The Classification of Anxiety, Depression, and Stress on Facebook Users Using the Support Vector Machine

By Dedy Agung Prabowo

The Classification of Anxiety, Depression, and Stress on Facebook Users Using the Support Vector Machine

Tsania Maulida Wijiasih[1*], Rona Nisa Sofia Amriza[2], Dedy Agung Prabowo[3]
[1]Information Systems Study Program, Faculty of Informatics, Institut Teknologi Telkom Purwokerto
[2]Information Systems Study Program, Faculty of Informatics, Institut Teknologi Telkom Purwokerto
[3]Informatics Engineering Study Program, Faculty of Informatics, Institut Teknologi Telkom Purwokerto Email: [1]18103123@ittelkom-pwt.ac.id, [2] rona@ittelkom-pwt.ac.id, [3]dedy@ittelkom-pwt.ac.id

Abstract– Social media remains an essential platform for connecting people with friends, family, and the world around them. However, when events spread on social media are primarily negative, it will cause depression, anxiety, and stress that tend to increase. This study aims to classify depression, anxiety, and stress using the Support Vector Machine. The data in this study were obtained from active Facebook users using the Depression Anxiety Stress Scale (DASS 21) questionnaire. This study adopted the Knowledge Discover Database process. The result of this study is an evaluation of the performance of the Support Vector Machine classification of depression, anxiety, and stress. The accuracy of the Support Vector Machine in this study is 98.96%.

Keywords – Support Vector Machine, DASS 21, Depression, Anxiety, Stress, Facebook

I. INTRODUCTION

Social media is a computer technology that facilitates sharing of ideas, thoughts, and information through the internet network [1]. Data reported by Internet World Stats states that Indonesia ranks third in the world's most prominent use of social media, Facebook, reaching 176.5 million users in June 2021, which is equivalent to 63.9% of the total population of Indonesia [2]. The 2014 Indonesia Family Life Survey (IFLS) surveyed 22,423 individuals in Indonesia; the survey found that one standard deviation of social media use was associated with a 9% increase in CES
D scores (Center For Epidemiological Studies Depression Scale) [3]. It proves that social media has a negative impact on mental health [3]. Social media itself is seen as social support among users. Still, it can harm mental health, specifically, those who already have a significant degree of depression, anxiety, and stress [4].

Furthermore, Tang, Wang, and Norman (2013) found that activities on social media such as sharing, liking, messaging, and other activities increased stress. Moreover, excessive use of social media Facebook has become a severe source of stress because people often share all kinds of feeds, stories, and comments, from economics, politics, and social issues to personal problems [5]. Another thing is the desire to upload the best photos of yourself to get compliments or likes, and the pressure of bringing out the best of yourself can make the Facebook user feel anxious. In addition to anxiety, friends' achievements on Facebook are one of the factors which affect a person's mental health condition [6]. From the problems above, a classification is needed to classify active Facebook users affected by depression, anxiety, or stress to achieve a good life balance. Positive mental health can help individuals work productively and reach one's full potential.

The initial stage of this research is to collect data. In collecting data, this research was conducted by distributing questionnaires. The questionnaire itself is a research instrument consisting of a series of questions or other types of instructions that aim to collect information from a respondent. Several studies have used previous questionnaires to assess levels of depression, anxiety, and stress, such as the Perceived Stress Scale (PSS-10) [7], Subjective Units of Distress Scale (SUDS)[8], the Hamilton Rating Scale for Depression (HAM-D) [9], Hamilton Anxiety Rating Scale (HAM-A) and Depression, Anxiety, and Stress Scale (DASS 21) [10]. DASS-21 used in this study is because it has been used in several studies and has high consistency [11].

Several previous researchers used the Support Vector Machine to classify depression, anxiety, and stress in conducting the classification. Research by Zhang et al. predicts Social Anxiety Disorder using the Support Vector Machine, and the results of this study show an accuracy of 76.25%. It shows that the Support Vector Machine makes a good diagnosis of the potential for Social Anxiety Disorder [12]. Subsequent research was conducted by Frick et al. using the Support Vector Machine to classify Social Anxiety Disorder, and the result is an accuracy of 72.6% [13]. Another study conducted by Pantazatos et al. used the Support Vector Machine and got high accuracy results of 89% [14]. Therefore, this study identifies the classification of depression, anxiety, and stress on social media Facebook using a Support Vector Machine; this model determines the distance using a support vector, so the computing process becomes faster and produces high accuracy in classification.

II. RESEARCH METHODOLOGY

The object of this research is the social media Facebook, and the subject of this research is active Facebook users. The respondents of this study were obtained by distributing Google Forms via social media such as Facebook and Twitter. The questionnaire contains

JISA (Jurnal Informatika dan Sains) e-ISSN: 2614-8404 Vol. 05, No. 01, June 2022 p-ISSN:2776-3234

the questions from Depression Anxiety Stress Scale 21

(DASS21).
previous process, it continues with data mining,
Figure 1 shows the research stages, starting with the

study of literature and data collection to achieve data results

that can be processed in the Knowledge Discovery

Database and then evaluate the performance of the Support

Vector Machine.

After the data transformation has been carried out in the extracting

potentially valuable patterns. At this stage, the Support

Vector Machine is applied.

(e). Data Evaluation

At this stage, the researcher evaluates the performance

A. Literature Study

Literature studies are carried out by reading scientific sources such as books and journals related to the research Total = (∑ ������ ��������) × 2

topic or research question. This stage aims to find how this research relates to existing knowledge.

B. Data collection

Data collection in this study was carried out by distributing a Google Form containing a 2 Depression Anxiety and Stress Scale (DASS 21) questionnaire to active Facebook users using Convenience Sampling. The

Male Female

from the classification result. The evaluation in this research uses a confusion matrix. The output of this stage is accuracy, precision, f1, and recall.

III. Results And Discussion

The total respondents to the DASS 21 questionnaire were 193 respondents with, 67 male and 126 female. It can be seen in the image below:

GENDER

Figure1. Research Stages

Figure 2. Gender

questionnaire was shared on several social media

platforms, such as Facebook and Twitter.

C.KDD

Most respondents belonged to the age group of 20-30

years, 131 people, followed by 13-20 years, 41 people, 40-

50 years, 11 people, then 50 years, 11 people, and 30-40 years total of six people. It can be seen in the image

below:

The Knowledge Discovery Database in this in the KDD process research transforms data into valuable 6owledge. The context of this research is the classification of depression, anxiety, and stress in Facebook users. The stages

AGE

are explained below. (a). Data Selection

13-20y 20-30y 30-40y 40-50y>50y

The researcher selects the data for the classification process at this data selection stage. The data used comes from Depression Anxiety Stress Scale 21 questionnaire. However, this data is not in accordance with the classification process,

so the researcher needs to select the appropriate data.

(b). Data Preprocessing

In this data selection stage, noise or irrelevant data is removed from the previous data collection. This stage is

Figure 3. Age

necessary so that there is no duplication of data, inconsistent data, or correcting errors in the data. The results of the DASS 21 questionnaire are then labeled using the formulation below:

The total value of each item calculates by performing an addition to all of the sub-items and then multiple by two. After the total value of each item is obtained, the next step is to compare each item's value, and the highest value of the item is chosen to be a label.

(c). Data Transformation

After cleaning the data, then continued with the Data Transformation stage. In this stage, we change the data format, structure, or value into the form required in the data mining process.

(d). Data Mining

Respondents with jobs as students or college students dominate in this study. It can be seen in the image below:

Total depression [1] = (∑������ �������� ��������������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

Total Anxiety [1] = (∑������ �������� �����������������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

Total Stress [1] = (∑������ �������� ����������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

PROFESSION

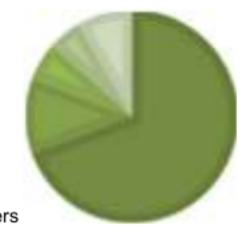

Figure 4. Profession

D4 2 0 1
D5 2 1 1
D6 2 1 1
D7 2 0 1
A1 2 1 0
Anxiety
A2 2 0 0
A3 3 1 1
A4 2 1 1
A5 2 1 1
A6 3 1 1
A7 2 0 1
S1 2 1 0
Stress
S2 2 1 1
S3 2 1 1
S4 3 0 1 S5 3 0 1
S6 3 0 1

Total depression [2] = (∑������ �������� ��������������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

Total Anxiety [2] = (∑������ �������� �����������������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

Total Stress [2] = (∑������ �������� ����������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

The domicile of the respondents is very diverse, from the islands of Java, Sumatra, and Kalimantan to Sulawesi.

DOMICILE

Total depression [3] = (∑������ �������� ��������������) × 2 (��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

Jawa Tengah Jawa Barat

S7 2 0 1 Total Anxiety [3] = (∑������ ��������
������������������) × 2

Jawa Timur Banten

CLASS	1	2	3

Sumatera Utara ~~Sumatera Barat~~

(��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

Sumatera Selatan Kalimantan Selatan

Total Stress [3] = (∑������ �������� ����������) × 2

Sulawesi Selatan Sulawesi Utara

(��1 + ��2 + ��3 + ��4 + ��5 + ��6 + ��7)× 2

DIY DKI Jakarta

= 28

=32

=34

Label[1] = Stress

Figure 5. Domicile

=8

The data used is a DASS 21-question instrument. There

= 10

are 21 questions, with 7 for depression questions, 7 for anxiety questions, and 7 for stress questions. The total number of respondents is 261, but only 193 can be

=6

categorized as users who experience depression, anxiety, and stress

Label[2] = Anxiety

The following process is labeling. Labeling is done on the respondent's data obtained from the previous process. Labeling was performed using a DASS score of 21.

=16

Table1. Sample of Respondent Data

= 10

Mental Illness	Items	Respondent 1	Respondent 2	Respondent 3
	D1	2	1	2
	D2	2	1	1
Depression	D3	2	0	1

= 12 Label[3] = Depression

Calculation of each depression, anxiety, and stress item was carried out using a DASS score of 21. Furthermore, each total mental illness was multiplied by two, and then compared the results of the calculation of each item. Respondent 1 was labeled depression because the calculated DASS 21 score for depression was more significant than the calculated DASS 21 score for anxiety and stress. Respondent 2 was labeled anxiety, and

������������

������������������

=����

���� + ����× 100%

=55

~~55 + 0~~ × 100% = 100%

��������������������

� ������������ =����

���� + ����× 100%

respondent three was labeled depression. The calculation was carried out on all 193-respondent data.

= 94.44%

������������

������������

After the data of 193 respondents were labeled depression, anxiety, and stress, it was continued by selecting the data to be used for processing in data mining.

= 98.21%

~~55 + 1~~ × 100%

Name, email address, gender, age, occupation, and domicile data on the questionnaire results were deleted. The label data is then transformed from numeric to categorical. Furthermore, the data is processed using the Support Vector Machine to produce a classification of depression, anxiety, and stress. Based on data of 193 respondents who have been tested, results of the calculation of f1, precision, recall, and accuracy are obtained. The Support Vector Machine model produces an accuracy of 98.96%, F1 is 95.75%, precision is 99.15%, and recall is 97.26% (Table 2).

Table2. Result

�����������������

��� ��1 = 2

×��������

��������

�� ×

��������

����

��������

��������

�� +

��������

Classification Mental illness
Accuracy F1 Precision Recall

SVM	Depres sion Anxiety Stress	98.96%	95.75 %	99.15%	97.26 %

The result of each item was evaluated using a confusion matrix. The mathematic formulation for accuracy, precision, recall, and f1 for each item score is defined below:

��������������������

���������������� = ���� + ����

���� + ���� + ���� + ���� × 100%

= 118 + 73

~~118 + 1 + 1 + 73~~ × 100%

=118

~~118 + 1~~ × 100% = 99.15%

��������������

������������ =����

���� + ����× 100% =17

=����

~~17 + 1~~ × 100%

���� + ����× 100% =55

����

= 2 ×99.15 × 99.15

~~99.15 + 99.15~~ = 99.15%

�������������� ��1 = 2

×������������������ ×

������������

������������������ +

������������

= 2 ×89.47 × 97.44

~~89.47 + 97.44~~ = 93.28%

������������ ��1 = 2

×������������������ ×

������������

������������������

+ ������������

= 2 ×91.66 × 98.21

~~91.66 + 98.21~~ = 94.82%

Where TN (True Negative) is a negative data with true value, FP (False Positive) is negative data that identifies as positive data. For the more TP (True Positive) have actual

= 98.96%

��������������

���������������� = ���� + ����

���� + ���� + ���� + ����× 100%

= 17 + 173

~~17 + 173 + 2 + 1~~ × 100%

= 98.44%

������������

���������������� = ���� + ����

���� + ���� + ���� + ���� × 100%

= 55 + 137

$\frac{}{55 + 137 + 0 + 1} \times 100\%$
$= 99.48\%$

��������������������

������������������

=����

���� + ����× 100%

=118

$\frac{}{118 + 1} \times 100\% = 99.15\%$

��������������

������������������

=����

���� + ���� × 100%

=117

$\frac{}{117 + 2} \times 100\% = 98,31\%$

value, and FN (False Negative) is a positive data identified as a negative data.

Based on the Support Vector Machine classification method, the result shows that depression accuracy is 98.96%, which explains that 118 instances labeled as depression have a correct value. Based on the Support Vector Machine classification method, the result shows that depression accuracy is 98.44%, which explains that 17 instances labeled as depression have a correct value. Based on the Support Vector Machine classification method, the result shows that depression accuracy is 99.48%, it explains that 55 of instances labeled as depression have a correct value.

The second confusion matrix score is precision. The one instance cannot be labeled as depression; it shows that the precision of depression is 99.15%. Then, two instances cannot be labeled as anxiety; it indicates that anxiety's precision is 89.47%. All of the instances labeled as stress show that the precision of stress is 100%.

The recall value for depression is 99.15% because one instance didn't classify as depression. Next, recall's anxiety is 94.44% because one instance didn't classify as anxiety. Finally, the recall value for stress was 98.21% because one instance wasn't classified as stress.

The last confusion matrix is F1. 99.15% for depression, 93.28% for anxiety and 94.82% for stress.

IV. Conclusion

From the test results, the Support Vector Machine model has high accuracy because it has advantages such as determining the distance using a support vector to make the

computing process faster [15]. The Support Vector Machine creates a decision function or hyperplane that differentiate between categories. The resulting decision function or hyperplane will be used to predict a predetermined class, so the classification accuracy is high [16]. The Support Vector Machine method is the best method for classifying depression, anxiety, and stress in Facebook users. It was shown with 98.96% accuracy. Furthermore, this research can be developed by adding other data mining methods such as naïve Bayes, random forest, and decision three to see the comparative performance of the model. Furthermore, other social media use can be subject to future research, for example, Twitter users, Instagram users, and YouTube users.

References

[1] M. Dollarhide, “Social Media Definition,” Investopedia, 2021.
[2] Internet World Stats, “Internet World Stats,” Internet World Stats, 2015.
[3] S. Sujarwoto, G. Tampubolon, and A. C. Pierewan, “A Tool to Help or Harm? Online Social Media Use and Adult Mental Health in Indonesia,” Int. J. Ment. Health Addict., vol. 17, no. 4, pp. 1076–1093, 2019, doi: 10.1007/s11469-019-00069-2.
[4] P. D. Lauren Reining, M.A., Michelle Drouin, “College Students in Distress: Can Social Media be a Source of Social Support?,” Park. Heal. Res. Repos., 2018.
[5] L. Weng and F. Menczer, “Topicality and impact in social media: Diverse messages, focused messengers,” PLoS One, vol. 10, no. 2, pp. 1–17, 2015, doi: 10.1371/journal.pone.0118410.
[6] S. Budury, A. Fitriasari, and K. -, “Penggunaan Media Sosial Terhadap Kejadian Depresi, Kecemasan Dan Stres Pada Mahasiswa,” 2019. doi: 10.36376/bmj.v6i2.87.
[7] E. H. Lee, “Review of the psychometric evidence of the perceived stress scale,” Asian Nurs. Res. (Korean. Soc. Nurs. Sci)., vol. 6, no. 4, pp. 121–127, 2012, doi: 10.1016/j.anr.2012.08.004.
[8] L. Horwitz, “Book Reviews: The Practice of Supportive Psychotherapy.,” J. Am. Psychoanal. Assoc., vol. 36, no. 1, pp. 197–199, 1988, doi: 10.1177/000306518803600115.
[9] D. Carrozzino, C. Patierno, G. A. Fava, and J. Guidi, “The hamilton rating scales for depression: A critical review of clinimetric properties of different versions,” Psychother. Psychosom., vol. 89, no. 3, pp. 133–150, 2020, doi: 10.1159/000506879.
[10] M. HAMILTON, “the Assessment of Anxiety States By Rating,” Br. J. Med. Psychol., vol. 32, no. 1, pp. 50–55, 1959, doi: 10.1111/j.2044-8341.1959.tb00467.x.
[11] S. Verma and A. Mishra, “Depression, anxiety, and stress and socio-demographic correlates among general Indian public during COVID-19,” Int. J. Soc. Psychiatry, vol. 66, no. 8, pp. 756–762, 2020, doi: 10.1177/0020764020934508.
[12] W. Zhang et al., “Diagnostic prediction for social anxiety disorder via multivariate pattern analysis of the regional homogeneity,” Biomed Res. Int., vol. 2015, 2015, doi: 10.1155/2015/763965.
[13] A. Frick et al., “Classifying social anxiety disorder using multivoxel pattern analyses of brain function and structure,” Behav. Brain Res., vol. 259, pp. 330–335, 2014, doi: 10.1016/j.bbr.2013.11.003.
[14] S. P. Pantazatos, A. Talati, F. R. Schneier, and J. Hirsch, “Reduced anterior temporal and hippocampal functional connectivity during face processing discriminates individuals with social anxiety disorder from healthy controls and panic disorder, and increases following treatment,”

Neuropsychopharmacology, vol. 39, no. 2, pp. 425–434, 2014, doi: 10.1038/npp.2013.211.
[15] V. Kecman, “Support Vector Machines – An Introduction 1 Basics of Learning from Data,” StudFuzz, vol. 177, pp. 1–47, 2005.
[16] V. N. Vapnik, Statistics for Engineering and Information Science Springer Science+Business Media, LLC. 2000.

The Classification of Anxiety, Depression, and Stress on Facebook Users Using the Support Vector Machine

ORIGINALITY REPORT

10%

SIMILARITY INDEX

PRIMARY SOURCES

1	Mengqi Xing, Jacklynn M. Fitzgerald, Heide Klumpp. "Classification of Social Anxiety Disorder With Support Vector Machine Analysis Using Neural Correlates of Social Signals of Threat", Frontiers in Psychiatry, 2020 Crossref	45 words — 1%
2	Ana Opanković, Milan Latas, Ivan Ristić, Stefan Jerotić, Zoran Bukumirić, Nikola Lalović, Srđan Milovanović. "Predictors of depression, anxiety and stress during the first wave of the covid-19 pandemic: The results of an online survey in Serbia", Engrami, 2021 Crossref	40 words — 1%
3	Samsudin Samsudin, Riri Syafitri Lubis. "Development Of Alumni Portal Application Based Android", Sinkron, 2022 Crossref	36 words — 1%
4	Darsol Seok, Reza Tadayonnejad, Wan-wa Wong, Joseph O'Neill, Jeff Cockburn, Ausaf A. Bari, John P. O'Doherty, Jamie D. Feusner. "Neurocircuit dynamics of arbitration between decision-making strategies across obsessive-compulsive and related disorders", NeuroImage: Clinical, 2022 Crossref	30 words — 1%

5 Atika Ratna Dewi, Ridho Ananda, Utti Marina Rifanti. "DYNAMIC ANALYSIS OF THE COVID-19 MODEL WITH ISOLATION FACTORS", BAREKENG: Jurnal Ilmu Matematika dan Terapan, 2022 — 29 words — 1%
Crossref

6 Warawut Narkbunnum, Kittipol Wisaeng. "Prediction of Depression for Undergraduate Students Based on Imbalanced Data by Using Data Mining Techniques", Applied System Innovation, 2022 — 24 words — 1%
Crossref

7 Ghanim Kanugrahan, Alfan Farizki Wicaksono. "Sentiment Analysis of Face-to-face Learning during Covid-19 Pandemic using Twitter Data", 2021 8th International Conference on Advanced Informatics: Concepts, Theory and Applications (ICAICTA), 2021 — 20 words — 1%
Crossref

8 Faisal Dharma Adhinata, Diovianto Putra Rakhmadani. "Prediction of Covid-19 Daily Case in Indonesia Using Long Short Term Memory Method", Teknika, 2021 — 17 words — 1%
Crossref

9 Jeitler, Michael, Stefan Brunnhuber, Larissa Meier, Rainer Lüdtke, Arndt Büssing, Christian Kessler, and Andreas Michalsen. "Effectiveness of Jyoti Meditation for Patients With Chronic Neck Pain and Psychological Distress—A Randomized Controlled ClinicalTrial", Journal of Pain, 2015. — 16 words — 1%
Crossref

10 Kartik Singhai, Mukesh Kumar Swami, Naresh Nebhinani, Ashu Rastogi, Edward Jude. "Psychological adaptive difficulties and their management during COVID-19 pandemic in people with diabetes mellitus", — 11 words — < 1%

Diabetes & Metabolic Syndrome: Clinical Research & Reviews, 2020
Crossref

11	Itamar Kahn, Jessica R. Andrews-Hanna, Justin L. Vincent, Abraham Z. Snyder, Randy L. Buckner. "Distinct Cortical Anatomy Linked to Subregions of the Medial Temporal Lobe Revealed by Intrinsic Functional Connectivity", Journal of Neurophysiology, 2008 Crossref	10 words — < 1%
12	Wenjing Zhang, Xun Yang, Su Lui, Yajing Meng, Li Yao, Yuan Xiao, Wei Deng, Wei Zhang, Qiyong Gong. "Diagnostic Prediction for Social Anxiety Disorder via Multivariate Pattern Analysis of the Regional Homogeneity", BioMed Research International, 2015 Crossref	10 words — < 1%
13	Zhijun Yao, Mei Liao, Tao Hu, Zhe Zhang, Yu Zhao, Fang Zheng, Jürg Gutknecht, Dennis Majoe, Bin Hu, Lingjiang Li. "An Effective Method to Identify Adolescent Generalized Anxiety Disorder by Temporal Features of Dynamic Functional Connectivity", Frontiers in Human Neuroscience, 2017 Crossref	9 words — < 1%
14	Emily A. Boeke, Avram J. Holmes, Elizabeth A. Phelps. "Toward Robust Anxiety Biomarkers: A Machine Learning Approach in a Large-Scale Sample", Biological Psychiatry: Cognitive Neuroscience and Neuroimaging, 2019 Crossref	8 words — < 1%

EXCLUDE QUOTES	OFF	EXCLUDE SOURCES	OFF
EXCLUDE BIBLIOGRAPHY	OFF	EXCLUDE MATCHES	OFF