ISSN 2085-4579

Sentiment Analysis of User Satisfaction Towards Sales Promotion of Gojek Application Service Using SVM

Calandra Alencia Haryani *[1], Gabrielle Florencia[2], Andree E. Widjaja[3], Hery[4], Ferry V. Ferdinand[5]
[1,2,3,4] Department of Information Systems, Faculty of Computer Science, Universitas Pelita Harapan, Tangerang, Indonesia
*Email: calandra.haryani@uph.edu
[5] Department of Mathematics, Faculty of Science and Technology, Universitas Pelita Harapan, Tangerang, Indonesia

Accepted on November 10th, 2023
Approved on November 28th, 2023

***Abstract*— Customer satisfaction is a crucial factor for the sustainability of a company, including tech companies like Gojek. This condition exists because customer satisfaction affects the continuation of customers purchasing or reusing the company's services. Promotions or offers are commonly applied strategies by companies to enhance attractiveness among competitors and improve customer satisfaction. Therefore, this research aims to understand customer sentiments towards promotions or offers provided by the Gojek company to develop better promotional strategies by increasing customer satisfaction. The data source for this research comes from user tweets on the social media platform Twitter which has now changed to "X". The method used to determine the sentiment of this data is the support vector machine.**

***Index Terms*—user satisfaction; promotion; sales; sentiment analysis; suport vector machine;**

I. INTRODUCTION

The rapid advancement of technology has brought about significant changes across various sectors of society, including the business world. One example is the emergence of companies that embrace technology concepts, such as Gojek. Gojek is a private company that harnesses the advancement of information and communication technology to establish an app-based online transportation service business. This company has been established and has pioneered online transportation in Indonesia since 2010 [1].

Several companies now offer online transportation services like Gojek. This condition is due to the convenience and efficiency of online transportation in connecting potential passengers with transportation drivers, leading to a booming trend of using online transportation apps to support daily life within the community [2]. Therefore, Gojek must implement the right competitive strategies to compete with other competitors in the online transportation market.

According to Putra and Seminari [3], the key to winning the competition is for companies to recognize, understand, and satisfy the desires and needs of their customers. By doing this, customers will not look to or switch to other competitors offering similar products and services, which increases the likelihood of repeat purchases. Furthermore, understanding customer satisfaction can also be a benchmark and foundation for improvement in the quality of products or services and company marketing strategies [4]. One factor that can influence customer satisfaction and act as an attraction for customers includes price related elements, such as discounts [5].

This research aims to analyze the sentiments of Gojek customers regarding sales promotions in the form of discounts offered by the company. The results of this study will assist Gojek in understanding customer satisfaction from a promotional or offer perspective and help the company implement better promotional strategies to stay competitive against other competitors. The sentiment is collected from Twitter, an interactive social media platform that allows its users to voice opinions or critiques on matters in real time [6]. It's important to note that this research was conducted before Twitter changed its name to "X" so the data still references the original "Twitter" name. The sentiment analysis will be performed using the Support Vector Machine (SVM) classification method. This method has been selected for its effectiveness in high-dimensional spaces, which is typical of text data. Additionally, it can identify intricate decision boundaries through kernel functions.

II. LITERATURE

A. Customer Satisfaction

Customer satisfaction is a customer's feeling that come out by comparing what they expected with what they received, be it feelings of disappointment or happiness. There are five aspects that influence

customer satisfaction: product quality, service quality, emotions, price, and cost. Customer satisfaction is closely related to customer loyalty, profit, and market share [5].

B. Sales Promotion

Sales promotion is one of the trend tools in current marketing practices. Sales promotions are effective in grabbing the attention of customers and target markets, and can encourage them to make purchases. Sales promotions provide benefits that impact customer evaluations of a company's products or services psychologically (utilitarian benefits), such as in terms of cost savings, usefulness, and convenience, and emotionally and intrinsically (hedonic benefits), such as entertainment and the feelings of satisfaction obtained. The utilitarian benefits of this sales promotion become one of the things that can help the company achieve customer satisfaction [7].

Sales promotions for marketing new products or services can be in the form of coupons, samples, and refills. Meanwhile, for marketing old products or services, the company can give discounts. A discount or price cut is a form of price reduction given by the seller to the buyer directly for a product or service in a certain period based on policies and objectives set by the seller [8].

C. Sentiment Analysis

Sentiment analysis is an analytical approach aimed at determining the subjectivity of a text [9][10][11]. Through this analysis, the sentiment expressed from a text can be identified as positive, negative, or neutral [12][13][14]. For this research, only positive and negative sentiments were identified.

D. Support Vector Machine

The Support Vector Machine (SVM) is one of the methods of supervised machine learning and can be defined as a binary classification algorithm. SVM is a probabilistic-based binary classifier, a classifier that makes predictions based on inputs that match predefined classes. Compared to other classification algorithms, SVM classifiers are one of the most popular because of their accuracy and speed, making them most suitable for analyzing and classifying tweets based on positive and negative polarities, providing more appropriate and significant output [15][16].

The main principle of SVM is to find the hyperplane, a line in two dimensions, or a flat plane in multiple dimensions, which is best for separating two types of classes in the input space. There are three types of SVM kernel functions used for mapping the initial dimension or lower dimension of a dataset to a new dimension or higher dimension, namely the linear kernel function (1), gaussian radial basic function (RBF) (2), and polynomial (3)[17].

$$\boldsymbol{K}(\mathrm{x},\mathrm{y}) = x \,.\, y \quad (1)$$

$$\boldsymbol{K}(x_i, x_j) = \exp\left(-\frac{-\|x_i - x_j\|^2}{2\sigma^2}\right) \quad (2)$$

$$\boldsymbol{K}(x_i, x_j) = \left(\left(x_i . x_j\right) + \mathrm{c}\right)^d \quad (3)$$

III. METHODOLOGY

Based on research goals and previous studies on sentiment analysis [10][11][12]. There are five stages in this research methodology that can be shown in Fig. 1. Initially, the research will focus on Data Collection, wherein relevant data will be accumulated, ensuring a comprehensive and representative analysis. After data accumulation, Data Labeling will be carried out, where each piece of data will be classified as positive and negative. This step is crucial, as the labeling quality will significantly influence the ensuing classification model's accuracy. The third phase, Data Preprocessing, involves refining the collected data by removing noise, cleaning, and other necessary adjustments, ensuring the data's readiness for efficient model training. The next phase involves Model Training and Evaluation, where the processed data will be trained utilizing the Support Vector Machine (SVM) classification algorithm using Linear Kernel Function. Upon model training, an evaluation will be conducted to ascertain its accuracy and readiness for sentiment analysis. The results of the research will then be visualized, providing a clear and comprehensible depiction of the sentiments discovered. To streamline this entire process and guarantee seamless integration across all stages, the author has opted to utilize the Anaconda application, explicitly leveraging the Jupyter Notebook platform for its flexibility and support for various libraries essential for such research.

Fig. 1. Methodology

A. Data Collection

During the Data Collection phase, data crawling is conducted to gather tweets from Gojek application users on the social media platform, Twitter, utilizing the Twitter API. The collected data comprises tweets where users have tagged Gojek's official Twitter account "@gojekindonesia" and contain keywords such as "promo","potongan","diskon" and/or "voucher". This process uses the Twitter API to collect data from Twitter. Once the tweet data is obtained, these datasets' files will be merged. Then, all the datasets will be checked, and any duplicate tweet data identified will be removed.

B. Data Labeling

During the Data Labeling phase, tweets are labeled to categorize them as either positive or negative sentiments. Tweets categorized as positive are labeled "1" and represent tweets expressing advantages or

customer satisfaction regarding Gojek's promotional features or sales promotions. Then, tweets categorized as negative are labeled "2" and represent tweets raising questions about customer issues with the sales promotion features, as well as statements concerning the shortcomings and customer dissatisfaction regarding Gojek's sales promotions. This labeling process utilized annotation techniques carried out by author.

C. Data Preprocessing

In the data preprocessing stage, the labeled tweet data will be transformed or cleaned from irrelevant data to facilitate and enhance the accuracy of the sentiment analysis in the subsequent stage [18]. This preprocessing phase is executed utilizing several available Python libraries. The preprocessing steps include case folding, filtering, normalization, stopword removal, and stemming shown in Fig. 2.

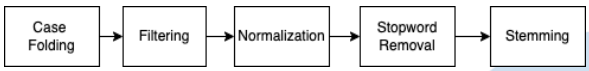

Fig. 2. Preprocessing Steps

- Case Folding

Case folding is the process of converting all letters in a dataset to lowercase. This is done to ensure that the algorithm treats different cases of the same word equally, such as "Minimum","minimum", and "MINIMUM" which should all be considered the same word.

- Filtering

Filtering is the process of cleaning text by removing components that are considered to have no meaning or relevance for sentiment analysis. These components include characters, punctuation, emoticons, and URLs. By filtering out these elements, the focus is placed on the significant parts of the text that contribute to understanding sentiment, thereby improving the quality of the analysis. The filtering process can significantly reduce noise in the data and help highlight the sentiment-bearing words.

- Normalization

Normalization is the process of converting words in the text into their standard form. In this study, the standard reference for the normalization process is the Indonesian Language Dictionary, and the author has created a list containing non-standard words along with their standard forms, which are stored in a glossary file at "normalisasi.csv" shown in Fig. 3.

```
def normalization(text):
    reader = csv.reader(open('normalisasi.csv', 'r'))
    d = {}
    for row in reader :
        k,v = row
        k = k.lower()
        v = v.lower()
        d[k] = v
    pat = re.compile(r"\b(%s)\b" % "|".join(d))
    text = pat.sub(lambda m: d.get(m.group()), text)
    return text
```

Fig. 3. Python Code for Normalization

- Stopword Removal

Stopword removal is the process of filtering out words that do not carry meaning in the text. These meaningless words include conjunctions, adverbs, affixes, and so on. This research utilizes the Python library "Sastrawi" (StopWordRemover) along with supplemental stopwords data acquired from various sources and through an analysis of tweet data. These additional stopwords are merged into a file named "more-stopwords.txt".

- Stemming

Stemming is the process of transforming words into their base form within a text [1] [19] [20]. The research uses the Python library "Sastrawi" (Stemmer) for the stemming process.

D. Model

This research data will be used to train the sentiment analysis model after the tweet data has been cleaned through the preprocessing stage. The sentiment analysis model training is carried out by classifying each word into positive or negative sentiment. Then, sentiment analysis will be conducted on each tweet containing the categorized words using the SVM classification algorithm. The research employs the SVM algorithm due to its robust capability in navigating the complex text data derived from Twitter [16] [17]. SVM stands out as it can sift through vast quantities of text data while pinpointing patterns with precision [15].

After the model training process is complete, the model is evaluated to determine its accuracy level in performing sentiment analysis. This research uses the Python libraries "NLTK" and "SKLEARN" to conduct the training and evaluation processes for this sentiment analysis model.

E. Visualization

After the sentiment analysis model has been created, the final step is to visualize the training data from the previous stages. The visualizations to be created will present the words that frequently appear in tweet data with negative sentiment to make improvements for negative sentiment. This visualization process will be conducted using Python libraries, and the results will be displayed as pie charts and word clouds.

IV. RESULT

This result part consist of data collection, data labeling, data preprocessing, model and evaluation, and visualition.

A. Data Collection

The research obtained 566 tweet data entries for this study shown in Fig. 4. The data collection period from November 12, 2022 to December 6, 2022.

0	halo @gojekindonesia apakah ini akun saya kena soft banned? Sekarang sudah tidak pernah dapet voucher goride/gofood dan tidak bisa claim kode promo. Mohon bantuannya supaya bisa nyaman dan aktif kembali menggunakan gojek. Terima kasih
1	@gojekindonesia dear gojek, ini akun aku ke suspend apa gmn si? Kok sering bgt ga dapet voucher goride\udd83d\ude2d\udd83d\ude2d
2	@gojekindonesia kenapa voucher cashback 100% 7000 goclub belum masuk ya? bulan kemarin juga gitu gak masuk \udd83e\udd72 masa bulan ini juga gak?! padahal juragan loh????? https://t.co/CmYW0eaiTj
3	@daesugua @gojekindonesia @gojekpromo Voucher gojek bisa dijual?
4	@gojekindonesia kok pasang promo di gofood tiba-tiba dibatalin mulu kenapa ya?
...	...
562	halo @gojekindonesia apakah akun saya kena soft banned? Sekarang sudah tidak pernah dapet voucher goride/gocar/gofood dan tidak bisa claim kode promo. Mohon bantuannya supaya bisa nyaman dan aktif kembali menggunakan gojek. Terima kasih
563	halo @gojekindonesia apakah ini akun saya kena soft banned? Saya tidak pernah dapet voucher goride/gocar dan tidak bisa claim kode promo. Mohon bantuannya supaya saya bisa nyaman dan aktif menggunakan gojek. Terima kasih. https://t.co/5fR0eyjbQk
564	halo @gojekindonesia, apakah akun gojek saya terkena softbanned? karena tidak mendapat voucher mingguan goride, gocar, gofood, voc cashback tiap bulan, dan jika memasukan kode voucher promo itu tidak bisa digunakan. tolong dibantu, terima kasih.
565	Halo admin @gojekindonesia, apakah akun saya kena soft banned? Karena gabisa klaim kode promo, ga dapet voucher, promo resto di gofood juga ga muncul. SUDAH SAYA INFOKAN NO HP & EMAIL AKUN SAYA MELALUI DM Tolong bantuannya agar bisa nyaman menggunakan gojek kembali, Terima kasih
566	Halo admin @gojekindonesia, apakah akun saya kena soft banned? Karena sudah jarang dapet gofood, dan kode promo yang diklaim juga tidak muncul di aplikasi. Tolong bantuannya ya agar bisa nyaman menggunakan gojek kembali, terima kasih banyak.

Fig. 4. Data Collection

B. Data Labelling

Data labelling was conducted manually from the collected tweet data, with a significant majority displaying negative sentiment. Out of 566 tweets, 49 were identified as having positive sentiment, while the remaining 517 tweets were categorized as having negative sentiment. Fig. 5 is an example of annotated tweet data labelled according to its sentiment.

Polaritas	Data Tweet
1	"hari-hari naik @gojekindonesia with voucher \udd83d\ude47\u200d\u2640\ufe0f https://t.co/GbXu3F7z6N"
1	"terbaik emang @gofoodindonesia @gojekindonesia selalu kash gua voucher ataupun hp. bangga gua sama tim gojek indonesia punya anak negeri. tak lupa terimakasih buat @waroengsteak udah ngebuat makanan ini menjadi enak \udd83d\udc4d\udd83d\udc4d\udd83d\udc4d https://t.co/DRFF4rHmq"
1	"@gojekindonesia AAAAAA SENENG DPT VOUCHER GO FOOD HUHU MAACI LOVE UUUUUUuddd3e\udef6\udd83e\udef6\udd83e\udef6\udd83e\udef6\udd83e\udef6\udd83e\udef6"
1	"Sering-sering lah buat promo cashback makasih loo. Sangat membantu anak kossan sekali ini heheh luv. .@gopayindonesia @gojekindonesia https://t.co/uIxyvRFFKm"
2	"@gojekindonesia kenapa voucher cashback 100% 7000 goclub belum masuk ya? bulan kemarin juga gitu gak masuk \udd83e\udd72 masa bulan ini juga gak?! padahal juragan loh????? https://t.co/CmYW0eaiTj"
2	"halo @gojekindonesia apakah ini akun saya kena soft banned? Sekarang sudah tidak pernah dapet voucher goride/gofood dan tidak bisa claim kode promo. Mohon bantuannya supaya bisa nyaman dan aktif kembali menggunakan gojek. Terima kasih"
2	"Makin pelit aja ngasi diskon @gojekindonesia https://t.co/xCBrvNSgUB"
2	"Ampas @gojekindonesia promo gopay payday mcd dibatalin. Jadi males deh"

Fig. 5. Python Code for Data Labelling

C. Data Preprocessing

The data preprocessing stages in this research consist of case folding, filtering, normalization, stopword removal, and stemming. Then, the labeled tweet dataset is divided into training and testing data, with a distribution ratio of 70:30. Fig. 6.,Fig. 7., Fig. 8., Fig. 9., Fig. 10., and Fig. 11, are examples of the results of several tweet data that have gone through each stage of preprocessing.

```
2|"Makin pelit aja ngasi diskon @gojekindonesia https://t.co/xCBrvNSgUB"
2|"Ampas @gojekindonesia promo gopay payday mcd dibatalin. Jadi males deh"
```

Fig. 6. Data Before Preprocessing

```
2|"makin pelit aja ngasi diskon @gojekindonesia https://t.co/xcbrvnsgub"
2|"ampas @gojekindonesia promo gopay payday mcd dibatalin. jadi males deh"
```

Fig. 7. Data After Case Folding

```
2|makin pelit aja ngasi diskon
2|ampas promo gopay payday mcd dibatalin jadi males deh
```

Fig. 8. Data After Filtering

```
2|semakin pelit saja memberi diskon
2|sampah promo gopay payday mcd dibatalkan jadi malas deh
```

Fig. 9. Data After Normalization

```
2|pelit memberi diskon
2|promo gopay payday mcd dibatalkan malas
```

Fig. 10. Data After Stopword Removal

```
2|pelit beri diskon
2|promo gopay payday mcd batal malas
```

Fig. 11. Data After Stemming

D. Model and Evaluation

After training the model using the SVM classification algorithm an accuracy rate of approximately 93% was achieved.

E. Visualization

From the tweets that were obtained, processed, and analyzed, the visualization results show that there are ten words commonly found in tweets with negative sentiment which are "promo", "voucher", "use", "minimum", "gofood", "goride", "code", "softbanned", "get", and "gocar". The the visualization result can be seen in Fig. 12.

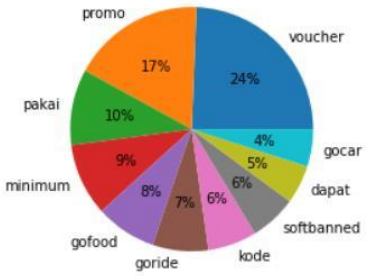

Fig. 12. The visualization results are presented as a pie chart and a word cloud.

From the obtained visualization, there are still numerous complaints about Gojek's sales promotions, particularly regarding policy, usage, and acquisition of promotions. The types of promotions most frequently discussed include those for GoFood, GoRide, and

GoCar services. To increase the efficiency of Gojek's promotions, it is essential to have transparent communication of the terms, simplify the redemption process for better usability, and put in place a strong feedback mechanism to address customer issues.

V. CONCLUSION AND RECOMMENDATION

Based on the sentiment analysis conducted on the promotional sales of Gojek's app services, it is concluded there are still numerous complaints indicating customer satisfaction still needs to be optimal. From the data obtained, the company needs to revisit its policies, usage processes, and the acquisition of promotions or discounts offered to customers and consider better promotional sales strategies for services like GoFood, GoRide, and GoCar.

Future research could add neutral sentiment category and expand the dictionary used for normalization and stopwords removal during the preprocessing phase. This would likely result in a more accurate sentiment analysis. Additionally, assessing sentiment based on service quality categories could offer deeper insights [21]. By categorizing sentiments into specific aspects of service, organizations can discern which areas are receiving negative feedback and require improvement. This approach allows for a targeted response to enhance overall customer satisfaction. It would also be valuable to explore and compare different classification algorithms in sentiment analysis to identify the most effective and precise algorithm for the scope of the research, especially in the context of service quality theory.

REFERENCES

[1] E.Dwiantoand and M.Sadikin,"Analisis sentimen transportasi online pada twitter menggunakan metode klasifikasi naïve bayes dan support vector machine," Jurnal Format, vol. 10, no. 1, pp. 94-100, 2021.

[2] G. N. Tejawulan, "Pengaruh orientasi pasar dan inovasi produk terhadap kinerja pemasaran (studi kasus gojek)," Journal of Economics and Business Aseanomics, vol. 6, no. 2, pp. 94-100, 2021.

[3] K. A. G. K. Putra and N. K. Seminari, "Kualitas produk, kualitas layanan, dan kewajaran harga berpengaruh terhadap kepuasan pelanggan the old champ café," E- Jurnal Manajemen Universitas Udayana, vol. 9, no. 10, pp. 3423-3442, 2020.

[4] I.K.DewiandA.Kusumawati,"Pengaruhdiskonterhadap keputusan pembelian dan kepuasan pelanggan bisnis online," Jurnal Administrasi Bisnis, vol. 56, no. 1, pp. 155- 163, 2018.

[5] J. Siska and T. Purba, "Pengaruh diskon dan kualitas pelayanan terhadap kepuasan pelanggan pada pt wellindo blast media di kota batam," eCo-Buss, vol. 5, no. 1, pp. 14- 25, 2022.

[6] S. Mandasari, B. H. Hayadi and R. Gunawan, "Analisis sentimen pengguna transportasi online terhadap layanan grab indonesia menggunakan multinomial naive bayes classifier," Jurnal Teknologi Sistem Informasi dan Sistem Komputer TGD, vol. 5, no. 2, pp. 118-126, 2022.

[7] S. K. Sinha and P. Verma, "Impact of sales promotion's benefits on perceived value: does product category moderate the results?," Journal of Retailing and Consumer Services, vol. 52, 2020.

[8] E. W. Putra, S. Kumadji and E. Yulianto, "Pengaruh diskon terhadap minat beli serta dampaknya pada keputusan pembelian," Jurnal Administrasi Bisnis, vol. 38, no. 2, pp. 184-193, 2016.

[9] M. Wongkar and A. Angdresey, "Sentiment analysis using naive bayes algorithm of the data crawler : twitter," International Conference on Computer Science, Information Technology, and Electrical Engineering (ICOMITEE), 2019.

[10] Cahyani, R. R., & Cahyani, R. (2020). Analisis Sentimen Pada Media Sosial Twitter Terhadap Tokoh Publik Peserta Pilpres 2019. *MATICS*, *12*(1), 79. https://doi.org/10.18860/mat.v12i1.8356

[11]. Pratama, S. F., Andrean, R., & Nugroho, A. (2019). Analisis Sentimen Twitter Debat Calon Presiden Indonesia Menggunakan Metode Fined-Grained Sentiment Analysis. *JOINTECS (Journal of Information Technology and Computer Science)*, *4*(2), 39. https://doi.org/10.31328/jointecs.v4i2.1004

[12] A. Erfina and Y. H. Putra, "Irony sentence detection techniques using fuzzy historical classifier," IOP Conference Series: Materials Science and Engineering, vol. 662, no. 6, 2019.

[13] Fitriyyah, S. N. J., Safriadi, N., & Pratama, E. E. (2019). Analisis Sentimen Calon Presiden Indonesia 2019 dari Media Sosial Twitter Menggunakan Metode Naive Bayes. *Jurnal Edukasi Dan Penelitian Informatika (JEPIN)*, *5*(3), 279. https://doi.org/10.26418/jp.v5i3.34368

[14] Andika, L. A., Azizah, P. A. N., & Respatiwulan, R. (2019). Analisis Sentimen Masyarakat terhadap Hasil Quick Count Pemilihan Presiden Indonesia 2019 pada Media Sosial Twitter Menggunakan Metode Naive Bayes Classifier. *Indonesian Journal of Applied Statistics*, *2*(1), 34. https://doi.org/10.13057/ijas.v2i1.29998

[15] A. Gupta, P. Tyagi, T. Choudhury and M. Shamoon, "Sentiment analysis using support vector machine," International Conference on contemporary Computing and Informatics (IC3I), pp. 49-53, 2019.

[16] P. Kumar, T. Choudhury and S. Rawat, "Analysis of various machine learning algorithms for enhanced opinion mining using twitter data streams," International Conference on Micro-Electronics and Telecommunication Engineering (ICMETE), pp. 265-270, 2016.

[17] M. N. Muttaqin and I. Kharisudin, "Analisis sentimen aplikasi gojek menggunakan support vector machine dan k nearest neighbor," UNNES Journal of Mathematics, vol. 10, no. 2, pp. 22-27, 2021.

[18] H. Parveen and S. Pandey, "Sentiment analysis on twitter data-set using naive bayes algortihm," 2nd International Conference on Applied and Theoretical Computing and Communication Technology (iCATccT), 2016.

[19] M. I. Petiwi, A. Triayudi and I. D. Sholihati, "Analisis sentimen gofood berdasarkan twitter menggunakan metode naïve bayes dan support vector machine," Jurnal Media Informatika Budidarma, vol. 6, no. 1, pp. 542-550, 2022.

[20] T. Rahman, F. E. M. Agustin and N. F. Rozy, "Normalization of unstructured indonesian tweet text for presidential candidates sentiment analysis," The 7th International Conference on Cyber and IT Service Management (CITSM), 2019.

[21] Haryani A., C., Tohari, H., Marhamah, & Nurrahman A., Y. (2018). Sentimen Analisis Kepuasan Pelanggan E-commerce Menggunakan Lexicon Classification dengan R. *Konferensi Nasional Sistem Informasi (KNSI)* , 189–196.