Journal of Student Research Exploration

https://shmpublisher.com/index.php/josre
p-ISSN 2746-7686 | e-ISSN 2746-0991

Optimising SVM models in text mining to see the sentiments and user complaints of DANA mobile application through play store reviews

Arell S. Biyantoro[1], Budi Prasetiyo[2]
[1, 2] Department of Computer Science, Universitas Negeri Semarang, Indonesia

Article Info

Article history:

Received July 9, 2024
Revised August 24, 2025
Accepted August 27, 2025

Keywords:

Application
Dana
SVM
Play store
Review

ABSTRACT

Dana is a mobile electronic wallet application available for download on Google Play Store. Users can rate and comment on this application directly through the review section on the platform. By utilizing these user reviews, research can be conducted to identify the main complaints experienced by Dana application users. This research uses Support Vector Machine (SVM) sentiment analysis to classify reviews and Latent Dirichlet Allocation (LDA) to map negative comment topics. LDA extracts several representative words or tokens that are grouped to form specific themes. The findings show that the most common sources of user complaints are related to transaction issues, premium features, and app updates. These insights can provide valuable input for developers to improve the overall quality and user experience of the Dana app.

1. Introduction

Dana is one of the well-known online payment service providers and e-wallet applications in Indonesia to help people in doing transactions [1]. This application offers convenience and quickness in accessing transaction-related needs such as transfers, payments, and investments so that they can have an impact on productivity, security of use, and well-being [2]. Even by using an electronic-based payment system such as Dana can minimise and slow down the inflationary path due to the uncontrolled circulation of banknotes [3]. With these advantages that Dana offers, they are able to become a digital wallet application that has a growth of around 4% in difficult times, COVID-19 [4].

In order to be used, the Dana application is published through the Play Store, one of Google's services that provides various digital content, such as digital books, films, and applications [5]. In addition to being able to publish Dana, the Google Play Store has a review feature that is very useful for the development of

applications and content in it. This review feature allows users of applications published on the Google Play Store to give criticism, suggestions, and compliments [6]. In this Google Play Store review feature, users can also provide their level of satisfaction through stars that have a scale of 1 to 5 [7].

One of the vital things in application development is user review of the application itself. Ratings can direct application development in a better direction, especially in fulfilling the needs of its users [8]. Rating-based assessment also allows users to evaluate in advance the application to be used according to their needs [9]. This is also very easy to do because the rating is clearly displayed on the same page where the application will be downloaded [9].

This review and rating feature will be used to find out the biggest complaints from Dana application users. By using the Support Vector Machine (SVM) method, it is expected to provide maximum results for research. SVM is one model that is quite effective for sentiment classification in high data dimensions [10].

2. Method

This research on the sentiments and biggest complaints of Dana application users is quantitative research from the scores given by users and qualitative from user comments. This qualitative and quantitative research itself is the concept of big data or data that is widespread and abundant in online media, marked by its very large size [11]. By conducting such mixed data collection, it can make the results more accurate and can present appropriate results [12]. A more complete flow explanation can be seen in the flow below.

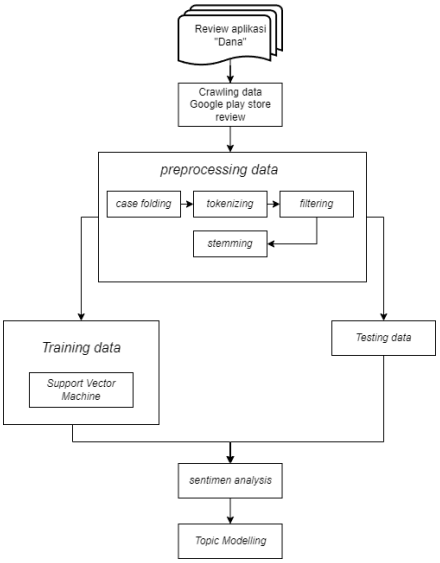

Figure 1. Research method

Data Crawling

The definition of data crawling is a technique or a way to retrieve information in the form of content or the like through the API provided on certain web pages [13].

In this case, researchers took Dana application review data from the PlayStore website. By using the Google-Play-Scraper API which can be used freely and easily through the Python language [14]. Utilising this, researchers can easily collect data about the assessments and ratings given to the Dana application by the users.

Preprocessing

Preprocessing is an elementary method commonly used in text data processing, especially in sentiment research [15]. This method is an important step that must be taken in the text mining process because it will greatly affect the accuracy and even distribution of data [16]. Preprocessing itself is divided into several commonly used steps, namely case folding, tokenising, filtering, and finally stemming [16].

The first step in preprocessing is Case folding which is the stage of changing each alphabet into lowercase letters so that each word in the data has a fixed meaning or is not case sensitive [17]. Then, there is tokenising which is a stage to separate each word in a whole sentence so that it becomes. The purpose of this tokenizing is to explore every word contained in a sentence in the form of text [18]. Next there is stopwords removal filtering which is part of preprocessing with the aim of removing words that are not important or not related to the research context. The stopwords removal/filtering process can be very helpful for researchers conducting sentiment classification research [19]. The last step of preprocessing is to perform stemming. stemming is a technique of changing each word back to its basic form [20].

Support Vector Machine (SVM) Modelling

As already mentioned, Support Vector Machine (SVM) is one of the classification models that is often relied upon by researchers to conduct sentiment classification research. This is inseparable from the Support Vector Machine (SVM) which is strong and can always provide good calculations [21]. Not only in classification, Support Vector Machine (SVM) is also suitable for regression and rigidity detection. Based on Support Vector Machine (SVM), a simple and traditional technique used to learn or know a pattern is to minimise empirical risk in optimising training set work. Support Vector Machine (SVM) performs pattern recognition on two classes of points. The trick is to determine the determining point of the training set [22].

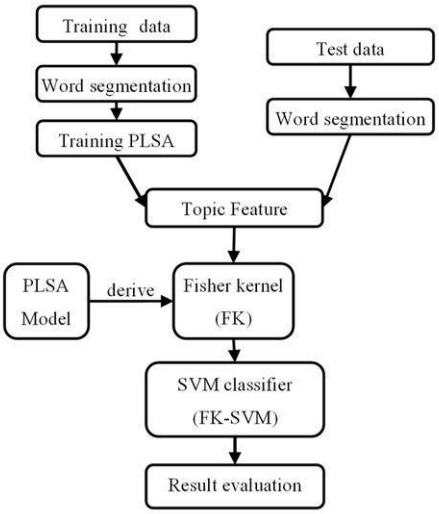

Figure 1. Support vector machine workflow [23]

Latent Dirichlet Allocation (LDA)

LDA is modelling to find topics in text data with a generative probabilistic model. The history of Latent Dirichlet Allocation (LDA) began in 2003 with the introduction of Blei, Ng, and Jordan. This model is based on the concept of documents being represented as random topics. A topic itself is a popular topic in the document and is characterised by a fairly even distribution of words [24].

So basically, Latent Dirichlet Allocation (LDA) is a model of theme selection based on the probability or intention of the word in a document. The higher the probability of a word, the higher the topic associated with that word [24]. LDA work flow can be seen in Figure 2.

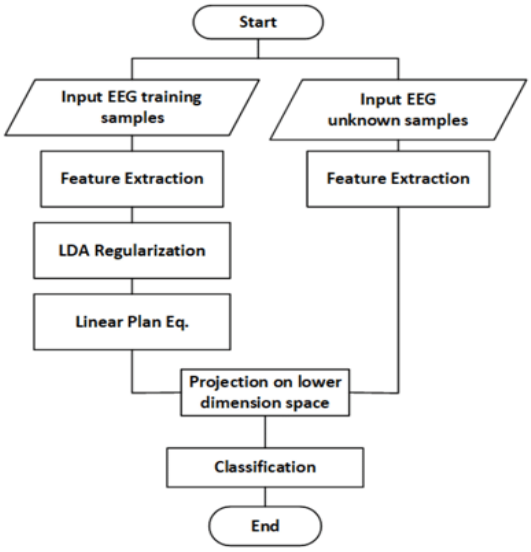

Figure 2. LDA workflow [25]

3. Results and Discussion

Through these methods and steps, researchers have implemented it into calculations using a Google tool called Colab. Colab is a Python-based Jupyter

notebook software that can be accessed online through a browser [26]. Then an API called google-play-scraper is used to retrieve around 350 data reviews of the Dana application with MOST_RELEVANT filtering or the most presenting comments on the Dana application itself.

Data Crawling

From the process of crawling 350 comment and rating data that best suits the Dana application, a dataset is obtained with reviewId for the identity of a comment, userName for the user's name, userImage for the user's profile, content for the comment written by the user, score for the number given by the user, thumbsUpCount for likes on the comment, reviewCreatedVersion for which version of the Dana app the comment was made on, at for when the comment was made, replyContent for replies to the comment, repliedAt for when the reply was made, and appVersion for what version of the Dana app the comment was made on.

After obtaining a complete dataset regarding the Dana application review, the researcher searched it again to only take important attributes in it. The important attributes taken are user comments (content) and the number of stars given by users (score). Then for the needs of sentiment calculation, researchers changed the score from a scale of 1-5 to Positive, Neutral, and Negative sentiments. The change is done automatically using the Python language and applying the if else concept. The calculation is negative comment sentiment if the score given by the user is 2 or 1, the comment sentiment will be neutral if the score given is 3, and positive comment sentiment if the score given is 4 or 5.

Table 1. Dana application review data

Index	Score	Sentiment	Comment
0	1	Negative	Penangan. Laporan lambat hanya bisa via chatbot tidak langsung customer service nya , komplain terkait saldo tidak bertambah setelah top up proses nya 1 bulan gak ada hasil di riwayat transaksi sudah berhasil namun saldo di aplikasi tidak bertambah
1	3	Neutral	Kok sekarang..makin kesini makin sering gangguan ya.. minta kirim dana dari teman..karna mau dipakai darurat status sukses,tp saldo nggak masuk. Dari sini bisa disimpulkan aplikasi dana belum bisa untuk diandalkan. Iya kalau gangguan cuma 30-60menit..kadang sampai 24jam. Saya tambah bintangnya jika aplikasi ini sudah benar" bagus/tidak sering gangguan. Terima kasih.
2	3	Neutral	Makasih min udah diperbaiki tetapi masih ada bug pembayaran saat kubuka e commerce di konfirmasi pembayaran diterakhirnya dan udah masuk kode virtual dengan benar tetapi kenapa masih ada bug lagi, saldo saya udah sesuai masuk jumlah syarat pembayaran tetapi kenapa transaksi ditolak dibilang jumlah kurang pembayaran

			kurang dari syarat minimal pembayaran, padahal saldo saya lebih dari syarat minimal (banyak)🙏🙏🙏🙏 Mohon diperhatikan dan ditindaklanjuti !!!! Saya mau beli barang jadi susah
3	5	Positive	Saran aja untuk pihak Dana 1.untuk lebih di ringanin aplikasi nya karena terlalu berat, sering terjadi bug karena berat. 2.kuota untuk gratis transfer antar bank nya setiap bulan agar diperbanyak lagi, atau kalau bisa gratis atau menggunakan admin bulanan seperti pada bank 3.Scan qris suka bug, saya harap bisa diperbaiki lagi agar tidak sering bug
4	4	Positive	Bintang 4 dulu ya, secara tampilan dll sudah oke tapi untuk klik" ke menu lain agak berat/lama padahal secara memori masih cukup luas jaringan bagus+wifi dan satu lagi ini yang penting menurut saya. Coba di buatkan pin atau kode lain pas mau masuk ke apk jangan langsung masuk ke beranda karna disini lah pintu masuk transaksi para pengguna. Terima kasih
...	...	...	...

Preprocessing

In the Preprocessing stage, researchers perform Case Folding, Tokenising, Filtering, and Stemming steps to produce data that is ready to be processed through the Support Vector Machine (SVM) model. Table 2 below will display some data results from the preprocessing that has been conducted.

Table 2. Preprocessing result

Index	Score	Sentiment	Comment
0	1	Negative	nang lapor lambat chatbot langsung customer service komplain kait saldo tambah proses hasil riwayat transaksi hasil saldo aplikasi tambah
1	3	Neutral	sekarangmakin kesini ganggu kirim dana temankarna pakai darurat status suksestp saldo nggak masuk simpul aplikasi dana andal ganggu 30-60menitkadang 4jam bintang aplikasi bagus tidak ganggu terima kasih
2	3	Neutral	makasih udah baik bayar buka commerce konfirmasi bayar akhir udah masuk kode virtual saldo udah sesuai masuk syarat bayar transaksi tolak bilang bayar syarat minimal bayar saldo syarat minimal mohon perhati ditindaklanjuti beli barang susah
3	5	Positive	saran dana 1untuk ringanin aplikasi berat berat kuota gratis transfer bank banyak gratis admin bulan bank 3scan qris suka harap baik

4	4	Positive	bintang tampil klik menu berat lama memori luas jaring baguswifi coba buat kode masuk langsung masuk beranda karna pintu masuk transaksi guna terima kasih
...	...	...	...

Implementation of Support Vector Machine (SVM)

In this section, Support Vector Machine (SVM) is used to examine the valuation of Dana Apps by users. Researchers apply split data by utilising sentiment classes that have been created. The dataset that has been preprocessed is then split into two parts, namely training data and also testing data with a division ratio of 9:1. The ratio means that the data is split into 90% training data and 10% as testing data.

After splitting the data, the data is then processed through a Support Vector Machine (SVM) to find out its accuracy. The result obtained through this sentiment classification modelling is 74.28% with a total distribution of results, namely 258 negative sentiments, 48 positive sentiments, and 44 neutral sentiments.

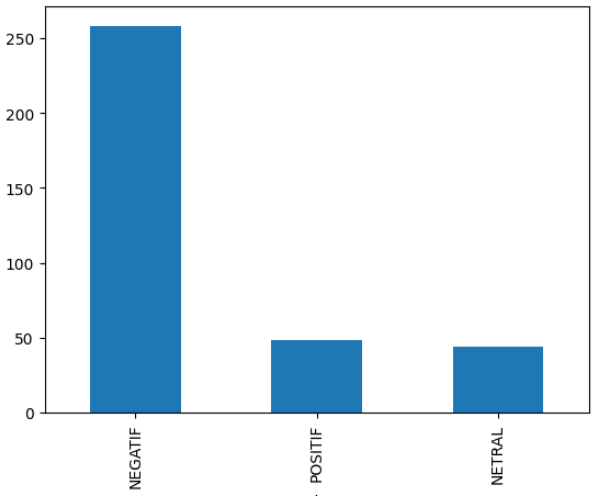

Figure 2. Sentiment distribution

Topic Modeling Latent Dirichlet Allocation (LDA)

From the dataset of Dana Application reviews that have been processed previously, researchers took content data or comments with negative sentiments to find what words appear most often in negative sentiments with a note excluding the word "Dana" which is the name of the application itself. Table 3 below shows the frequency of the top 10 tokens/words in negative reviews.

Table 3. Preprocessing result

Index	Word	Frequency
0	aplikasi	183
1	saldo	152
2	transaksi	141

3	uang	98
4	masuk	85
5	udah	69
6	akun	66
7	premium	64
8	tolong	63
9	upgrade	63

Finally, researchers used Latent Dirichlet Allocation (LDA) to estimate the top 5 themes regarding the biggest complaints of Dana App users through negative sentiment comments. The results are users complaining about the transaction, balance, premium and application performance features after the update. Table 4 below is the result of the top 5 themes in the negative sentiment comment data in the Dana Application.

Table 4. Latent dirichlet allocation

Index	Results of Coherence Values	Estimated Topic
0	0.038*"transaksi" + 0.038*"saldo" + 0.032*"uang" + 0.023*"aplikasi"	Transaction issues that impact the balance in the Dana account
1	0.024*"upgrade" + 0.016*"transaksi" + 0.015*"tolong" + 0.014*"udah"	Application update issue on transaction feature
2	0.032*"aplikasi" + 0.026*"saldo" + 0.019*"udah" + 0.017*"premium"	Relating to the premium features of the Application
3	0.030*"aplikasi" + 0.023*"saldo" + 0.018*"transaksi" + 0.017*"pakai"	Relating to Dana balances used for transactions
4	0.035*"aplikasi" + 0.021*"akun" + 0.017*"uang" + 0.014*"transaksi"	Account on Dana Application

4. Conclusion

The conclusions that can be drawn from the results of research on sentiment and complaints of Dana Application users are as below:

1. Through sentiment analysis of the Support Vector Machine (SVM) model to 350 review data, the Dana Application sentiment is dominantly negative with a total of 258 comments, followed by positive 48 and neutral 44 comments.
2. The sentiment accuracy obtained from the Support Vector Machine (SVM) model is 74.28% with a training and testing data division ratio of 9:1.
3. With Latent Dirichlet Allocation (LDA) modelling on comments that have negative sentiments, 5 main topics were found regarding user complaints about the Dana Application.
4. The most common negative topics are those that contain problems in the transaction section of the Dana Application, complaints about the balance in the account, premium features, to updates / updates to the Dana Application.

The advice needed in this research is the need to increase the number of data samples to be varied, so that it is expected to increase the accuracy of sentiment analysis.

REFERENCES

[1] A. Gunawan, N. Wahyuni, and V. Nursekha, "Kualitas Pelayanan Aplikasi Dana Terhadap Kepuasan Konsumen," *J. Integr. Syst.*, vol. 4, no. 2, pp. 181–198, Dec. 2021, doi: 10.28932/jis.v4i2.3861.

[2] R. Ardila and T. Sutabri, "Analisis Sudut Pandang Mahasiswa Dalam Performa dan Efisiensi Aplikasi Dana Dalam Kehidupan Sehari Hari Menggunakan Metode Deskriptif," *IJM Indones. J. Multidiscip.*, vol. 2, no. 1, 2024.

[3] L. Pransiska and F. A. Nofirda, "Dampak Ekonomi Digital terhadap Penggunaan Aplikasi DANA pada Gen Z di Kota Pekanbaru," *J. Pendidik. Tambusai*, vol. 7, no. 3 SE-Articles of Research, pp. 26604–26609, Nov. 2023, doi: 10.31004/jptam.v7i3.10901.

[4] F. A. Larasati, D. E. Ratnawati, and B. T. Hanggara, "Analisis Sentimen Ulasan Aplikasi Dana dengan Metode Random Forest," *J. Pengemb. Teknol. Inf. dan Ilmu Komput.*, vol. 6, no. 9 SE-, pp. 4305–4313, Sep. 2022, [Online]. Available: https://j-ptiik.ub.ac.id/index.php/j-ptiik/article/view/11562

[5] N. Agustina, D. H. Citra, W. Purnama, C. Nisa, and A. R. Kurnia, "Implementasi Algoritma Naive Bayes untuk Analisis Sentimen Ulasan Shopee pada Google Play Store," *MALCOM Indones. J. Mach. Learn. Comput. Sci.*, vol. 2, no. 1, pp. 47–54, Apr. 2022, doi: 10.57152/malcom.v2i1.195.

[6] S. Fransiska, R. Rianto, and A. I. Gufroni, "Sentiment Analysis Provider by.U on Google Play Store Reviews with TF-IDF and Support Vector Machine (SVM) Method," *Sci. J. Informatics*, vol. 7, no. 2, 2020.

[7] S. A. Harun, Hapsawati Taan, and Djoko Lesmana Radji, "Pengaruh Kualitas Pelayanan Terhadap Kepuasan Pelanggan Maxim Bentor Gorontalo (Studi Pada Mahasiswa Fakultas Ekonomi Ung)," *JEMSI (Jurnal Ekon. Manajemen, dan Akuntansi)*, vol. 10, no. 1, pp. 167–174, Feb. 2024, doi: 10.35870/jemsi.v10i1.1814.

[8] H. Setiawan and D. Novita, "Analisis Kepuasan Pengguna Aplikasi KAI Access Sebagai Media Pemesanan Tiket Kereta Api Menggunakan Metode EUCS," *J. Teknol. Sist. Inf.*, vol. 2, no. 2, pp. 162–175, Sep. 2021, doi: 10.35957/jtsi.v2i2.1375.

[9] R. Rofik, D. A. A. Pertiwi, and M. A. Muslim, "Improvement accuracy of gradient boosting in app rating prediction on google playstore," *J. Numer. Optim. Technol. Manag.*, vol. 1, no. 2, 2023.

[10] D. B. Santoso, A. Munna, and D. H. Untari Ningsih, "Improved playstore review sentiment classification accuracy with stacking ensemble," *J. Soft Comput. Explor.*, vol. 5, no. 1, pp. 38–45, Mar. 2024, doi: 10.52465/joscex.v5i1.247.
[11] S. E. Zaluchu, "Strategi Penelitian Kualitatif dan Kuantitatif Di Dalam Penelitian Agama," *Evang. J. Teol. Injili dan Pembin. Warga Jemaat*, vol. 4, no. 1, p. 28, Jan. 2020, doi: 10.46445/ejti.v4i1.167.
[12] F. F. Roji, N. G. Ginasta, Y. Cahyan, D. Rahayu, and D. Ramdani, "Review Analysis of SatuSehat Application Using Support Vector Machine and Latent Dirichlet Allocation Modeling," *J. RISTEC Res. Inf. Syst. Technol.*, vol. 4, no. 1, 2023.
[13] J. Budiarto, "Identifikasi Kebutuhan Masyarakat Nusa Tenggara Barat pada Pandemi Covid-19 di Media Sosial dengan Metode Crawling," *JTIM J. Teknol. Inf. dan Multimed.*, vol. 2, no. 4, pp. 244–250, Feb. 2021, doi: 10.35746/jtim.v2i4.119.
[14] R. Eliviani and D. D. Wazaumi, "Exploring Sentiment Trends: Deep Learning Analysis of Social Media Reviews on Google Play Store by Netizens," *Int. J. Adv. Data Inf. Syst.*, vol. 5, no. 1, pp. 62–70, Mar. 2024, doi: 10.59395/ijadis.v5i1.1318.
[15] S. Khomsah and Agus Sasmito Aribowo, "Text-Preprocessing Model Youtube Comments in Indonesian," *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, vol. 4, no. 4, pp. 648–654, Aug. 2020, doi: 10.29207/resti.v4i4.2035.
[16] A. E. Budiman and A. Widjaja, "Analisis Pengaruh Teks Preprocessing Terhadap Deteksi Plagiarisme Pada Dokumen Tugas Akhir," *J. Tek. Inform. dan Sist. Inf.*, vol. 6, no. 3, Dec. 2020, doi: 10.28932/jutisi.v6i3.2892.
[17] I. Nuritha, A. A. Arifiyanti, and V. P. Widartha, "Analysis of Public Perception on Organic Coffee through Text Mining Approach using Naïve Bayes Classifier," in *2018 2nd East Indonesia Conference on Computer and Information Technology (EIConCIT)*, IEEE, Nov. 2018, pp. 153–158. doi: 10.1109/EIConCIT.2018.8878572.
[18] T. Verma, R. Renu, and D. Gaur, "Tokenization and Filtering Process in RapidMiner," *Int. J. Appl. Inf. Syst.*, vol. 7, no. 2, pp. 16–18, Apr. 2014, doi: 10.5120/ijais14-451139.
[19] J. Kaur, "STOPWORDS REMOVAL AND ITS ALGORITHMS BASED ON DIFFERENT METHODS," *Int. J. Adv. Res. Comput. Sci.*, vol. 9, no. 5, pp. 81–88, Oct. 2018, doi: 10.26483/ijarcs.v9i5.6301.
[20] K. Kevin, M. Enjeli, and A. Wijaya, "Analisis Sentimen Pengunaaan Aplikasi Kinemaster Menggunakan Metode Naive Bayes," *J. Ilm. Comput. Sci.*, vol. 2, no. 2, pp. 89–98, Jan. 2024, doi: 10.58602/jics.v2i2.24.
[21] "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics," *Cancer Genomics Proteomics*, vol. 15, no. 1, Jan. 2018, doi: 10.21873/cgp.20063.
[22] S. Yue, P. Li, and P. Hao, "SVM classification:Its contents and challenges," *Appl. Math. J. Chinese Univ.*, vol. 18, no. 3, pp. 332–342, Sep. 2003, doi: 10.1007/s11766-003-0059-5.
[23] K.-X. Han, W. Chien, C.-C. Chiu, and Y.-T. Cheng, "Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter DataSet," *Appl. Sci.*, vol. 10, no. 3, p. 1125, Feb. 2020, doi: 10.3390/app10031125.
[24] H. Jelodar *et al.*, "Latent Dirichlet allocation (LDA) and topic modeling: models, applications, a survey," *Multimed. Tools Appl.*, vol. 78, no. 11, pp. 15169–15211, Jun. 2019, doi: 10.1007/s11042-018-6894-4.
[25] H. A. Abdulkareem and M. Z. Al-Faiz, "OFFLINE LINEAR DISCRIMINANT ANALYSIS CLASSFICATION OF TWO CLASS EEG SIGNALS," *Iraqi J. Inf. Commun. Technol.*, vol. 2, no. 3, pp. 1–10, Dec. 2019, doi: 10.31987/ijict.2.3.71.
[26] M. Canesche, L. Braganca, O. P. V. Neto, J. A. Nacif, and R. Ferreira, "Google Colab CAD4U: Hands-On Cloud Laboratories for Digital Design," in *2021 IEEE International Symposium on Circuits and Systems (ISCAS)*, IEEE, May 2021, pp. 1–5. doi: 10.1109/ISCAS51556.2021.9401151.