

Received : Dec 12, 2023
Accepted : Dec 20, 2023
Published : Mar 28, 2024

Conference on Management, Business,
Innovation, Education and Social Science
<https://journal.uib.ac.id/index.php/combines>

Design of a Video Game Review System Using Naïve Bayes Algorithm

Diny Anggraini Adnas¹, Eric Sanjaya²

E-mail: diny.anggriani@uib.ac.id ¹, 2031027.eric@uib.edu²

¹ Faculty of Computer Science, Universitas Internasional Batam, Batam City, Indonesia

² Faculty of Computer Science, Universitas Internasional Batam, Batam City, Indonesia

Abstract

The rapid development of the technology marks the emergence of social medias. Collaboration, two-way communication, and friendships with other social media users are facilitated through functions such as comments (comment/retweet) and likes or dislikes. This research aims to examine how video game reviews are perceived within a system based on data obtained from Kaggle. Data mining processes will be conducted using machine learning methods, specifically the Naïve Bayes Classifier algorithm. This classification method is based on probability and Bayes' theorem, assuming attribute independence. The research results indicate that the use of Naïve Bayes shows a significance level of 0.46%.

Keywords: Review, Naïve Baiyes, Data Mining, Machine Learning, System

Introduction

With the rapid growth of technology(Nikan & Adnas, 2022), different forms of social media have emerged, promoting collaboration, two-way communication, and friendships among users through features such as chats. According to a January 2014 Global Web Index survey, Indonesia had 72,700,000 internet users out of a population of 251,160,120. As a widely sought-after type of entertainment, the world of video games is not lagging behind in technical innovation. Games with features that are enjoyable, motivating, creatively addictive, will have people find pleasure in engaging in this activity(Yulianto et al., 2018). Video games, which involve user interface interaction via video devices, are a notable feature of this age(Abdullah et al., 2022).

A website can be considered the most popular medium on the internet. Almost all internet users consistently utilize website media, whether for searching or disseminating information on the internet(Yusup et al., 2021).

This industry has bright possibilities, with video game sales providing a profitable route for developers to earn quickly. As a result, proper data is critical for developers to capitalize on this fast-paced market. A game's quality extends beyond its design and graphics; features such as sound, plot, and gameplay are critical in improving a game's overall quality(Hisham et al., 2022). As a result, data processing must be digitized in order to give timely and reliable

information, this way, consumers will understand the game more before purchasing impulsively, and the developer will know what the potential consumers are looking for (Anwar et al., 2020). The goal of this study is to create and build a system capable of detecting reviews from a list of games on the market using sentiment analysis. Data mining is a part of the Knowledge Discovery in Databases (KDD) process, which consists of several steps, such as data selection, pre-processing, transformation, data mining, and result evaluation. KDD is also frequently referred to as a database (Zai, 2022). Machine learning can be defined as the application of computer and algorithmic methods to somehow master the mathematical learning derived from data and make predictions about the future (Roihan et al., 2019). This system's data source is sourced from Kaggle. Machine learning will be used to mine data along with the Naïve Bayes Classifier method based on probability and Bayes' theorem, assuming attribute independence (Parlina et al., 2019).

Naive Bayes is one of the popular algorithms employed for data mining purposes due to its user-friendly nature, fast processing, easy implementation with a relatively simple structure, and a high level of effectiveness (Khotimah & Utami, 2022). The Naïve Bayes algorithm assumes that all attributes are independent or free from the influence of the class variable (Kausar et al., 2023). The Naive Bayes algorithm is widely utilized for addressing prediction problems in the form of classification (Sinaga et al., 2022). Therefore, the Naive Bayes Classifier algorithm is the appropriate choice due to its ability to achieve high accuracy and processing speed when dealing with large training datasets. It performs well in classification even when using small training data (Mantik et al., 2022).

Natural Language Processing (NLP) sentiment analysis will assist the researcher in categorizing tweets and evaluating responses. Given the changing gaming business, the researcher believes it is vital to examine video game suggestions. The goal is to determine whether a video game's quality is high enough to suggest to potential consumers. Data collection and interviews with respondents are part of this analysis. As a result, the research title has been chosen as "**Design of a Video Game Review System Using Naïve Bayes Algorithm.**"

Literature Review

(Irawati Setiawan et al., n.d.) carried out a study in 2021 that serves as the key reference for the researcher in this study. Their study focuses on a system that can detect the veracity of news distributed via the Twitter social media network by employing the Naive Bayes algorithm. The study concerns the prevalence of negative news, which can have a negative impact on society, potentially leading to strife. The public's proclivity to believe unconfirmed news is a major source of concern. Provocations or hate speech directed at certain parties or groups are common in such news, contributing to social divisions. To calculate incoming data, the Naive Bayes method is used, and the research results show that opinions in tweets attained the best accuracy of 93.33% with an 8-fold model. While the lowest accuracy of 85% was obtained with a 5-fold model. The average accuracy across 10 folds was 89%.

Another 2020 study by (Suharno et al., 2020) examines the Scrum method's adoption in a project management information system. Nyoman's research emphasizes the necessity for firms in the digital business to incorporate information technologies into their organizational operations. Scrum and SDLC Waterfall software development approaches are used in the study. The purpose is to evaluate the efficacy of Scrum in digital sectors and companies, comprehend the function of project management systems in Scrum management, and decide whether Scrum

genuinely works in software development for digital enterprises. The research findings suggest that Scrum may be efficiently adopted in software development in digital firms, which are noted for their speed and simplicity.

Also in 2020, Yaulie D. Y. Rindengan and Arie S did a sentiment analysis to better understand public perceptions on social media advances. Tweets were divided into good, negative, and neutral categories using lexical analysis and a R programming language-built system. The study's goal was to collect data on popular sentiment on Twitter following the release of the 2019 election results. The hashtag #JokowiAminSudahMenang received 54% positive feedback, while the hashtag #HidupTakSeanjingItu received 11%. Negative feedback was prevalent in hashtags #02KasetRusak and #PrabowoBukanPemimpin, both of which had 51% negative feedback, while #98jagademokrasi had 10% negative feedback.

Lastly, In 2022, (R.H. Zer et al., 2022) and colleagues completed a study employing the C4.5 method based on PSO to provide a more substantial grasp of machine learning with website programming. The study sought to determine the key obstacles impeding students' grasp of website programming at AMIK Tunas Bangsa by categorizing the most significant aspects. The findings indicate that machine learning can be used to classify data and solve problems in a variety of sectors with great accuracy and precision. According to the study, the C4.5 method based on PSO with a machine learning technique can analyze website programming understanding with an accuracy of 83.00%, with the highest node being the Ease Factor.

In summary, the reviewed studies highlight critical dimensions of information technology's influence on society. Esther Irawati Setiawan and colleagues' 2021 study emphasizes the urgency of accurate information dissemination in the age of social media, showcasing the Naive Bayes algorithm's effectiveness in detecting false news on Twitter. Similarly, Nyoman Gunantara and Made Sudarma's 2020 research advocates for the seamless integration of the Scrum method in software development for digital businesses, emphasizing adaptability to the evolving technological landscape.

Furthermore, Yaulie D. Y. Rindengan and Arie S's 2020 sentiment analysis study unveils the power of social media in shaping public opinion following significant events. Lastly, R.H. Zer P, Hayadi B, and colleagues' 2022 research demonstrates the potential of machine learning, particularly the C4.5 method based on PSO, in addressing educational challenges. Together, these studies offer valuable insights into the nuanced dynamics of technology and society, underlining the importance of accuracy, adaptability, and a comprehensive understanding as we navigate the complex terrain of the digital age. These findings provide a foundation for future research and practical applications in an ever-changing technological landscape.

Research Methods

This study is an applied and experimental study that aims to determine the findings of the Naive Bayes algorithm and analyze its correctness in generating video game review data. The researchers put these findings to the test by comparing predictions made by the Naive Bayes algorithm and experimental approaches. The author outlines and depicts each phase of the activities in the flowchart below.

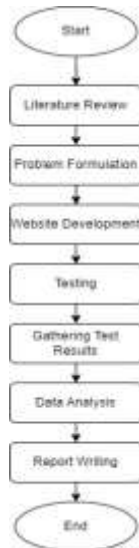


Figure 3.1 Research Flowchart

According to the flowchart above, the study begins with a search for numerous journal references on issues connected to this research and the creation of a literature review. Following the completion of the literature evaluation, the research moves on to the formulation of the research problem. Application development occurs after the definition of the research problem. As a data mining tool, the website is built with the Nave Bayes method. The site is then tested to ensure that it runs well. Data is gathered through interviews with video game users utilizing experimental methods. The collected data is initially examined, and then a data reconciliation process between the data received from the application results and the data obtained from the interviews with video game players is performed. The final stage after completing all steps is to compose the research report.

Meanwhile, the author will use the Scrum approach to establish a video game recommendation website throughout the website development phase. Figure 3.2 depicts the Scrum method's framework.



Figure 3.2 Scrum Method's Framework

The planned information system project begins in the Pre-game stage, with the researcher compiling a list of needs and planning activities for the subsequent phases. This stage produces two results: User Stories and a Product Backlog. Following that, the project

enters the game development stage, sometimes known as the sprint stage. The sprint stage is divided into sections. The sprint planning stage is the initial stage. Sprint planning is a step of planning that addresses which tasks will be completed and when they will be completed. The sprint backlog is the stage that comes after the sprint planning stage. The sprint backlog stage is a Sprint Planning component that provides a prioritized list of functionalities, as well as a brief description of the actions required to produce the functioning version of the product, usually at the start of the sprint backlog creation process. This is not an entire list, and it will be updated when new pages are created. The sprint review is the following step. The sprint review is an activity that takes place at the end of the daily scrum meeting to examine the work results. The Sprint Retrospective is the sprint's final step, in which the progress of product requirements is addressed and scheduled on the same day.

Results and Discussion

This system is separated into two roles: the administrator role and the user role. An explanation of each system interface will begin with the login page, as seen in Figure 3.3:



Figure 3.3 Login Page Interface

To gain access to the system, users can commence the login procedure, which will then proceed with username and password validation, as shown in Figure 3.3 above. Users can also create an account by clicking the register button and entering the needed information. Email, name, username, password, and password confirmation are all required information for the registration procedure. The admin will be taken to the admin dashboard page after successfully logging in. The admin dashboard is a page used to manage the website's content, as seen in Figure 3.4 below.



Figure 3.4 Dashboard Page

Upon successful login, the system displays many menus to the administrator, as seen in Figure 3.4. The dashboard is the first menu page. The admin receives an explanation of how to utilize the website on this page, which is offered in a concise description. This guide is meant to aid the administrator in running the website and reducing confusion during system use. The second menu describes the "Add Game" page that will be displayed on the website. Figure 3.5 depicts the "Add Game" page as follows.



Figure 3.5 Add Game Page

Figure 3.5 shows how, in order to add game data, the admin must fill out various essential fields. These comprise the title of the game, its category, console, developer, release date, and image URL. The admin can then confirm the data addition by clicking the submit

button. Furthermore, the admin can alter game data by going to the "Edit Game" menu page. Figure 3.6 depicts the appearance of the "Edit Game" option as follows.



Figure 3.6 Edit Game Page

When users browse the website, the user interface displays a catalog of games available in the system. The following is an explanation of the website's primary view in Figure 3.7.



Figure 3.7 Website Main Page

When a user wishes to check reviews for a specific game, they can click on the "Search Data" button and enter the game's name. After that, the system will search for and match the user's input. If the system finds the requested game, the reviews for that game will be displayed. Figure 3.8 is an explanation of the review display.



Figure 3.8 Review Page

The author's primary contribution to the project centered on the design and implementation of a Video Game Review System, wherein He played a pivotal role in integrating the Naïve Bayes Algorithm for sentiment analysis and review classification. The Naïve Bayes Algorithm was selected for its efficiency and effectiveness in handling large datasets, making it particularly well-suited for analyzing the plethora of user reviews within the gaming industry. This algorithmic approach not only enhances the accuracy of sentiment analysis but also provides insightful classifications of user sentiments towards video games. By leveraging the Naïve Bayes Algorithm, The system aims to revolutionize the way user feedback is processed and interpreted. The implications of this study extend to game developers, marketers, and consumers, offering a more nuanced understanding of user sentiments and influencing

Commented [TW1]: No discussion to explain your contribution and study implication

purchasing decisions. While the author encountered some challenges during the implementation, such as fine-tuning the algorithm for gaming-specific nuances, the author's efforts have paved the way for a robust Video Game Review System with promising implications for the gaming industry.

Conclusions

According to the author's research, the scraped data effectively acquired information from 6001 game users and extracted their remarks. The Naive Bayes Classifier (NBC) was used in the analysis. This Naive Bayes Classifier technique was used to collect a large volume of comment or opinion data. However, due to its accuracy value of 0.48, it was discovered that employing this algorithm for game reviews may not be appropriate. For future research, the author advises comparing two or more methodologies to evaluate which method is more appropriate, accurate, and successful for further sentiment analysis data processing.

References

- Abdullah, H. A., Risky, D., Putra, D., & Azhar, Y. (2022). *ANALISA PENJUALAN VIDEO GAME MENGGUNAKAN METODE ENSEMBLE*. 12(3), 8–16. <https://jurnal.umj.ac.id/index.php/just-it/index>
- Anwar, S., Andrawina, L., & Rizana, A. F. (2020). PERANCANGAN SISTEM INFORMASI UNTUK PENGELOLAAN DATA WARGA DALAM TINGKAT RT DENGAN METODE SCRUM INFORMATION SYSTEM DESIGN FOR MANAGING DATA OF CITIZENS IN THE NEIGHBORHOOD LEVEL WITH SCRUM METHOD. *E-Proceeding of Engineering*, 7(2), 6137–6146.
- Hisham, M. R., Pratama, J., Andito, L., Kho, A., & Wijaya, H. (2022). Analisa Klasifikasi Genre Game PC Terpopuler. *Journal of Information System, Graphics, Hospitality and Technology*, 4(01), 27–31. <https://doi.org/10.37823/insight.v4i01.145>
- Irawati Setiawan, E., Johanes, S., Tandy Hermawan, A., & Yamasari, Y. (n.d.). Deteksi Validitas Berita pada Media Sosial Twitter dengan Algoritma Naive Bayes. *JOURNAL OF INTELLIGENT SYSTEMS AND COMPUTATION*, 55–60.
- Kausar, A., Irawan, A., Fernando, I., Teknik Informatika, P., Sistem Komputer, P., Informatika, P., & Teknologi Informasi Universitas Serang Raya, F. (2023). *IMPLEMENTASI ALGORITMA NAÏVE BAYES CLASSIFIER UNTUK PENILAIAN KINERJA DOSEN*. 10(2).
- Khotimah, A. C., & Utami, E. (2022). COMPARISON NAÏVE BAYES CLASSIFIER, K-NEAREST NEIGHBOR AND SUPPORT VECTOR MACHINE IN THE CLASSIFICATION OF INDIVIDUAL ON TWITTER ACCOUNT. *Jurnal Teknik Informatika (JUTIF)*, 3(3). <https://doi.org/10.20884/1.jutif.2022.3.3.254>
- Mantik, J., Novitasari, I., Basuki Kurniawan, T., Arrova Dewi, D., Ilmu Komputer, F., Bina Darma, U., Selatan, S., Ilmu Komputer, M., Nusa Putra, U., Barat, J., Komputer, T., & Vokasi, F. (2022). Analisis Sentimen Masyarakat Terhadap Tweet Ruang Guru Menggunakan Algoritma Naive Bayes Classifier (NBC). In *Jurnal Mantik* (Vol. 6, Issue 3). Online.
- Nikan, M. U., & Adnas, A. (2022). The Effect of Online Video Marketing Promotion and Brand Awareness on Consumer Purchasing Decisions at MSMEs Make Di Fried Chicken. In *JURNAL INFOKUM* (Vol. 10, Issue 5). <http://infor.seaninstitute.org/index.php/infokum/index>
- Parlina, I., Yusuf Arnol, M., Febriati, N. A., Dewi, R., Wanto, A., Lubis, M. R., & Susiani. (2019). Naive Bayes Algorithm Analysis to Determine the Percentage Level of visitors the Most Dominant Zoo Visit by Age Category. *Journal of Physics: Conference Series*, 1255(1). <https://doi.org/10.1088/1742-6596/1255/1/012031>

- R.H. Zer, P. P. P. A. N. W. F. I., Hayadi, B. H., & Damanik, A. R. (2022). PENDEKATAN MACHINE LEARNING MENGGUNAKAN ALGORITMA C4.5 BERBASIS PSO DALAM ANALISA PEMAHAMAN PEMROGRAMAN WEBSITE. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(3). <https://doi.org/10.23960/jitet.v10i3.2700>
- Roihan, A., Abas Sunarya, P., & Rafika, A. S. (2019). IJCIT (Indonesian Journal on Computer and Information Technology) Pemanfaatan Machine Learning dalam Berbagai Bidang: Review paper. In *IJCIT (Indonesian Journal on Computer and Information Technology)* (Vol. 5, Issue 1).
- Sinaga, S., Sembiring, R. W., & Sumarno, S. (2022). Penerapan Algoritma Naive Bayes untuk Klasifikasi Prediksi Penerimaan Siswa Baru. In *Journal of Machine Learning and Data Analytics (MALDA)* (Vol. 1, Issue 1).
- Suharno, H. R., Gunantara, N., & Sudarma, M. (2020). Analisis Penerapan Metode Scrum Pada Sistem Informasi Manajemen Proyek Dalam Industri & Organisasi Digital. *Majalah Ilmiah Teknologi Elektro*, 19(2), 203. <https://doi.org/10.24843/mite.2020.v19i02.p12>
- Yulianto, F., Utami, Y. T., & Ahmad, I. (2018). GAME EDUKASI PENGENALAN BUAH-BUAHAN BERVITAMIN C UNTUK ANAK USIA DINI. In *Jurnal Nasional Pendidikan Teknik Informatika /* (Vol. 7, Issue 3).
- Yusup, A., Prayoga, H., Sari, N., & Maudunah Urubah, D. (2021). Perancangan Sistem Informasi Pembayaran Piutang Berbasis Website pada PT. Wira Sarana Karawang. *JUSTIFI*, 1(1). <http://jurnalilmiah.id/index.php/justifi>
- Zai, C. (2022). IMPLEMENTASI DATA MINING SEBAGAI PENGOLAHAN DATA. *Portaldata.Org*, 2(3).