JOURNAL OF DATA SCIENCE | Vol.2022:05
eISSN:2805-5160

Movie Recommendation System Based on Sentiment Analysis on Movie Reviews

Astried[1], Tri Basuki Kurniawan[2*], Mohd Zaki Zakaria[3], Misinem[4], Mohamed Fakhrul Faris bin Mohamed Fuad[3]

[1]Faculty of Mathematics & Science, University of Riau, Pekanbaru, Riau, Indonesia
[2]Faculty of Computer Science, University of Bina Darma, Palembang, Indonesia
[3]Faculty of Computer & Mathematics Sciences, University Technology Mara, Malaysia
[4]Faculty of Vocational, University of Bina Darma, Palembang, Indonesia

[*]**Email**: tribasukikurniawan@binadarma.ac.id

Abstract

A movie review plays a significant role in determining whether the movie is recommended to them. Nowadays, movie reviews are filled with paid, sarcasm, and fake reviews that give users mixed feelings about the movie. With a movie recommendation system based on sentiment, the movie review is analyzed with specific keywords that give the user an absolute result on whether it is recommended or not recommended. The system uses three classifiers, Naïve Bayes, Support Vector Machines (SVM), and Deep Learning to determine the best classifier that gives better accuracy for user purposes. The core approach of this project is to provide the user with a simplified view that analyses all accumulated reviews into one single view. The project results show that SVM produces the best results with 81.17% accuracy. That result is because of the nature of the classification that works best in categorization. This project includes future works, adding more lists of movies and user input for better interactivity between users and machines.

Keywords

Movie, Review, Naïve Bayes, Support Vector Machines, Deep Learning

Introduction

Movies, also known as films, are visual communication that uses moving pictures and sound to tell stories or inform. People around the world watch movies as a type of entertainment. Most movies are created to be shown in movie theatres. A moviegoer is a person who watches movies. The review or critique acquired from them is known as a movie review. These movie reviews can be divided into two categories, good or bad reviews, where a good review is mostly positive while a bad review is mostly negative about the movie.

Submission: 2 June 2022; **Acceptance:** 10 June 2022

Movies, also known as films, are visual communication that uses moving pictures and sound to tell stories or inform. People around the world watch movies as a type of entertainment. Most movies are created to be shown in movie theatres. According to Internet Movie Database (IMDb), currently, there are a total of 15 popular genres of movies comprising comedy to superhero genres. This domain was chosen because it is fascinating as it consists of a vast amount of data to be explored.

In a new age of technology, all information can be found at the fingertips. While it can benefit the user, some hurt them. For example, most movie reviews on popular movie review websites such as Internet Movie Database (IMDb) or Rotten Tomatoes are unreliable or untrusted. This situation is because the study is constructed by people with different beliefs and personalities, thus affecting their judgment criteria. The reviews also were built in the form of marketing, so they will gain audiences to come and watch their movies. Besides, existing movie review sites make it hard for a user to decide because many reviews are presented in a complex way. These factors make it hard to determine whether the movie will suit the taste and personality of the user who will watch the movie.

Twitter can present the data more accurately, allowing 280 characters per tweet (Gligorić, Anderson, & West, 2020). It means the system will perform tasks faster in detecting keywords for sentiment analysis. The tweets also need to mention at least a movie involved in the project. It also must come from a legitimate account at a specific time. To compare with conventional review sites, in this paper, three (3) movies were chosen with different review scopes from IMDb (bad, neutral, and good movie ratings) to be analyzed in this system.

Sentiment Analysis

Sentiments are the users' emotions toward specific entities such as products, events, issues, or services labeled as excellent, good, neutral, or bad. Sentiment analyses are examinations of user emotions based on online feedback. They are also known as sentiment mining, opinion extraction, or review mining (Kaur, 2017). Sentiment analysis can be complicated because user identification and user group extraction are required to do the analysis (Li, Cui, Shen, & Ma, 2016).

Currently, there are two (2) types of sentiment analysis: basic classification. It is the most popular type of classification because it is easy to differentiate. For example, a restaurant review like positive, negative, or neutral (Pandarachalil, Sendhilkumar, & Mahalakshmi, 2015). Then, an advanced classification involves a more detailed level of variety to provide an accurate opinion. For example, human emotions include happiness, sadness, confusion, and anger (Parveen & Pandey, 2016).

Naïve Bayes

Naïve Bayes Classification is a supervised learning method that captures uncertainty about the model by determining probabilities) (Parveen & Pandey, 2016) This algorithm was created by Thomas Bayes, who proposed the Bayes Theorem to determine the positive possibility of a particular corpus in a document. This model utilizes a bag of word feature extraction. This classification can be applied using tools trained by the SentiWordNet dictionary, available

publicly. It helps determine the hypothesis's exact probabilities and better manage noise in the input.

Support Vector Machines

Support Vector Machines (SVM) are supervised learning models with learning algorithms that analyze data and recognize overall patterns. The basic SVM takes a set of data input and then predicts each given input, which outputs the two possible classes, thus making it a non-probabilistic binary linear classifier. SVM is similar in functional form to neural networks and radial basis functions. (G K & Sukumaran, 2013) Thus making it suitable as a training algorithm for learning classification and regression rules from a set of data. An application example is polynomial, multilayer perceptron, and radial basis, from neural networks and radial basis function, respectively.

Deep Learning

Deep Learning, also known as deep structured learning or hierarchical learning, is a machine learning method based on an artificial neural network. This classification can be supervised or unsupervised (Kahou et al., 2013). It is formed by a composition of multiple nonlinear transformations to achieve a more abstract and helpful representation. This classification is used chiefly in speech recognition, which yields better results than other classifications.

Dynamic Convolutional Neural Network

As the name suggests, other classification, such as Dynamic Convolutional Neural Network (DCNN), is a part of neural network classification, which uses a convolution architecture that changes between broad convolution layers with dynamic pooling layers calculated by k-max pooling. Wide Convolution can be obtained from convolving a matrix of weights m∈ R^(d×m) with the matrix of activations at the layer below (Kalchbrenner, Grefenstette, & Blunsom, 2014). It is proven that the DCNN algorithm provides more accurate results than the Naïve Bayes algorithm (Ratnawati & Winarko, 2018).

Web Scraping

Web scraping is a technique to extract structured data from websites. The goals are to pull out valuable data to utilize other usages, such as improving other functions. The process involves extracting data from internet pages, recognizable proof, matching and combining comparative data, opinion extraction from online sources, idea progression, metaphysics, or learning mix (A. Yani, Pratiwi, & Muhardi, 2019). There are various ways to scrap data by using a programming language such as Python, C/C++, Java, R, or Ruby. Currently, Python is the most popular programming language for web scraping as many plugins are available such as BeautifulSoup and Scrapy. A technology solution to extract data from a website in the most efficient way, automated, and presented in a structured format is web scraping tools (Public, Information, Topic, & No, 2015). Figure 1 shows a few tools that are currently available to use.

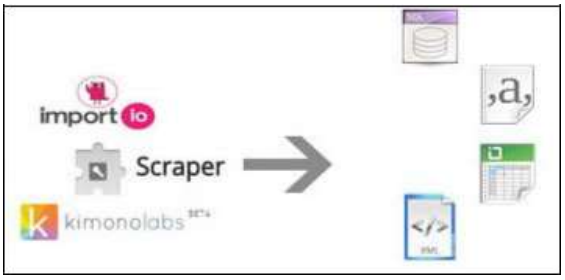

Figure 1. Online tools are available to perform data scraping into a various file formats

Sentiment Analysis of Movie Reviews and Blog Posts

In (Singh, Piryani, Uddin, & Waila, 2013) utilize the Internet Movie Database (IMDb) for sentiment analysis. The user reviewing movies on this site often ranges from a group of people who is diverse. It recently watched the movie or regular to the particular movie. The website proves the benefits of the review, which reflects the opinion of a large number of users. On the word lists, SentiWordNet is used because it is available for free to the public. A polarity value between -1 to +1 is given for every term indicating all words' positivity or negativity. The score is then aggregated to make use of classification efficiently. This method is also known as data pre-processing. The method of the project is by setting up variables among SentiWordNet methodology, which utilizes adjectives or adverbs, or both. The conclusion of the final results shows that the adjective shows the best performance in accuracy, thus concluding that the adjective plays an essential role in sentiment analysis.

Research Methodology

Methodology explains the task involved in every phase that needs to be done to achieve them. The steps need to be done to deliver an expected outcome. This project methodology consists of 7 stages: preliminary study and knowledge acquisition phases to achieve the first objective, data collection and data analysis steps to achieve the second objective, and system design. Subsequent system development and system testing and evaluation phases to accomplish the third objective are shown in figure 2.

Preliminary Study

A preliminary study is the first phase to complete the first objective, identifying people's preferences when watching the movie and creating a personalized movie for them. In this phase, an extensive understanding of movies is required. The study can be done by reading some literature reviews of journal articles, presentations, and e-books. Movie review websites such as IMDb.com and RottenTomatoes.com can be a great place to extract data from them. The problem statement, objective, scope, significance, and aim are formulated by exploring literature reviews.

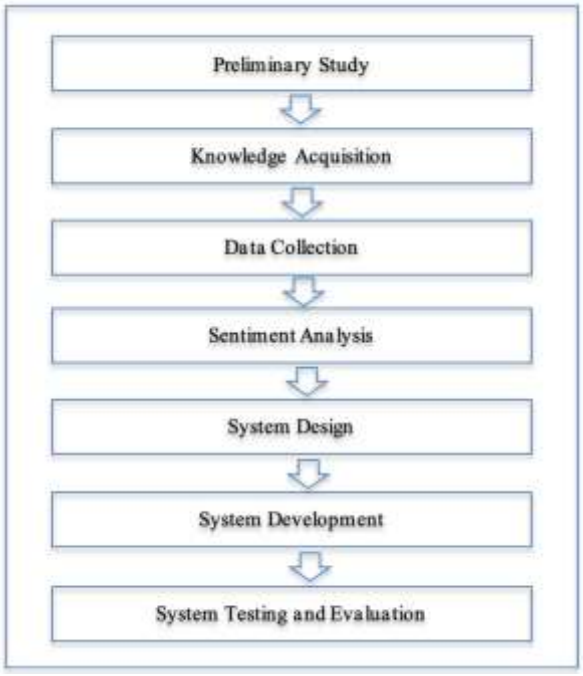

Figure 2. Phases on methodology approach

Knowledge Acquisition

Knowledge acquisition is when knowledge and information about the project are collected. The second phase is also needed to complete the first objective. Activities done in this phase are thoroughly gathering knowledge from the literature review from the library and the internet. Library research is done at Tun Abdul Razak Library at UiTM Shah Alam. The literature review can be obtained from Google Scholar, ScienceDirect, Scopus, and other journal-related articles for online research.

Data Collection

Data collection is the third phase, choosing a suitable approach to collect and retrieve data. In this case, IMDb.com is used as it is the most popular movie review website. An online tool, the Data Miner extension for Google Chrome is used as an extractor tool to get structured data from websites. The data is then divided into two (2) files, one file with a class column and another with no class. It is also known as training and testing set, respectively. The scraped data involves five (5) genres; action, adventure, animation, comedy, and romance, with two (2) movies in each genre. The review of the movie is being gathered for analysis. The movie review will include a few essential keywords to classify it into different sentiments.

Sentiment Analysis

Sentiment analysis is the fourth phase which is data gathered from the previous step is processed to be analyzed, thus producing helpful information. This process includes data pre-processing that involves data cleaning and data integration. RapidMiner is used for sentiment classification. The data is being read by the model and by performing data pre-processing to obtain the polarity and confidence of the review.

System Design
System design is the fifth phase where architecture, interface, data and modules are defined to satisfy specified requirements. Data cleaned from the previous step will undergo system design, presenting valuable information.

System Testing and Evaluation
System testing and evaluation are the last phases where tuning system and documentation activity be done. The public will evaluate moviegoers to emulate the accuracy of the real world.

Results and Findings

This chapter highlights the evaluation conducted to determine the reliability and accuracy of the prototype. It includes data results, comparison with each other, discussion on the classifier used, and an overall summary of results and findings.

Data Parameters
Ten (10) movies are used for sentiment analysis in this project. All the movies are divided into five (5) genres, with two (2) movies in each genre. The review data is collected from every movie. A mix of "good" and "bad" movies is composed to determine whether the rating is proportional to the sentiment analysis result. The genre used is portrayed in the following table. Next, the data is summed, adding up to 23k reviews, as shown in table 1.

Table 1. The genre of movies with titles and ratings

Genre	Title	IMDb rating
Action	Ghostbusters (2016)	5.2 / 10
	Avengers: Infinity War (2018)	8.5 / 10
Adventure	Eragon (2006)	5.1 / 10
	The Lord of the Rings: The Fellowship of the Ring (2001)	8.8 / 10
Comedy	Guardians of the Galaxy (2014)	8.1 / 10
	Deadpool (2006)	8.0 / 10
Animation	Frozen (2013)	7.5 / 10
	Spider-Man: Into the Spider-Verse (2018)	8.5 / 10
Romance	Titanic (1997)	7.8 / 10
	Beauty and the Beast (2017)	7.2 / 10

Data Processing Result
The result analysis is produced from the RapidMiner tool and Python shown here. The first classifier used is Naïve Bayes and uses the earlier dataset, as shown in table 2.

Table 2. Confusion matrix and performance vector for Naïve Bayes

n=2804	False Negative	False Positive	
True Negative	497	60	557
True Positive	1758	489	2247

Accuracy	35.16%
Precision	22.04%
Recall	89.23%

The next classifier used is SVM with the same review dataset. The training and testing process result can be seen in table 3. Compared to Naïve Bayes, this classifier produced significantly higher accuracy at 81.17%. The False Negative value is 76, far lesser than Naïve Bayes.

Table 3. Confusion matrix and performance vector for SVM

n=2804	False Negative	False Positive	
True Negative	105	452	557
True Positive	76	2171	2247

Accuracy	81.17%
Precision	58.01%
Recall	18.85%

Deep Learning was the third classifier used for the training and testing process with the same review dataset. This classifier depends on a multilayer feed-forward artificial neural network trained with stochastic inclination plunge utilizing back-propagation. The activation function used by the neurons in the concealed layers was "Rectifier," which picks the limit of (0, X), where X is the information value. The parameter setting for epochs is 10. Epoch is a process of iteration that happens during the learning process. Based on table 4, the accuracy produced by this classifier is 76.43% which is close to the accuracy of the SVM classifier.

Table 4. Confusion matrix and performance vector for Deep Learning

n=2804	False Negative	False Positive	
True Negative	245	312	557
True Positive	349	1898	2247

Accuracy	76.43%
Precision	41.25%
Recall	43.99%

Results Comparison

The results for the overall movie are recorded and summarized on the chart, as shown in figure 3.

Figure 3. Charts of accuracy from 3 classifiers for all movies

Based on the data results (figure 3), SVM has shown the best consistency and accuracy among the three classifiers used in the project. It is because SVM is suitable for the categorization of data. On the other hand, Naïve Bayes only gives better results when the data have distinct features in small data sets. Deep Learning provides a good accuracy based on the results because of the increasingly using many layers atop each other or a deep learning structure. In other words, it learns based on every layer to produce the best results.

Word Frequency

On the other side, another system based on Python is used to develop the most frequent word mentioned in the review of the movie. It is known as a word cloud. It ensures that the movie's review is viewed as a whole picture with a jumbled word, where the biggest is the most popular keyword while the smallest is the least. For example, the entire movie reviews in a word cloud would look like in figure 4.

Figure 4. Word cloud for all of the review dataset

The terms' movie', 'good', and 'film' is the biggest in the word cloud because most reviewers frequently mention the terms. Since most of the movie reviews are picked from well rating movies, the probability of the reviewer inputting the 'good' term in the review is high. Next, Table 5 shows all the frequent adjective words mentioned in a review for all movies.

Table 5. Table of word cloud with frequently used keyword

Movie Title	Word Cloud	Frequently Used Keyword
Ghostbusters (2016)		good, fun, bad, great
Avengers: Infinity War (2018)		amazing, best, great, good
Eragon (2006)		good, bad, horrible, great

The Lord of the Rings: The Fellowship of the Ring (2001)		good, great, epic, amazing
Guardians of the Galaxy (2014)		fun, good, entertaining
Deadpool (2016)		good, funny, fun, best
Frozen (2013)		best, good, overrated, great
Spider-Man: Into the Spider-Verse (2018)		Best, amazing, great, good
Titanic (1997)		love, epic, best, beautiful

Beauty and the Beast (2017)		good, original, beautiful, classic

Every movie review in the table above perfectly described the movie description and style. Therefore, this word cloud is displayed in the web-based system for users to view.

Conclusion

Firstly, the objective is to identify the keywords that help differentiate between a positive or negative sentiment analysis of movie reviews. This objective has been achieved while creating the model of the project in RapidMiner. SentiWordNet is being used as a word list to give confidence level in each word, whether it is positive or negative, to perform sentiment analysis.

Secondly, the objective is to compare the sentiment analysis techniques of a movie review. In this project, three classifiers have been used in the analysis, Naïve Bayes, SVM and Deep Learning. SVM gives the best results among the three making it the most suitable classifier in this project.

References

A. Yani, D. D., Pratiwi, H. S., & Muhardi, H. (2019). Implementasi Web Scraping untuk Pengambilan Data pada Situs Marketplace. *Jurnal Sistem Dan Teknologi Informasi (JUSTIN)*, *7*(4), 257. https://doi.org/10.26418/justin.v7i4.30930

G K, G. K., & Sukumaran, S. (2013). A study on classification techniques in data mining. In *2013 4th International Conference on Computing, Communications and Networking Technologies, ICCCNT 2013*. https://doi.org/10.1109/ICCCNT.2013.6726842

Gligorić, K., Anderson, A., & West, R. (2020). *Adoption of Twitter's New Length Limit: Is 280 the New 140?* (September). Retrieved from http://arxiv.org/abs/2009.07661

Kahou, S. E., Pal, C., Bouthillier, X., Froumenty, P., Gulcehre, C., Memisevic, R., … Wu, Z. (2013). Combining modality specific deep neural networks for emotion recognition in video. In *ICMI 2013 - Proceedings of the 2013 ACM International Conference on Multimodal Interaction*. https://doi.org/10.1145/2522848.2531745

Kalchbrenner, N., Grefenstette, E., & Blunsom, P. (2014). A convolutional neural network for modelling sentences. *52nd Annual Meeting of the Association for Computational Linguistics, ACL 2014 - Proceedings of the Conference*, *1*, 655–665. https://doi.org/10.3115/v1/p14-1062

Kaur, R. (2017). *Sentiment Analysis of Movie Reviews: A Study of Machine Learning Algorithms with Various Feature Selection Methods*. International Journal of Computer Sciences and Engineering.

Li, H., Cui, J., Shen, B., & Ma, J. (2016). An Intelligent Movie Recommendation System through Group-level Sentiment Analysis in Microblogs. *Neurocomputing*, *210*. https://doi.org/10.1016/j.neucom.2015.09.134

Pandarachalil, R., Sendhilkumar, S., & Mahalakshmi, G. S. (2015). Twitter Sentiment Analysis for Large-Scale Data: An Unsupervised Approach. *Cognitive Computation*, *7*(2), 254–262. https://doi.org/10.1007/s12559-014-9310-z

Parveen, H., & Pandey, S. (2016). *Sentiment analysis on Twitter Data-set using Naive Bayes algorithm*. https://doi.org/10.1109/ICATCCT.2016.7912034

Public, E., Information, S., Topic, P., & No, R. (2015). *Web S craping : Applications and T ools*. 1–31.

Ratnawati, F., & Winarko, E. (2018). Sentiment Analysis of Movie Opinion in Twitter Using Dynamic Convolutional Neural Network Algorithm. *IJCCS (Indonesian Journal of Computing and Cybernetics Systems)*, *12*(1), 1. https://doi.org/10.22146/ijccs.19237

Singh, V., Piryani, R., Uddin, A., & Waila, P. (2013). *Sentiment analysis of movie reviews: A new feature-based heuristic for aspect-level sentiment classification*. https://doi.org/10.1109/iMac4s.2013.6526500