 0 0 0,54406 0,54406 Anies 0 0 0 0 0 0,84509 0 Penonton 0 0 0 0 0 0 0,84509 Musim 0 0 0 0 0 0 0,84509 3) The division of test data and training data is necessary. The data partitioning stage uses 20% of the vectorized data as test data and 80% as training data, resulting in a total of 3941 data terms. The division of test and training data is performed using the train_test_split module from the sklearn.model_selection library. The dataset is split into two parts, 2 with a total of 733 data points allocated for testing. 4) Classification Modelling Following the previous stage of dividing the data into test data and training data, a total of

733 test data will undergo sentiment classification modelling to assess the expected and actual results in the positive and negative classes. The outcomes of the modelling are depicted in Figure 2.

 Figure 2 displays the outcomes of the mutinomial Naive Bayes classification modelling. Following the implementation of multinomial Naïve Bayes classification modelling, a subsequent round of classification modelling 2 was conducted using the RandomForest method in order to assess and compare the efficacy of the two algorithms. The outcomes of the classification modelling using the RandomForest algorithm are displayed in Figure 3. Figure 3. Random Forest Classification Modelling Results. 5) Evaluation of the Model Following the completion of the modelling stage, we proceed to the model evaluation stage, where we employ a confusion matrix to assess the accuracy of the constructed model. The confusion matrix will provide information on several key metrics, including accuracy, recall, precision, and F1-Score. The evaluation results, obtained from the confusion matrix, are presented in Table 8. Table 8. Summary of Confusion Matrix Results Multinomial Naive Bayes RandomForest Accuracy (%) 83 % 81 % Precision (%) 83 % 86 %

 Recall (%) 81 % 73 % F1-Score (%) 82 % 77 % 6) Discussion The findings from the classification modelling conducted using two algorithms indicate that the mult inomial Naive Bayes algorithm outperforms the RandomForest approach in terms of accuracy, specifically achieving a value of 83%. 1 The multinomial Naive Bayes method accurately predicts 733 test data points, consisting of 292 positive values and 317 negative values. The precision value for both positive and negative classes is 83%. The recall achieved for negative data was 84%, and for positive data, it was 83%. Similarly, the F1-Score acquired for negative data was 84% and for positive data was

83%.Subsequently, 2 the multinomial Naive Bayes model was employed to conduct sentiment analysis on additional data. This entailed incorporating 161 new data points that had successfully undergone preprocessing and TF-IDF word weighting. The data will subsequently undergo modelling, yielding sentiment analysis results that encompass both 2 positive and negative sentiments present in the text. Table 9 displays the recently uploaded test data. Table 9. New Test Data User_id Review in Bahasa Valen Radja Naitili Sebagai anak bangsa saatnya kita saling mendukung. Kita harus bersatu untuk kemajuan Indonesia, JIS Emang mantap Lukman Projayan jis adalah produk anak bangsa alias lokal pride .. tanpa tenaga kuli asing... mantap man STADION TERBAIK BELUM ADA DI NEGARA2 ASIA TENGGARA BAHKAN MASUK 10 BESAR STADIIN TERMEWAH DIDUNIA INI SUATU KEBANGGAAN ANAK BANGSA Agung Wahyu JIS adalah Stadion yg menelan biaya lebih dari 4 Trilyun lebih... Wan Adzro Salam dari Kuala Lumpur.. Stadium dah bagus. Harap penyokong indonesia berdisiplin dan open minded. Azharu Rosyidin Pernah berkunjung ke JIS dan ku akui memang keren, cuma saran kepada Jakpro mohon diperhijau lagi area sekitarnya, supaya makin syahdu. Giovanny Ini baru stadion berkelas amiin ya Allah SKY Figurefun Stadium jis bagus cantik gilerr Sipri m Satu-satunya stadion di Indonesia berkelas Dunia, hebat maha karya anak bangsa Indonesia. Kreeeennnn Pro maju Bagus dan keren banget gan.semoga stadion jis kelak bisa menjadi tuan rumah piala dunia.amin Subsequently, the data underwent sentiment classification modelling utilising 1 the multinomial Naive Bayes modelling approach. The outcomes derived from the examination of fresh data are displayed in Table 10. Table 10. New Data Test Results. User_id Review in Bahasa Sentiment Class Valen Radja Naitili Sebagai anak bangsa saatnya kita saling mendukung. Kita harus bersatu untuk kemajuan Indonesia, JIS Emang mantap Positif Lukman Projayan jis adalah produk anak bangsa alias lokal pride .. tanpa tenaga kuli asing... mantap Positif man STADION TERBAIK BELUM ADA DI NEGARA2 ASIA TENGGARA BAHKAN MASUK 10 BESAR STADIIN TERMEWAH DIDUNIA INI SUATU KEBANGGAAN ANAK BANGSA Positif Agung Wahyu JIS adalah Stadion yg menelan biaya lebih dari 4 Trilyun lebih...

Negatif Wan Adzro Salam dari Kuala Lumpur.. Stadium dah bagus. Harap penyokong indonesia berdisiplin dan open minded. Positif Azharu Rosyidin Pernah berkunjung ke JIS dan ku akui memang keren, cuma saran kepada Jakpro mohon diperhijau lagi area sekitarnya, supaya makin syahdu. Positif Giovanny Ini baru stadion berkelas amiin ya Allah Positif SKY Figurefun Stadium jis bagus cantik gilerr Positif Sipri m Satu-satunya stadion di Indonesia berkelas Dunia, hebat maha karya Psotif

 anak bangsa Indonesia. Kreeeennnn Pro maju Bagus dan keren banget gan.semoga stadion jis kelak bisa menjadi tuan rumah piala dunia.amin Positif The analysis of the new test data revealed that the modelling process successfully classified 110 instances as positive feelings and 51 instances as negative sentiments. The data will be represented graphically in the form of a diagram, as depicted in Figure 4. Figure 4. Visualisation of Modelling Results Test Data. The upcoming visualisation will present a wordcloud showcasing the commonly occurring words in a set of 161 new test data points. Figure 5 displays the outcomes of a visual representation without words, specifically focusing on good emotions. There are many terms that are displayed in larger size, indicating that they are [2] the most commonly used words. Commonly used words include 'jis','son', 'country','stadium', an'country', 'stadium', and 'majestic'. The term 'jis' frequently occurs in all favourable reviews. The term 'jis' is commonly employed by individuals as an acronym for Jakarta International Stadium, or JIS, for the purpose of facilitating recall. The terms 'child' and 'country' frequently appear in reference [1] to the Jakarta International Facility (JIS), which is a football facility constructed to international standards by the youth of the nation. These words are frequently employed to express satisfaction and admiration for the development of the Jakarta International Stadium (JIS). The term'majestic' is frequently employed to express admiration for a grandiose football stadium. Figure 5. Positive Sentiment Wordcloud. Figure 6 displays the frequently used terms associated with unfavourable attitudes in assessments [9] of the Jakarta International Stadium (JIS) on Google Maps. The terms'stadium', 'jis', 'world cup', and'standard' are frequently employed. The term 'world cup' is frequently employed to express popular disillusionment since [1] the Jakarta International Stadium (JIS) was unable to host international football competitions owing to encountered issues. The term 'standard' is frequently employed due to issues arising from the evaluation conducted by the Indonesian Football Federation, also known as PSSI, which determines that the Jakarta International Stadium does not meet the FIFA standard for a football stadium.

 Figure 6. Negative Sentiment Wordcloud 5. Conclusion The research findings indicate that reviews on Google Maps exhibit a spectrum of attitudes, encompassing both positive and negative sentiments. The collected review data will be utilised for sentiment analysis modelling through machine learning. A total of 2971 data points have been acquired and manually assigned sentiment labels. [2] The sentiment analysis of Google Maps reviews was successfully conducted by employing machine learning modelling and the multinomial naive Bayes algorithm. A total of 2971 data points were collected throughout the data collection procedure. These data points were subsequently subjected to modelling during the data preprocessing stage. Subsequently, the TF-IDF approach is employed to assign weights to [6] the words in the dataset. Following this, the dataset is split into training data and testing data, with a ratio of 80:20. The application [2] of multinomial naive Bayes modelling resulted in an accuracy rate of 83%. The RandomForest method achieved an accuracy rate of 81% after the modelling process. Based on the accuracy results achieved by both algorithms, the modelling was deemed successful. However, [1] the multinomial Naive Bayes algorithm has shown higher effectiveness compared to the RandomForest approach. The developed model was subsequently applied to sentiment analysis on a novel dataset consisting of 161 unlabeled data points. [2] The multinomial naive Bayes model accurately classified

sentiment in the review sentences, with 110 reviews classified as positive and 51 reviews classified as negative. BIBLIOGRAPHY

[1] . “Stadion,” Wikipedia. Dec. 17, 2022. [2] 1 . “Profil JIS: Stadion Mewah Jakarta yang disebut PSSI tak Standar FIFA,” CNN Indonesia, Sep. 12, 2022. [3] . A. J. Fadhillah, “Analisis Sentimen Terhadap Gerai Minuman Xing Fu Tang Indonesia Menggunakan Data Ulasan Google,” 2022. [4] . A. Erfina and D. S. Utami, “Analisis Sentimen Objek Wisata Bali Di Google Maps Menggunakan Algoritma Naive Bayes,” 2022. [5] . M. Permana, S. Widiastuti, and S. Saepudin, “ANALISIS SENTIMEN PENGGUNAAN APLIKASI VIDEO CONFERENCE PADA ULASAN GOOGLE PLAY STORE MENGGUNAKAN METODE NBC (NAIVE BAYES CLASSIFIER),” JURSISTEKNI (Jurnal Sistem Informasi dan Teknologi Informasi), vol. 5, pp. 178– 191, Jan. 2023. [6] . D. R. Lazuardi, T. A. Munandar, H. Harsiti, Z. Mutaqin, and R. N. Hays, “Sentiment analysis of public opinions on the welfare of honorary educators using Naive Bayes,” in IOP Conference Series: Materials Science and Engineering, Institute of Physics Publishing, May 2020. doi: 10.1088/1757-899X/830/3/032018.

[7] . A. P. Giovani, A. Ardiansyah, T. Haryanti, L. Kurniawati, and W. Gata, “ANALISIS SENTIMEN APLIKASI RUANG GURU DI TWITTER MENGGUNAKAN ALGORITMA KLASIFIKASI,” Jurnal Teknoinfo, vol. 14, no. 2, p. 115, Jul. 2020, doi: 10.33365/jti.v14i2.679. [8] . R. Alamsyah, T. Ai Munandar, F. Nidaul Khasanah, and S. Setiawati, “Sentiment Analysis Destinasi Wisata Kabupaten Bekasi Berdasarkan Opini Masyarakat Menggunakan Naive bayes,” Journal DINDA, vol. 2, no. 2, pp. 64–74, 2022, [Online]. Available: http://journal.ittelkom-pwt.ac.id/index.php/dinda [9] . M. A. Fauzi and S. Adinugroho, “Analisis Sentimen Pariwisata di Kota Malang Menggunakan Metode Naive Bayes dan Seleksi Fitur Query Expansion Ranking Text clustering View project Twitter Sentiment Analysis View project,” 2018. [Online]. Available: https://www.researchgate.net/publication/322959527 [10] . J. Winahyu and I. Suharjo,

"Aplikasi Web Analisis Sentimen Dengan Algoritma Multinomial Naïve Bayes," Kumpulan Artikel Mahasiswa Pendidikan Teknik Informatika (KARMAPATI), vol. 10, no. 2, 2021. [11] . N. S. Wardani, A. Prahutama, and P. Kartikasari, "ANALISIS SENTIMEN PEMINDAHAN IBU KOTA NEGARA DENGAN KLASIFIKASI NAÏVE BAYES UNTUK MODEL BERNOULLI DAN MULTINOMIAL," Jurnal GAUSSIAN, vol. 9, pp. 237–246, 2020, [Online]. Available: https://ejournal3.undip.ac.id/index.php/gaussian/ [12] . H. Herwanto, N. L. Chusna, and M. S. Arif, "Klasifikasi SMS Spam Berbahasa Indonesia Menggunakan Algoritma Multinomial Naïve Bayes," JURNAL MEDIA INFORMATIKA BUDIDARMA, vol. 5, no. 4, p. 1316, Oct. 2021, doi: 10.30865/mib.v5i4.3119. [13] . G. A. Buntoro, "Analisis Sentimen Calon Gubernur DKI Jakarta 2017 Di Twitter," 2017. [Online]. Available: https://t.co/jrvaMsgBdH [14] . W. Yulita et al., "Analisis Sentimen Terhadap Opini Masyarakat Tentang Vaksin Covid-19 Menggunakan Algoritma Naïve Bayes Classifier," JDMSI, vol. 2, no. 2, pp. 1–9, 2021. [15] . E. M. Sipayung, H. Maharani, and I. Zefanya, "PERANCANGAN SISTEM ANALISIS SENTIMEN KOMENTAR PELANGGAN MENGGUNAKAN METODE NAIVE BAYES CLASSIFIER," 2016. [Online]. Available: http://ejournal.unsri.ac.id/index.php/jsi/index [16] . B. Mas Pintoko and K. Muslim, "Analisis Sentimen Jasa Transportasi Online pada Twitter Menggunakan Metode Naïve Bayes Classifier," Bandung, Dec. 2018. Accessed: Jun. 20, 2023. [Online]. Available: https://openlibrarypublications.telkomuniversity.ac.id/index.php/engineering/article/view/7447/7336 [17] . R. Mitchell, Web Scraping with Python: Collecting More Data from the Modern Web, 2nd ed. United States Of America: O'Rielly Media, 2018. [18] . C. S. Sriyano and E. B. Setiawan, "Pendeteksian Berita Hoax Menggunakan Naive Bayes Multinomial Pada Twitter dengan Fitur Pembobotan TF-IDF," 2021. [19] . V. Amrizal, "PENERAPAN METODE TERM FREQUENCY INVERSE DOCUMENT FREQUENCY (TFIDF) DAN COSINE SIMILARITY PADA SISTEM TEMU KEMBALI INFORMASI UNTUK MENGETAHUI SYARAH HADITS BERBASIS WEB (STUDI KASUS: HADITS SHAHIH BUKHARIMUSLIM)," JURNAL TEKNIK INFORMATIKA, vol. 11, no. 2, pp. 149–164, Nov. 2018, doi: 10.15408/jti.v11i2.8623. [20] . G. Singh, B. Kumar, and A. Tyagi, "Comparison between Multinomial and Bernoulli Naïve Bayes for Text Classification," International Conference on Automation, Computational and Technology Management (ICACTM), Apr. 2019. [21] . C. D. Manning, Prabhakar. Raghavan, and Hinrich. Schütze, Introduction to information retrieval. Cambridge University Press, 2008. [22] . N. L. P. M. Putu, Ahmad Zuli Amrullah, and Ismarmiaty, "Analisis Sentimen dan Pemodelan Topik Pariwisata Lombok Menggunakan Algoritma Naive Bayes dan Latent Dirichlet Allocation," Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi), vol. 5, no. 1, pp. 123–131, Feb. 2021, doi: 10.29207/resti.v5i1.2587.

Sources

1	https://ejurnal.jejaringppm.org/index.php/jitcsa/article/view/118 INTERNET 20%	
2	https://www.mdpi.com/2076-3417/13/7/4550 INTERNET 2%	
3	https://www.learndatasci.com/glossary/tf-idf-term-frequency-inverse-document-frequency/ INTERNET 1%	
4	https://www.geeksforgeeks.org/understanding-tf-idf-term-frequency-inverse-document-frequency/ INTERNET 1%	
5	http://ejurnal.jejaringppm.org/index.php/jitcsa/article/download/118/63 INTERNET 1%	
6	https://medium.com/@mistrytejasm/understanding-tf-idf-term-frequency-inverse-document-frequency-for-machine-learning-f5f54c1ae897 INTERNET 1%	
7	https://ejurnal.jejaringppm.org/index.php/jitcsa/article/download/118/63/501 INTERNET <1%	
8	https://www.analyticsvidhya.com/blog/2021/01/a-guide-to-the-naive-bayes-algorithm/ INTERNET <1%	
9	https://constructionreviewonline.com/biggest-projects/jakarta-international-stadium-project-timeline-and-all-you-need-to-know/ INTERNET <1%	

EXCLUDE CUSTOM MATCHES	ON
EXCLUDE QUOTES	OFF
EXCLUDE BIBLIOGRAPHY	OFF