SENTIMENT ANALYSIS ON OAKWOOD RESIDENCE CIKARANG HOTEL REVIEWS ON TRIPADVISOR WEBSITE USING K-NEAREST NEIGHBOR ALGORITHM

Oleh
Rizki Nurul Nugraha[1)], Eviana Tarigan[2)], Yuni Trisnawati[3)]
[1,2,3]Universitas Nasional Jakarta, Universitas Nasional Jakarta
E-mail: [1]rizki.nurul@civitas.ac.id, [2]evianatarigan10@gmail.com ,
[3]yunitrisnawati2815@gmail.com

Abstract

Tripadvisor is used not only as a platform to find and book hotels but also as a comparison material before buying, reviews written for a hotel will greatly influence the decisions of potential guests. One of the hotels that received the majority of Excellent ratings is Oakwood Residence Cikarang. This research has several stages, the first is called the Preprocessing stage. The second stage is to use the TF-IDF method, namely to do word weighting, the last stage is to classify the data using the K-Nearest Neighbor method. The test results related to sentiment analysis at Oakwood Residence Cikarang with the K-NN algorithm get an average accuracy of k = 3 of 90%.

Keywords: K-Nearest Neighbor, Sentiment Analysis, reviews, Tripadvisor

PENDAHULUAN

The pace of development of the times is increasingly developing followed by the development of information technology facilities that are easily accessible anywhere, anytime. The ease of finding information is a challenge for business people to compete to present products and services that are in accordance with customer expectations. Business competition is getting tougher with the provision of facilities to easily access various information both for the services or products that the company offers. Companies are required to remain creative but maintain an original or distinctive feature for their identity. The rapid development of this technology can make everyone connected, from the unknown to the familiar or from the ignorant to the knower. We can analyze or find out the value of a product by searching for it on the internet. From this, of course, there will be a positive or negative perception of various products or companies because they can be assessed from all over the world. All business sectors can develop their business through this digitalization, one of which is the hospitality industry. There are various websites to applications developed to make it easier for potential tourists to determine both attractive destinations, restaurants to be visited, flights that suit the budget to various choices of lodging / hotels available.

A hotel can be defined as a residence or building that has the main business in providing lodging for the public or the public in general and has service services beverages, food, room services, washing and also facilities (Prakoso, 2017). According to Rendrawan & Ngr Yudha Martin Mahardika (2020) the characteristics of the service or service industry are different from the manufacturing industry. The interaction between the provider and the recipient of the service is one of the characteristics of the service. Basically, a hotel is a property that prioritizes professional service in addition to hotel facilities and hotel guests expect quality facilities and services that match their expectations.

The development of technology, which is currently very messaged, will certainly cause various impacts, both positive and negative. From the ease of sharing information, a customer of a product or service can easily share they experience from their business managers can find out how satisfied customers are with the product or service offered.

The Theory of the Expectancy Disconfirmation Model regarding customer satisfaction proposed by Oliver & Sharpe (2014) the results of the analysis of Demolingo & Sriwulandari (2022) explains that consumer satisfaction and dissatisfaction are the result of a comparison between consumer desires before buying and those achieved after doing so. Rahma also revealed that customer satisfaction is one of the non-financial performance indicators used by organizations that can help them achieve their goals. When customers' needs and desires are met, satisfaction is a reflection of how they react that is exposed in the results of the analysis.

Customer satisfaction according to Kotler is the level of a person's feelings after comparing the performance or results he feels with expectations (Kotler, 2002). Whereas customer satisfaction is defined by Richard F. Gerson as the consumer's perception that his expectations have been met or exceeded. Consumer satisfaction, according to Howard and Sheth, is the cognitive situation of the buyer regarding equality or disproportion between the results obtained and the sacrifices made (Gerson, 2002). Irawan defines consumer satisfaction as the accumulation of consumers or customers who use products and services (services). Consumers are satisfied if the quality of the product is satisfactory after they have purchased it. As a result, every new transaction or experience will have an impact on customer satisfaction. Satisfied consumers will share their tastes and experiences with other consumers (Irawan, 2018).

Opinion from Liga Suryadana et al., (2018) regarding the potential for the development of online business in Indonesia which is increasingly grounded and developing very rapidly, the internet is one that has a hand in business growth, especially Indonesian online business. This rapid development is due to a more natural way of internet access for the public these days. The increase in the number of users began in 2012, and this is one of the forerunners of the proliferation of online businesses. The existence of the Covid-19 world disaster can also be one of the causes of the rapid development of business digitally because there are many restrictions on activities in groups / in groups, so people will spend a lot by surfing the internet.

One of the websites that is active and often used for consideration in choosing lodging services is Tripadvisor.com. In February 2000, Tripadvisor was founded in Needham, Massachusetts, in the United States. Tripadvisor was originally a B2B (business-to-business) website. The main purpose is to assemble manufacturers and manufacturers, this website is used by people traveling to discuss their insights or experiences where Tripadvisor is a review medium which is one of the options for tourists to share their hotel-related reviews that are publicly accessible from the internet (Sarudin et al., 2021).

Hotel is a commercial business engaged in services that offer services to guests who stay, in a hotel organization there are various departments such as front office services, housekeeping, food and beverage and recreation. From various reviews on the Tripadvisor website, one of the ones that received an excellent rating is Oakwood Residence Cikarang in 2022. Oakwood Residence Cikarang is a four-star class hotel located in the Jababeka industrial business district, has approximately 335 apartments ranging from Studio type with one, two to three bedrooms. Of course, equipped with various facilities for a comfortable and safe lifestyle, Oakwood Residence Cikarang is the perfect choice for tourists both with business, vacation purposes, with a short or long-term stay.

The ease of accessing information, expressing opinions and expressing feelings related to something experienced through a media such as social media, websites, digital applications or reviews on e-commerce and other online platforms is a natural activity to occur or it can be said to be a form of activity that has become a trend today. To group reviews or reviews at Oakwood Residence Cikarang, the use of sentiment analysis is considered appropriate.

Sentiment analysis is often interpreted as opinion mining. Sentiment analysis can be done automatically, which this is one of the various advantages of sentiment analysis, another advantage is that it can save time and effort.

One of the decision-making methods that can be used is K-Nearest Neighbor, a method for classifying objects based on learning data that is closest to the object (Liwei Kuang, 2008). K-NN is also incorporated in the supervised learning method, which will be the result of an example of a query classified using the majority of categories in the K-NN algorithm. Based on the results of the analysis (Permana et al., 2021) the best class that appears will later become the result of a classification that will also be used in this study.

1.1 Sentiment Analysis

Rozi argues in (Cahyani & Pandu Adikara, 2019) that sentiment analysis is a technique that automatically interprets, extracts, and processes text data to determine the type of sentiment in the text. Meanwhile, the type of art itself is classified as positive and negative. And decision-making on an object can be understood more easily through the sentiment process using that sentiment analysis. According to (Wilson et al., 2005) sentiment analysis is in charge of identifying positive and negative opinions, emotions, and evaluations known as sentiment analysis.

In addition to opinions regarding sentiment analysis can be done automatically, opinions (Setiowati & Helen, 2018) reveal sentiment analysis is used to process user recommendations by determining whether the suggestion/opinion is good, negative, or neutral. Determining the polarity of sentiment can be done manually, but as opinion data from the public increases, it takes longer to manually identify the polarity of views. In addition to using the K-NN method, automatic sentiment analysis for data can also use Naïve Bayes, J48, Random Forest and Support Vector Machine (SVM).

Sentiment analysis and opinion summary Sentiment classification is the process of categorizing an entire document based on how people feel about a particular object. Feature-based Sentiment classification, on the other hand, takes into account opinions about the features of a particular object. Opinion summarization differs from traditional text summarization in that only product features expressed by customers are mined. Opinion summaries do not summarize reviews by selecting subsets or rewriting a few original sentences from the review to capture key points, such as traditional text summaries (Chandrasekaran & Vinodhini, 2012). Sentiment analysis, also known as opinion mining, is the study of people's opinions, sentiments, evaluations, judgments, attitudes, and emotions towards entities such as products, services, organizations, individuals, problems, events, topics, and their attributes. It represents a large problem space (Liu, 2012).

Previous research used as a reference that also discussed sentiment analysis through sentiment review grouping, including (Suryadi et al., 2021) in his article reviewed reviews on the website www.traveloka.com using the Naive Bayes algorithm, as well as by (Setiawan et al., 2022) which discussed sentiment analysis using the SVM algorithm. As for sentiment analysis using the K-NN algorithm method at one of the hotels in Surabaya, namely by (Permana et al., 2021). More details of some of the journals that researchers use as references can be seen in the following table:

Tabel 1.

Previous research on the use of K-Nearest Neighbor

Judul	Penulis: Penerbit	Hasil Pembahasan	Akurasi K-Nearest Neighbor
Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel	Siti Ernawati, Risa Wati JURNAL KHATULISTIWA INFORMATIKA, VOL. VI, NO. 1 JUNI 2018	Pengklasifikasi teks review agen travel dengan pengklasifikasi K-Nearest Neighbor, menggunakan 100 review positif dan 100 review negatif serta enam kata yang berhubungan dengan sentimen yaitu Fast, Good, Great, Bad, Cencel dan Wait. Nilai akurasi yang dihasilkan mencapai 87.00% dengan nilai AUC sebesar 0.916 masuk kedalam kelompok Excellent Classification.	87.00 %

Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine	Riza Cahyani, Indriati, Putra Pandu Adikara Jurnal Pengembangan Teknologi Informasi dan Ilmu Komputer 2019	Klasifikasi menggunakan Boosting Weighted ELM. Proses klasifikasi menghasilkan sentimen yang berupa kelas positif atau kelas negatif. Nilai f-measure tertinggi sebesar 0,953. Nilai optimal parameter yang didapatkan antara lain nilai C = 16, L = 64, dan weak learner = 256 system analisis sentimen yang telah dibangun memperoleh bahwa hasil pengujian menggunakan	95.3 %	Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop! Surabaya	Kevin Permana, Zuda Pradana Putra, Aryo Nugroho: Jurnal Ilmu Komputer dan Bisnis (JIKB) 2022	Booking.com sebesar 87,58%, dan untuk Agoda.com sebesar 98,83%. Hasil yang diperoleh dari analisis sentimen terhadap hotel POP! Surabaya menggunakan algoritma K-Nearest Neighbor memperoleh hasil akurasi optimal pada nilai k=5 sampai dengan k=9 yaitu sebesar 93,3%	93,30%
Analisis Sentimen Review Hotel menggunakan Algoritma Naïve Bayes Classifier	Suryadi, Abdurrahman Ridho, Murhaban. TECHSI: Vol. 13, No. 2, Oktober 2021	algoritma naïve bayes classifier memberikan hasil pengujian klasifikasi dengan akurasi yaitu sebesar 92,47%, presisi 97,33%, recall 91,98%, dan f-measure yaitu 94,56%. Algoritma naïve bayes classifier menghasilkan akurasi performansi yang lebih baik dalam penerapan proses klasifikasi	92,47%	Analisis Sentimen Ulasan Pengunjung Objek Wisata Gunung Bromo pada Situs Tripadvisor	Amalia Anjani Arifiyanti, Mochammad Fuad Pandji Fikri, Bagus Utomo: Jurnal Sistem Informasi dan Telematika (Telekomunikasi, Multimedia dan Informatika) 2022	Dari tiga algoritma pengklasifikasi yang diuji pada penelitian ini, tingkat akurasi tertinggi model klasifikasi dihasilkan oleh algoritma Decision Tree sebesar 91% dan diikuti oleh Naïve Bayes dan Logistic Regression yang masing-masingnya 88%. Tingkat precision, recall, dan f1-measure untuk Decision Tree masing-masingnya sebesar 0,95, 0,62, dan 0,68. Dari hasil yang didapatkan, model klasifikasi perlu ditingkatkan performanya karena model klasifikasi memiliki kecenderungan prediksi ke kelas sentimen positif.	88%
Analisis Online Review Tripadvisor.com Terhadap Pembelian Produk Jasa Akomodasi Di Hotel Manhattan	Rendy Sarudin dan Achmad Ismail: Jurnal Hospitality dan Pariwisata 2021	Review online mempunyai pengaruh yang positif terhadap minat beli. Dengan hasil koefisien determinasi sebesar 51,80 % maka review online untuk Manhattan Hotel yang ada di Tripadvisor.com memiliki kontribusi terhadap minat beli.	51.80%				
Klasifikasi Analisis Sentimen Mengenai Hotel di Yogyakarta	Yuliana Setiowati, Afrida Helen: Scan Jurnal Teknologi Informasi dan Komunikasi	Nilai recall menunjukkan berapa persen data terklasifikasikan dengan benar oleh sistem. Dari hasil analisa menunjukkan bahwa metode SVM unggul dibandingkan dengan metode lainnya, selanjutnya diikuti oleh model Naïve Bayes, KNN, Random Forest dan J48	-			Pelanggan ulasan online diproses menggunakan analisis sentimen dan teks pertambangan. Temuan menunjukkan bahwa pelanggan sering meninjau pengalaman mereka di mana mereka tinggal berdasarkan masa tinggal terakhir mereka. Algoritma Decision Tree lebih baik untuk mengklasifikasikan	
Analisis Sentimen Hotel di Nusa Tenggara Barat Menggunakan Algoritma SVM	Rahmad Arif Setiawan, Adhien Kenya Anima Estetikha, Erry Maricha Oki Nurharyanto, Yoga Asmara, Azis Wahyudi:	Hasil penelitian ini menunjukkan jumlah sentimen positif sebesar 84,97% dan sentimen negatif sebesar 15,03%, dengan nilai akurasi sebesar 92,32%, presisi 93,34% dan nilai recall 92,32%.	92,32%	Tourist sentiment analysis on Tripadvisor using text mining: A case study using hotels in Ubud, Bali	Nasa Zata Dina: African Journal of Hospitality, Tourism and Leisure. 2020	hasil analisis sentimen daripada Algoritma Naïve Bayes dalam bidang akurasi. Namun, di bidang presisi dan ingatan, Nave Algoritma Bayes seringkali lebih baik daripada Algoritma Pohon Keputusan.	90%
Analisis Sentimen Review Pelanggan Hotel Menggunakan Metode K-Nearest Neighbor (KNN) (Studi Kasus: Hotel.com, booking.com, Agoda.com)	Rahma Nimas Safitri. 2020	Hasil dari penelitian yang telah dilakukan untuk situs Hotels.com persentase komentar positif sebesar 32% dan persentase untuk sentimen negatif adalah 68%, untuk situs Booking.com persentase sentimen positif adalah 29% dan untuk sentimen negatif persentasenya sebesar 71%. Situs Agoda.com jumlah persentase positif sebesar 86% dan untuk sentimen negatif sebesar 14%. Berdasarkan uji validitas menggunakan 10-Fold Cross Validation didapatkan akurasi untuk Hotels.com sebesar 94,55%, untuk	87-94 %	Implementasi Algoritma K - Nearest Neighbor Terhadap Analisis Sentimen Review Restoran Dengan Teks Bahasa Indonesia	Risma Putri Fitrianti, Ana Kurniawati, Dina Agusten: Seminar Nasional Aplikasi Teknologi Informasi Yogyakarta. 2019	Pelanggan Tripadvisor cenderung menggunakan kata-kata seperti “malam”, “kolam renang”, dan “waktu” dalam sentimen negatif dinyatakan setelah atau selama tinggal. Implementasi algoritma K - Nearest Neighbor terhadap analisis sentimen review restoran dengan teks Bahasa Indonesia dapat membantu mengatasi hal tersebut dengan sistem otomatis yang dibuat. Hasil akurasi aplikasi pada analisis sentimen menggunakan algoritma K - Nearest Neighbor terhadap	96.61%

review restoran Taman Santap Rumah Kayu dengan confusion matrix sebesar 96,61%.

1.2 Tripadvisor

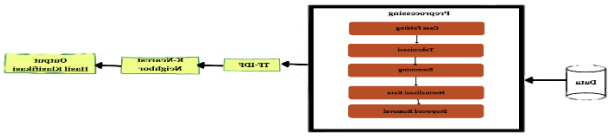
Revealed by Fergian in Sarudin et al., (2021) Tripadvisor is one of the largest travel websites in the world, allowing passengers to plan and book their trips to get the ideal vehicle. Tripadvisor was also founded in 2013 as a hotel and location search service for the most popular tours. Tripadvisor is also the world's most popular online travel site, with 48.5 million users, ranking first followed by Google, as can be seen in figure 1.

Gambar 1. Distribusi Ulasan Hotel menunjukan Tripadvisor setidaknya berada pada peringkat 3 besar

Tripadvisor was originally a B2B (business-to-business) site whose purpose was to bring manufacturers together with manufacturers. Over time, the site was used by the traveller community to share their experiences, and this is where Tripadvisor turned into a reviewer. In addition, there is review content about hotels that can be found through the Tripadvisor website in photos, prices, and descriptions of hotel amenities uploaded by the official Tripadvisor website or directly uploaded by the reviewing person

METHOD

The approach to this study uses descriptive qualitative, where qualitative research is not used in measuring large scale, how much (problems related to numbers and numbers). According to Anggito & Setiawan, (2018) Qualitative research is the collection of data in a natural setting with the intention of interpreting phenomena that occur where the researcher is a key instrument, data analysis is inductive / qualitative, and the results of qualitative research emphasize meaning rather than generalization.



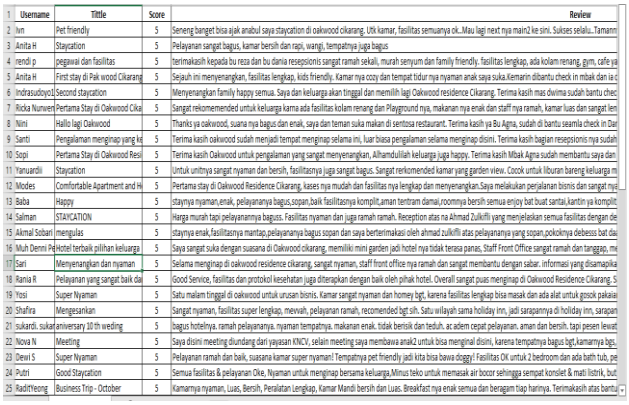
Tika (2005) revealed that descriptive research focuses more on outlining the situation or problem in its current state, but may also offer interpretation or analysis. The findings of this study concentrate on providing an overview of the current situation taken directly from the subject of the study. The qualitative research method that Bogdan and Taylor defined is a research approach that generates descriptive data of people and behaviors observed in the form of written or spoken words. The information generated on human behavior is presented in the form of words and images, in a structured and systematic manner (Moleong, 2002).

The purpose of qualitative research is to concentrate on the topic or issue raised in this study, namely political dynasties. According to Anggito & Setiawan (2018), qualitative research aims to gain a general understanding of social reality from the perspective of participants.

To complete this research, it went through several series of stages that were passed. There are six stages in this study, namely data collection, the second stage of preprocessing consisting of a case folding, tokenization, stemming, word normalization and stopwords, the third stage of TF-IDF, the fourth stage of K-Nearest Neighbor and the last stage is the result of classification.

2.1 Data collection

The data used in this study are various reviews or reviews that have been written by guests who want to share their experiences after staying at Oakwood Residence Cikarang. Of which there are at least 30 samples from tripadvisor reviews that can be seen in the following figure 3.



	Username	Title	Score	Review
2	Ivn	Pet friendly	5	Seneng banget bisa ajak anabul saya staycation di oakwood cikarang. Utk kamar, fasilitas semuanya ok..Mau lagi next nya mainz ke sini. Sukses selalu..Tamann
3	Anita H	Staycation	5	Pelayanan sangat bagus, kamar bersih dan rapi, wangi, tempatnya juga bagus
4	rendi p	pegawai dan fasilitas	5	terimakasih kepada bu reza dan bu dania resepsionis sangat ramah sekali, murah senyum dan family friendly. fasilitas lengkap, ada kolam renang, gym, cafe ya
5	Anita H	First stay di Pak wood Cikarang	5	Sejauh ini menyenangkan, fasilitas lengkap, kids friendly. Kamar nya cozy dan tempat tidur nya nyaman anak saya suka.Kemarin dibantu check in mbak dan ia c
6	indrasudoyo2	Second staycation	5	Menyenangkan family happy semua. Saya dan keluarga akan tinggal dan memilih lagi Oakwood residence Cikarang. Terima kasih mas dwima sudah bantu chec
7	Ricka Nunvert	Pertama Stay di Oakwood Cika	5	Sangat rekomemended untuk keluarga kama ada fasilitas kolam renang dan Playground nya, makanan nya enak dan staff nya ramah, kamar luas dan sangat len
8	Nini	Hallo lagi Oakwood	5	Thanks ya oakwood, suana nya bagus dan enak, saya dan teman suka makan di sentosa restaurant. Terima kasih ya Bu Agna, sudah di bantu seamla check in Dar
9	Santi	Pengalaman menginap yang ke	5	Terima kasih oakwood sudah menjadi tempat menginap selama ini, luar biasa pengalaman selama menginap disini. Terima kasih bagian resepsionis nya sudah
10	Sopi	Pertama Stay di Oakwood Resi	5	Terima kasih Oakwood untuk pengalaman yang sangat menyenangkan, Alhamdulilah keluarga juga happy. Terima kasih Mbak Agna sudah membantu saya dan
11	Yanuardii	Staycation	5	Untuk unitnya sangat nyaman dan bersih, fasilitasnya juga sangat bagus. Sangat rerkomended kamar yang garden view. Cocok untuk liburan bareng keluarga m
12	Modes	Comfortable Apartment and Hi	5	Pertama stay di Oakwood Residence Cikarang, kases nya mudah dan fasilitas nya lengkap dan menyenangkan.Saya melakukan perjalanan bisnis dan sangat nyu
13	Baba	Happy	5	staynya nyaman,enak, pelayananya bagus,sopan,baik fasilitasnya komplit,aman tentram damai,roomnya bersih semua enjoy bat buat santai,kantin ya komplit
14	Salman	STAYCATION	5	Harga murah tapi pelayanannya baguss. Fasilitas nyaman dan juga ramah ramah. Reception atas na Ahmad Zulkiffli yang menjelaskan semua fasilitas dengan de
15	Akmal Sobari	mengulas	5	staynya enak,fasilitasnya mantap,pelayananya bagus sopan dan saya berterimakasi oleh ahmad zulkiffli atas pelayananya yang sopan,pokoknya debesss bat da
16	Muh Denni Pe	Hotel terbaik pilihan keluarga	5	Saya sangat suka dengan suasana di Oakwood cikarang, memiliki mini garden jadi hotel nya tidak terasa panas, Staff Front Office sangat ramah dan tanggap, me
17	Sari	Menyenangkan dan nyaman	5	Selama menginap di oakwood residence cikarang, sangat nyaman, staff front office nya ramah dan sangat membantu dengan sabar. informasi yang disampaikа
18	Rania R	Pelayanan yang sangat baik da	5	Good Service, fasilitas dan protokol kesehatan juga diterapkan dengan baik oleh pihak hotel. Overall sangat puas menginap di Oakwood Residence Cikarang. S
19	Yosi	Super Nyaman	5	Satu malam tinggal di oakwood untuk urusan bisnis. Kamar sangat nyaman dan homey bgt, karena fasilitas lengkap bisa masak dan ada alat untuk gosok pakaia
20	Shafira	Mengesankan	5	Sangat nyaman, fasilitas super lengkap, mewah, pelayanan ramah, recomended bgt sih. Satu willayah sama holiday inn, jadi sarapannya di holiday inn, sarapan
21	sukardi. sukar	aniversary 10 th weding	5	bagus hotelnya. ramah pelayananya. nyaman tempatnya. makanan enak. tidak berisik dan teduh. ac adem cepat pelayanan. aman dan bersih. tapi pesen lewat
22	Nova N	Meeting	5	Saya disini meeting diundang dari yayasan KNCV, selain meeting saya membawa anak2 untuk bisa menginal disini, karena tempatnya bagus bgt,kamarnya bgs,
23	Dewi S	Super Nyaman	5	Pelayanan ramah dan baik, suasana kamar super nyaman! Tempatnya pet friendly! fasilitas OK untuk 2 bedroom dan ada bath tub, pe
24	Putri	Good Staycation	5	Semua fasilitas & pelayanan Oke, Nyaman untuk menginap bersama keluarga,Minus teko untuk memasak air bocor sehingga sempat konslet & mati listrik, but
25	RadiYeong	Business Trip - October	5	Kamarnya nyaman, Luas, Bersih, Peralatan Lengkap, Kamar Mandi bersih dan Luas. Breakfast nya enak semua dan beragam tiap harinya. Terimakasih atas bantu

Gambar 3. Data ulasan Oakwood Residence Cikarang sumber Tripadvisor

2.2 Preprocessing

The basic preparation or before carrying out another core process is the preprocessing stage or preprocessing data. According to Muljiawati in Permana et al., (2021) in running preprocessing requires a tool that one of them the author uses, namely Google Colab, which starts from Case Folding, Tokenization, Steaming, Word Normalization and Stopword. The essence of prepocessing can be carried out by reducing or eliminating inappropriate data or it can also be by changing the data into different forms so that it can be more easily processed on the system. Google Colab is a Python programming or coding environment language with a "notebook" format (comparable to Jupyter notebooks), or in other words, like Google gives us a Term Frequency - Inverse Document Frequency free computer to use to write programs or process the data we have.

2.3 Term Frequency - Inverse Document Frequency

TF–IDF or Term Frequency Inverse Document Frequency is an algorithmic method for calculating the weight of each commonly used word. Andayani and Ryansyah in Permana et al., (2021) explained that this method will calculate the value of Term Frequency (TF) and Inverse Document Frequency (IDF) in each word or token in each document used. Simply put, TF-IDF is used with the aim of being able to find out how often a word is in a document.

2.4 K-Nearest Neighbor

It is one of the methods that determines the algorithm of supervised or supervised. The main difference between supervised learning and unsupervised learning is that supervised learning aims to find new patterns in the data by connecting old and new data. Unsupervised learning aims to find patterns in the data even though the data does not yet have a pattern at the time of use. The purpose of the K-NN algorithm is to categorize new objects based on their attributes and training data (Hermawan & Agung, 2017).

In his book, Indriyanto (2021) revealed that K-Nearest Neighbor has several advantages, including toughness to noisy data training and effectiveness when training large data. K-Nearest Neighbor is a classification algorithm that utilizes all training data (complete storage). A guide to classify the accuracy of diagnostics using AUC, is presented below (Ernawati & Wati, 2018):

1. 0.90-1.00 = Excellent Classification;
2. 0.80-0.90 = Good Classification;
3. 0.70-0.80 = Fair Classification;
4. 0.60-0.70 = Poor Classification;
5. 0.50-0.60 = Failure.

This opinion classification procedure is in the form of sentiment analysis through text mining, so a method is needed that is able to classify opinions accurately. According to Feldman & Sanger, (2007) text mining is a knowledge-intensive process in which users interact with a collection of documents using a set of analysis tools on a regular basis.

K-Nearest Neighbor is a commonly used algorithm to classify data and text. The purpose of this algorithm is to classify objects by their attributes and training samples. The clasifier is not used to match anything and only on the memory given at the query point, the total object k will be found or the nearest training point of enquiry voting is used for the classification of the largest number of classifications of object classifications k (Ernawati & Wati, 2018). The decision making of the number of K or N in the

K-NN method is not decided in the exact formula, in general the consideration of determining K / N can be seen from the amount of data used or processed, if using data that is an odd number, K should be even, otherwise if the data processed is even then it can be determined that K is odd.

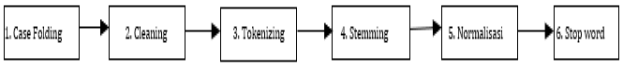
Opinion Helilintar et al., (2017) the k-Nearest Neighbor (k-NN) algorithm found the shortest distance between the data to be evaluated and its closest neighbor k (Neighbor) in the training data. In this method, it was also revealed by Deviyanto et al., (2018) that after the pre-processing stage, the data is ready to be processed with the K-Nearest Neighbor method. The K-Nearest Neighbor method divides the data into predetermined classes based on their distance/degree of resemblance closest to the existing dataset/trainer. Later, the data will be classified based on a number of "k" values that represent their closest distance to the training data.

RESULT AND DISCUSSION

After all the data is collected and processed, the results of the study can be presented with table media, images and graphs and others that can help readers understand the results and discussion in this study. This discussion begins by classifying the reviews that have been obtained from the Tripadvisor website regarding Oakwood Residence Cikarang where the method used is K-Nearest Neighbor as a system for the data processing process.

3.1 Preprocessing

The next stage after the data labeling is carried out, namely preprocessing where there are at least six series of processes passed, details can be seen through the following figure:



Gambar 4. Tahapan *preprocessing* menunjukan enam rangkaian proses

3.2.1 Case Folding

Case folding is although often overlooked, case folding is one of the easiest and most efficient text preprocessing techniques. Case folding is the process of converting all capital letters in a document to lowercase. The only accepted letters are 'a' to 'z'. other characters are omitted, and the remaining ones are considered limiters. Here is an example of the data before and after going through the case folding stage:

Tabel 2.
Tripadvisor Review Data Processing Results case folding stage

No	Sebelum Case Folding	Sesudah Case Folding
0	Seneng banget bisa ajak anabul saya staycation	seneng banget bisa ajak anabul saya staycation
1	Pelayanan sangat bagus, kamar bersih dan rapi	pelayanan sangat bagus kamar bersih dan rapi
2	terimakasih kepada bu reza dan bu dania reseps	terimakasih kepada bu reza dan bu dania reseps
3	Sejauh ini menyenangkan, fasilitas lengkap, ki	sejauh ini menyenangkan fasilitas lengkap ki
4	Menyenangkan family happy semua. Saya dan kelu	menyenangkan family happy semua saya dan kelu

3.2.2 Cleaning

Cleaning is the process performed on "damaged" or inaccurate records or files based on a set of records, tables, or databases that are found and will be repaired (or deleted) during data cleaning. Here is an overview of Google Colab during the cleaning process.

Figure 5. Code Pthyon on Google Colab shows the cleaning process on Tripadvisor Oakwood Residence Cikarang review data

3.2.3 Tokenizing

Tokenizing is the process by which the author collects all the words as well as removes punctuation marks or symbols that are not letters for example “, /. ;) and so on. Here is the result of the tokenizing process carried out with the help of the Google colab site or program.

['seneng', 'really', 'can', 'ajak', 'anabul', 'me', 'staycation', 'di', 'oakwood', 'cikarang', 'utk', 'room', 'facilities', 'everything', 'ok', 'want',

'again', 'next', 'nya', 'main2', 'ke', 'here', 'success', 'always', 'the garden', 'also', 'delicious', 'bgt', 'make', 'play2', 'and', 'relax2']

3.2.4 Normalisasi

Normalization of data at this stage is a process to ensure that all variables have the same range of values, with neither too large nor too small, in order to facilitate statistical analysis.

3.2.5 Stemming

Stemming is the process by which the author changes a token that has a basic word affix through the means of negating any token that has an affix. The most important purpose in this stemming process is to eliminate the initial and final affixes. The stemming process in this study can be seen from the picture below:

```
[31] #stemming (pembentukan kata dasar)
     from nltk.stem import PorterStemmer
     from nltk.stem import WordNetLemmatizer
     stemming = PorterStemmer

[32] def stem_list(row) :
       text = row['Review']
       stem = [stemming.stem(word) for ordw in text]
       return(stem)

       dfreviews_data['Review'] = dfreviews_data.apply(stem_list, axis=1)
       dfreviews_data.Review
```

Figure 6. The steaming process shows the python formula used in Google Colab for the stemming process

3.2.6 Stop word

Stop word is a process that aims to eliminate common words that have no meaning or information needed, including "that, from, with, to" and many other words. The stop word process in this study can be seen and illustrated from the following figure:

```
[33]  #stopword (menghapus kata yang tidak penting)
      from nltk.corpus import stopwords
      import nltk
      nltk.download('stopwords')
      #from nltk.tokenize import word_tokenize
      stops = set(stopwords.words('indonesian'))

[nltk_data] Downloading package stopwords to /root/nltk_data...
[nltk_data]   Unzipping corpora/stopwords.zip.

      dfreviews_data.head()
```

	Review	sentiment
0	seneng banget bisa ajak anabul saya staycation...	1
1	pelayanan sangat bagus kamar bersih dan rapi ...	1
2	terimakasih kepada bu reza dan bu dania reseps...	1
3	sejauh ini menyenangkan fasilitas lengkap ki...	1
4	menyenangkan family happy semua saya dan kelu...	1

Figure 7. The stop word process shows how the process of removing unimportant words in the scrape data

3.2 Data Labeling

After the preprocessing is completed, the next step is data labeling, in this process the researcher carries out data collection through a scraping process, where the data labeling process is carried out subjectively. The example of the result of labeling data is exposed as follows:

Tabel 3.

Oakwood Residence Cikarang in website Tripadvisor result

Review	Sentiment	Review	Sentiment
seneng banget bisa ajak anabul saya staycation di oakwood cikarang utk kamar fasilitas semuanya ok mau lagi next nya main2 ke sini sukses selalu tamannya juga enak bgt buat main2 dan santai2.	Positif	kamar yang nyaman untuk keluarga di saat pandemi saat cek in di berikan masker dan handstinezer dan keramahtamahan karyawan resepsionis tya dan tria perlu dipertankan pilihan makanan sarapan enak membuat tamu yang membawa keluarga lebih bahagia	Positif
pelayanan sangat bagus kamar bersih dan rapi wangi tempatnya juga bagus.	Positif	pelayana receptcion ny ramah kamar jg bersih fasilitas kolam renang jg nyaman view ny jg bagus breakfast nya jg enak enak cocoklah untuk liburan keluarga deket juga dari rumah ga perlu	Positif
terimakasih kepada bu reza dan bu dania resepsionis sangat ramah sekali murah senyum dan family friendly fasilitas lengkap ada kolam renang gym cafe yang banyak makanan dan unit nya bersih terimakasih oakwood residence cikarang.	Positif	Semua good good nyaman ramah kolam renang bersih cocok buat keluarga makananya enak tolong dipertahankan dan karyawan smua bagus bagus respon cepat lokasi dibelakang hotel hollydaiin cikarang	Positif
sejauh ini menyenangkan fasilitas lengkap kids friendly kamar nya cozy dan tempat tidur nya nyaman anak saya suka kemarin dibantu check in mbak dan ia dan waktu check out dibantu mbak akan mbak berdua	Positif	staycation yg menyenangkan kamar dan toilet bersih walau ada kendala pas masuk tv tidak menyala tp langsung di tangani dengan cepat seluruh staff ramah mulai dr receptions kak tria security dan petugas	Positif

ISSN 2722-9475 (Cetak)
ISSN 2722-9467 (Online)

ramah dan baik makasih ya buat team oakwood.		kebersihan yg kami temui selalu menyapa memberikan salam breakfast nya menu nya beragam enak dan nyaman nextime pasti balik lg terimakasih.	
menyenangkan family happy semua saya dan keluarga akan tinggal dan memilih lagi oakwood residence cikarang terima kasih mas dwima sudah bantu check in dan mbak agna sudah bantu proses check out fasilitas bagus dan lengkap.	Positif	tempat dan fasilitasnya enak dan nyaman pelayanannya juga bagus dari receptionistnya ibu tria dan pak izul juga okee pelayanannya managementnya pak edi juga ramah dan yang paling penting kebersihannya sangat terjaga bahkan bantal di kursi tamannya juga diganti.	Positif
sangat rekomemended untuk keluarga karna ada fasilitas kolam renang dan playground nya makanan nya enak dan staff nya ramah kamar luas dan sangat lengkap ada perlengkapan mandi dan kecantikan lainnya,	Positif	dalam kontrak kecepatan wifi adalah sebagian besar waktu hanya 1 dan dari waktu ke waktu kecepatannya kurang dari bahwa saya tidak dapat membuka situs web jadi saya harus menggunakan di kamar saya dan saya menelepon resepsionis berkali kali namun tidak ada perbaikan	Negatif

3.3 K-Nearest Neighbor

The next stage after completion of the preprocessing which uses data with .xlsx type files or data which is a review taken from the TripAdsivor website has a format from Microsoft Excel, the goal is to facilitate the input process in the system. As for the source code used and the test results are as follows:

```
#pembuatan moden KNN
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors = 3)
classifier.fit(X_train_cv, y_train)
```

```
KNeighborsClassifier(n_neighbors=3)
```

```
[48] #prediksi menggunakan data test
y_pred = classifier.predict(X_test_cv)
```

```
#menghitung confusion matrix
from sklearn.metrics import classification_report, confusion_matrix
cm = confusion_matrix(y_test, y_pred)
cr = classification_report(y_test, y_pred)
print(cm)
print(cr)
```

Figure 8. K-NN Source Code shows Google Colab processing data

The test results from the data above can be seen in the following figure:

```
[[0 1]
 [0 9]]
              precision    recall  f1-score   support

     Negatif       0.00      0.00      0.00         1
     Positif       0.90      1.00      0.95         9

    accuracy                           0.90        10
   macro avg       0.45      0.50      0.47        10
weighted avg       0.81      0.90      0.85        10
```

Figure 9. The results of the data test show the results of data processing with the K-NN method

The result above is the processing of data stored in Microsoft excel files, reviews at The Oakwood Residence Cikarang Hotel are processed using the K-NN method with the help of the Google Colab system. The amount of data taken from the Tripadvisor website is 50 reviews, of which at least 49 positive reviews and one negative review. The K-NN classification of the study consists of two steps, the first of which is to determine the parameter k. (the number of nearby neighbors). Tests ranging from k = 3 to k = 10 are needed to determine the value of the parameter k, where the value of k taken is an odd number with an average test sample of 85 percent 90 percent. Based on the results of the trial using a K value of 3.5.7, and 9 obtained the highest accuracy value in the use of the K value = 9, the K value used in this study was 9.

3.4 Sentiment Analysis Results

Tripadvisor is a website that can accommodate various reviews of various properties, accommodations or hotels in the world. The picture below shows a comparison of reviews to Oakwood Residence Cikarang.

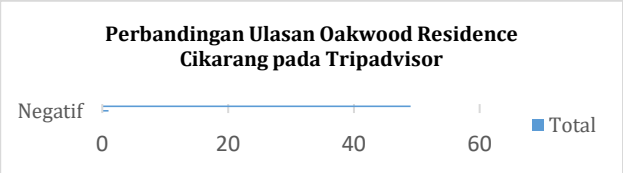


Figure 10. Oakwood Residence Cikarang reviews compare on Tripadvisor

Data shows that negative reviews predominate compared to positive reviews. The number of positive comments for Oakwood Residence Cikarang on Tripadvisor.com was 49 reviews. The number of negative reviews found only 1 review was disappointed with the performance of the wifi used at the time of the guest's stay.

A total of 49 positive reviews can certainly have a good influence on the manager of Oakwood Residence Cikarang because currently as revealed in the results of the analysis (Nugraha, 2018) that types of internet advertising can influence the decisions of customer purchases. Reviews given by guests who have stayed overnight can also be used as a dvertising medium for the managers of Oakwood Residence Cikarang in the process of developing the management of their property.

CONCLUSION AND SUGGESTIONS

This study successfully conducted a sentiment analysis on the Tripadvisor website for Oakwood Residence Cikarang using the K-Nearest Neighbor method. Based on the implementation and evaluation results using the value of K = 9, it can be concluded that Oakwood Residence Cikarang has a higher proportion of positive reviews than negative reviews. The cozy atmosphere and generous staff are often mentioned in guest-led reviews. It has a comparison that is sentiment which is 90% positive reviews and 10% negative reviews.

It can also be seen from the testing carried out if the trend of providing opinions with general media such as on Tripadvisor can grow a significant influence considering that currently the dissemination of information via the internet or online will be very fast and of course positive reviews will also bring a good image to the Oakwood Residence Cikarang Hotel and convince guests to return to stay there. Tripadvisor is one of the online platforms that travellers refer the most to in choosing a place to vacation or stay, the feedback provided by some guests is sometimes a benchmark for other guests to visit.

The advice that the author can give to managers is expected to continue to give appreciation by consistently replying to various reviews given by guests through Tripadvisor or various online platforms because it will provide a very significant experience for guests, namely knowing that the reviews written will be able to help the manager to continue to improve the quality of existing service and the connection of guests and managers will not gap so that guests will also be loyal. Good service quality will usually impress guests and revisit; guests will not hesitate to give a positive review to the hotel if the quality of service continues to improve.

The author realizes that the analysis of the researcher's sentiment still has many shortcomings. Therefore, several factors must be considered for the improvement of similar research in the future, namely expanding the amount of training data used in the classification process. Developing larger data models and subsequent research is expected to use two or more methods to improve system accuracy.

REFERENCES

[1] Anggito, A., & Setiawan, J. (2018). *Metodologi Penelitian Kualitatif*. Sukabumi: CV Jejak.

[2] Cahyani, R., & Pandu Adikara, P. (2019). Analisis Sentimen terhadap Ulasan Hotel menggunakan Boosting Weighted Extreme Learning Machine. *Jurnal Pengembangan Teknologi Informasi Dan Ilmu Komputer*, *3*(8), 2548–2964. http://j-ptiik.ub.ac.id

[3] Chandrasekaran, R. M., & Vinodhini, G. (2012). Sentiment Analysis and Opinion Mining: A Survey. In *International*

Journal of Advanced Research in Computer Science and Software Engineering (Vol. 2, Issue 6). www.zdnet.com

[4] Demolingo, R. H., & Sriwulandari. (2022). Analisis Artificial dan Natural Attraction terhadap Kepuasan Wisatawan. *Syntax Literate: Jurnal Ilmiah Indonesia*, *7*(3).

[5] Deviyanto, A., Didik Wahyudi, M. R., & Informatika UIN Sunan Kalijaga Yogyakarta Jl Marsda Adi Sucipto No, T. (2018). PENERAPAN ANALISIS SENTIMEN PADA PENGGUNA TWITTER MENGGUNAKAN METODE K-NEAREST NEIGHBOR. *Jurnal Informatika Sunan Kalijaga)*, *3*(1), 1–13. https://twitter.com/search?l=id&q=AHY%20since%3A2017-01-01%20until%3A2017-01-

[6] Ernawati, S., & Wati, R. (2018). Penerapan Algoritma K-Nearest Neighbors Pada Analisis Sentimen Review Agen Travel. *JURNAL KHATULISTIWA INFORMATIKA*, *VI*(1), 64–69. https://www.trustpilot.com/categories/tr

[7] Feldman, R., & Sanger, J. (2007). *The text mining handbook: advanced approaches in analyzing unstructured data*. Cambridge University Press.

[8] Gerson, R. F. (2002). *Tingkat Kepusan Pelanggan* (Widyaningrum, Ed.). PPM.

[9] Helilintar, R., Ramadhani, R. A. ·, & Rochana, S. (2017). *DATA MINING K-Nearest Neighbor* (D. P. Pamungkas, Ed.; 1st ed., Vol. 1). Fakultas Teknik Universitas Nusantara PGRI Kediri. https://www.researchgate.net/publication/321804055

[10] Hermawan, F., & Agung, H. (2017). Implementasi Metode K-Nearest Neighbor Pada Aplikasi Data Penjualan PT. Multitek Mitra Sejati. *Jurnal Sains Dan Teknologi: Kalbiscentia*, *4*(2), 103` – 109.

[11] Indriyanto, J. (2021). *ALGORITMA K-NEAREST NEIGHBOR UNTUK PREDIKSI NASABAH ASURANSI*. PT. Nasya Expanding Management (NEM).

[12] Irawan, B. S. (2018). *Manajemen Pemasaran Modern*. Liberty.

[13] Kotler, P. (2002). *Manajemen Pemasaran* (T. Hendra, Ed.). Jakarta: Prenhalindo.

[14] Liga Suryadana, M., Nurul Nugraha, R., & Tinggi Pariwisata Bandung, S. (2018). EFFECT OF APPLICATION OF RICH MEDIA ADVERTISING ON PURCHASE DECISIONS (STUDY OF TRAVEL COMPANIES). *Bisnis & Entrepreneurship*, *12*(2), 135–146.

[15] Liu, B. (2012). *Sentiment Analysis and Opinion Mining*. Morgan & Claypool Publishers.

[16] Moleong, L. J. (2002). *Metodologi Penelitian Kualitatif*. Remaja Rosdakarya.

[17] Nugraha, N. R. (2018). PENGARUH RICH MEDIA ADVERTISING TERHADAP KEPUTUSAN PEMBELIAN PAKET BULAN MADU BALI KEROBOKAN-SEMINYAK-PETITENGET DI BULANMADU.COM. *Jurnal Ilmu Manajemen Oikonomia*, *63*(1).

[18] Oliver, R. L., & Sharpe, J. W. E. (2014). *A Behavioral Perspective on the Consumer Second Edition Richard L Oliver* (2nd ed.). McGraw-Hill.

[19] Permana, K., Putra, Z. P., & Nugroho, A. (2021). Analisa Sentimen Pengunjung Hotel Dengan K-Nearest Neighbor Studi Kasus Hotel Pop! Surabaya. *Jurnal Ilmu Komputer Dan Bisnis (JIKB)*, *XII*(2), 159–166.

[20] Prakoso, P. A. (2017). *Front office praktis: administrasi dan prosedur kerja*. Yogyakarta: Gava Media.

[21] Rendrawan, G., & Ngr Yudha Martin Mahardika, A. (2020). Jenis Keluhan dan Cara Penanganan Keluhan di Hotel Holiday Inn Resort Baruna Bali. *JMPP*, *3*(1).

[22] Sarudin, R., Achmad Ismail, dan, Tanggal Bulan Tahun, D., & Tanggal Bulan Tahun, D. (2021). ANALISIS ONLINE REVIEW

TRIPADVISOR.COM TERHADAP MINAT PEMBELIAN PRODUK JASA AKOMODASI DI HOTEL MANHATTAN TRIPADVISOR.COM REVIEW ONLINE ANALYSIS ON THE INTEREST OF BUYING ACCOMMODATION SERVICES IN HOTEL MANHATTAN. *Jurnal Hospitality Dan Pariwisata*, *7*(4), 33–43. https://doi.org/10.30813/.v7i1.2634

[23] Setiawan, R. A., Estetikha, A. K. A., Nurharyanto, E. M. O., Asmara, Y., & Wahyudi, A. (2022). Analisis Sentimen Hotel di Nusa Tenggara Barat Menggunakan Algoritma SVM 149. *Jurnal Informasi Interaktif*, *7*(1), 149–155.

[24] Setiowati, Y., & Helen, A. (2018). KLASIFIKASI ANALISIS SENTIMEN MENGENAI HOTEL DI YOGYAKARTA. *Scan: Jurnal Teknologi Informasi Dan Komunikasi*, *13*(1). www.traveloka.com

[25] Suryadi, Ridho, A., & Murhaban. (2021). Analisis Sentimen Review Hotel Menggunakan Algoritma Naïve Bayes Classifier. *TECHSI - Jurnal Teknik Informatika*, *13*(2), 95–105. https://doi.org/10.29103/techsi.v13i2.5596

[26] Tika, Moh. P. (2005). *Metode Penelitian Geografi*. Jakarta: PT. Bumi Aksara.

[27] Wilson, T., Wiebe, J., & Hoffmann, P. (2005). *Recognizing Contextual Polarity in Phrase-Level Sentiment Analysis*. Association for Computational Linguistics. http://www.cs.pitt.edu/