

DAPA CA-2 Report

by Yash Bhargava

Submission date: 29-Apr-2024 08:44PM (UTC+0100)

Submission ID: 2365926056

File name: 20953_Yash_Bhargava_DAPA_CA-2_Report_67300_1641009119.pdf (1.5M)

Word count: 4041

Character count: 22101

1
National College of Ireland

Project Submission Sheet

Student Name: Prajwal Keshav Kongi, Prajwal Shashidhara, Yash Bhargava
Student ID: 22205314, 22209077, 22220861
Programme: MSc in Data Analytics **Year:** 2023 -2024
Module: Domain Application of Predictive Analytics
Lecturer: 29 Apr 2024 - 23:00
Submission Due Date:
Project Title: Predictive Analytics for Stress Detection: Using NLP Techniques on Human Stress Dataset
Word Count: 2899

I hereby certify that the information contained in this (my submission) is information pertaining to research I conducted for this project. All information other than my own contribution will be fully referenced and listed in the relevant bibliography section at the rear of the project.

ALL internet material must be referenced in the references section. Students are encouraged to use the Harvard Referencing Standard supplied by the Library. To use other author's written or electronic work is illegal (plagiarism) and may result in disciplinary action. Students may be required to undergo a viva (oral examination) if there is suspicion about the validity of their submitted work.

Signature: Prajwal Keshav Kongi, Prajwal Shashidhara, Yash Bhargava

Date: 29 Apr 2024

PLEASE READ THE FOLLOWING INSTRUCTIONS:

1. Please attach a completed copy of this sheet to each project (including multiple copies).
2. Projects should be submitted to your Programme Coordinator.
3. **You must ensure that you retain a HARD COPY of ALL projects**, both for your own reference and in case a project is lost or mislaid. It is not sufficient to keep a copy on computer. Please do not bind projects or place in covers unless specifically requested.
4. You must ensure that all projects are submitted to your Programme Coordinator on or before the required submission date. **Late submissions will incur penalties.**
5. All projects must be submitted and passed in order to successfully complete the year. **Any project/assignment not submitted will be marked as a fail.**

Office Use Only

Signature:	
Date:	
Penalty Applied (if applicable):	

AI Acknowledgement Supplement

Domain Application of Predictive Analytics

Predictive Analytics for Stress Detection: Using NLP Techniques on Human Stress Dataset

Your Name/Student Number	Course	Date
Prajwal Keshav Kongi /22205314, Prajwal Shashidhara /22209077, Yash Bhargava / 22220861	MSc in Data Analytics	29 Apr 2024

1 This section is a supplement to the main assignment, to be used if AI was used in any capacity in the creation of your assignment; if you have queries about how to do this, please contact your lecturer. For an example of how to fill these sections out, please click [here](#).

AI Acknowledgment

This section acknowledges the AI tools that were utilized in the process of completing this assignment.

Tool Name	Brief Description	Link to tool
Grammarly	Tool for checking basic grammatical errors, misspellings, and punctuations.	https://app.grammarly.com/

Description of AI Usage

This section provides a more detailed description of how the AI tools were used in the assignment. It includes information about the prompts given to the AI tool, the responses received, and how these responses were utilized or modified in the assignment. **One table should be used for each tool used.**

Grammarly	
Grammatical errors check, spellings check and punctuations	
[Insert Sample prompt]	[Insert Sample response]

Evidence of AI Usage

This section includes evidence of significant prompts and responses used or generated through the AI tool. It should provide a clear understanding of the extent to which the AI tool was used in the assignment. Evidence may be attached via screenshots or text.

Additional Evidence

Predictive Analytics for Stress Detection: Using NLP Techniques on Human Stress Dataset

1st Prajwal Keshav Kongi

Dept. of Computing

National College of Ireland

Dublin, Ireland

x22205314@student.ncirl.ie

2nd Prajwal Shashidhara

Dept. of Computing

National College of Ireland

Dublin, Ireland

x22209077@student.ncirl.ie

3rd Yash Bhargava

Dept. of Computing

National College of Ireland

Dublin, Ireland

x22220861@student.ncirl.ie

Abstract—Almost all researchers do sentiment analysis for the prediction of the stress of an individual in such a way that it can be detected at an early stage and diagnosed on time. Most people today use social media to express themselves [8] social media has become a part of their day-to-day life. Stress can have a positive as well as negative impact on an individual's life as it can cause many health-related issues such as anxiety, depression, obesity etc. So, it is essential to determine whether an individual has stress at an early stage. For that, we have used machine learning models to build a predictive model for the detection of stress [21] and on the key features and texts from Reddit by analysing the text using Natural Language Processing (NLP) for text processing and then further sending the cleaned text to the predictive model Random Forest for training so that the model can detect stress more precisely and effectively which can benefit many health organization and may contribute profitably to the business.

Index Terms—Reddit Human Stress Dataset, Natural Language Processing, Machine Learning

I. INTRODUCTION

Reddit is one of the most used and well-known online communities with tons of users who express their thoughts and emotions, share some content and build communities on many topics. Reddit's different subreddit structure which consists of various communities over different themes or interests can be used as a resource for human speech and content that can further be used for different tasks i.e. sentiment analysis, predictive analytics and model building. This generation spends most of their time on social media as they want to be socially active to stay updated. But sometimes it can also be a major concern [24] stress in an individual's life as most times individual uses social media such as Facebook, Instagram, X (formerly Twitter) and Reddit to express their emotions and thoughts on various topics. Stress detection is essential as it can be a major concern for many health-related issues such as anxiety, depression, cancer and many other severe health-related problems such that early detection and diagnosis are necessary. The ability to predict an individual's stress is crucial for promoting the mental health of the population. Stress can negatively impact the ability to think and capacity to make decisions of an individual which can also affect the business organization. So, in this research project, we will make use of predictive analytics techniques to detect the stress

of an individual which will further assist mental health organizations, and healthcare support systems to make informed business-driven decisions. In this study, we will make use of Natural Language Processing (NLP) and machine learning to detect whether an individual is stressed or not stressed by analysing their comments on Reddit which makes use of the Human Stress dataset which was collected from Kaggle and provides key business insights by building a predictive model using different machine learning techniques as this model will further provide great benefits to many healthcare institutions, corporate businesses. For instance, in the healthcare sector, this predictive model can be used for the early detection of stress so that it can be diagnosed on time, in corporate businesses, this model can be used by organisations to detect the stress of a worker and can manage the workload or provide medical assistance to the worker who is having higher risks of health-related issues.

II. HYPOTHESIS

4
Null Hypothesis (H0): There is no relationship between the length of the subreddit posts and the accuracy of stress prediction.

Alternate Hypothesis (H1): There is a relationship between the length of the subreddit posts and the accuracy of the stress prediction.

III. LITERATURE REVIEW

One of the research was done by Tanya Nijhawan in 2022 [1] in which they did sentiment and emotion analysis for the detection of stress by making use of posts and comments on social networking platforms by an individual [17] by analysing the tweets using machine learning and deep learning algorithms and Bidirectional Encoder Representations from Transformers (BERT) model for sentiment classification. Another research was conducted by Sunita Sahu in 2021 [2] in which they detected the emotions of the employees by analysing their facial expressions and monthly reviews on the company employee's portal as input and classified the seven different classes of emotions using machine learning techniques in which they have used SVM, Local Binary Pattern Histogram (LBPH) algorithm and CNN. Xiaohui Tao did another research [3] on

Twitter analysis for detecting depression on Social Networks based on tweets on Twitter and proposed a multi-step approach and applied Latent Dirichlet Allocation (LDA) algorithm for clustering of documents and evaluation of probabilities of similarity of the topics. In conclusion, they found a set of words with negative sentiments used by the people who were depressed. Another research was done by S. M. Chaware in 2020 [4] in which the authors used machine learning algorithms i.e., CNN to extract the text from Facebook posts TSVM for the classification of posts and the stress level and KNN for recommendation of nearby hospitals. Another research was conducted by S. N. Mounika. in 2019 [5] in which the authors detected the stress levels of the students by extracting the Facebook posts data using RNN and then further classification based on tweets on social media by analysing the text and emojis by implementing deep learning models such as CNN and RNN. Another research was done by Laszlo Nemes in 2020 [6] in this research the authors have extracted posts, tweets and comments from Twitter based on COVID-19 using NLP and Sentiment classification using RNN. Another research was done by C. Hutto and peers in 2018 [7] the authors presented a VADER which is the simple rule-based model for the sentiment analysis on social media tweets and the effectiveness was evaluated on benchmarks like LIWC, ANEW, SentiWordNet and machine learning techniques such as Naïve Bayes, Maximum Entropy and SVM. Finally, research was done by Syamsyah Jaya Putra 2018 [8] in which the authors presented a hybrid model for the social media sentiment analysis for the Indonesian Text in which they combined the lexical features and maximum entropy methods and employed machine learning techniques-based classification.

After thorough research of these papers, we have concluded that [1] [2] [3] [4], [5], [6] used automatic-based sentiment analysis in which they have done classification based on learning from previous data using machine learning techniques. In [7] the authors have used rule-based sentiment analysis in which they have presented a rule-based model VADER for sentiment analysis. Finally, in the study [8] the authors used a hybrid model combining rule-based and automated sentiment analysis in which they combined lexicon-based and machine learning-based methods. In our research, we are using automatic-based sentiment analysis using machine learning techniques to predict the stress by analysing the textual comments of Reddit, because by employing automatic-based models we are getting higher accuracy, and cost-effectiveness to build these models and flexibility can be achieved. These models can be easily trained on huge-scale datasets which can contribute to the organizations.

IV. METHODOLOGY

We have followed the CRISP-DM methodology for this project. CRISP-DM emphasizes business problems and defining project objectives. For this project of Predicting stress through texts, the methodology follows business and data un-

derstanding, followed by preprocessing using NLP techniques, modelling and evaluation.

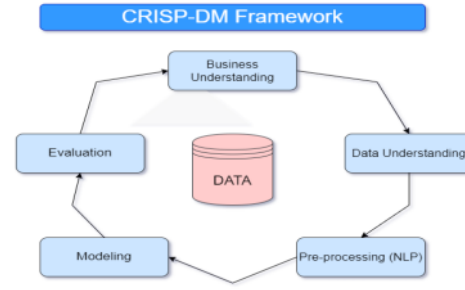


Fig. 1. CRISP-DM Framework

Fig. 1, shows the flow diagram of the CRISP-DM framework.

A. Business Understanding

The business understanding of this project includes identifying stress, and understanding and addressing it. Stress greatly impacts humans, like an individual's motivation, job performance, and overall well-being. The stress predictive model will be useful in various domains, including individuals seeking assistance, mental health professionals, researchers, and corporate organizations. We have used NLP techniques to preprocess the human text data and evaluation is done to focus on the reliability and effectiveness of the model in predicting stress. Deployment of this model in real-time will aid various businesses in providing insights into stress, and contribute to mental health support. This translates to increased productivity and higher work engagement.

B. Data Understanding

The dataset is derived from Kaggle and contains Reddit posts. It contains 2832 rows and 7 variables. The target variable is the 'stress' variable with binary values 0 and 1, where 1 indicates 'stress' and 0 is 'no stress'. Individuals have expressed their mental health problems through these posts, and the dataset contains data from various subreddit posts like relationships, PTSD, anxiety, domestic violence, etc.



Fig. 2. Distribution of Target variable Stress

In Fig. 2, the distribution of stress can be seen. Stress factors in the dataset are expressed as a percentage. The blue bar,

Subreddit	Count
ptsd	20.4
relationships	19.5%
anxiety	17.7%
domesticviolence	11.1%
assistance	10.2%
survivorsofabuse	8.6%
homeless	5.9%
almosthomeless	2.8%
stress	2.3%
food_pantry	1.3%

In Fig. 3, The frequency distribution of stress factors across different subreddit channels is summarized in the horizontal bar graph. The chart has one bar for each subreddit, and the length of the bar indicates how many times stress factors are mentioned there. PTSD-related subreddits account for 20.6% of users. The second most popular subreddits are relationships at 19.5%, followed by anxiety at 17.7%. Other subreddits include domestic violence, abuse survivors, homelessness, stress, getting mental health help, and even a food pantry.



In Fig. 4, the word cloud that displays the most recurring words in the indicated writings is shown. The size of a characteristic word is proportionate to the number of times it is stated in the texts that are labelled "not stressed.". This will enable us to locate the central ideas and matters that have been dealt with in all the unstressed outputs. Here, words like "know," "time," "life," and "better" are bigger and appear more often.

The NLP techniques have been applied to the data to pre-process the human language texts. The following steps are applied to the data:

- 2) **Removal of non-characters:** Special letters and characters like '!', '@', '#', and '_' are removed from the texts to eliminate unnecessary noise.
- 3) **Tokenization:** The texts are split into individual tokens using a tokenizer. Texts are divided into individual words or basic units.
- 4) **Removal of Stopwords:** Common English words like 'the', 'and', 'of', and 'in' are removed using a stopwords library. This step is important because these words hold no meaning ²³ classification.
- 5) **Stemming:** **Stemming is a process of reducing the word into its original or base form.** E.g. crying is simplified to crying. This helps the model to learn better.

text	processed_text
He said he had not felt that way before, suggested i go rest.	said had felt way suggested go rest so trigger ahead y
Hey there /assistance. Not sure if this is the right place to	hey r assistance sure right place post but go on cur
My mom then hit me with the newspaper and it shocked n	mom hit newspaper shocked that would this know d
until i met my new boyfriend, he is amazing, he is kind, he	met new boyfriend amazing kind sweet is good stud
October is Domestic Violence Awareness Month and i am, i	october domestic violence awareness month domes
i think he doesn't want to put in the effort for the relations	think want put the effort the relationship work both
It was a big company so luckily i didn't have to see him all t	big company luckily didn have see all time when d
It cleared up and i was okay but, On Monday i was thinking	cleared okay on monday thinking human how brain i
i actually give an assistant half my emergency fee to come	actually give assistant half emergency fee come help
just feel like the street life has fucked my head up there. There's	just feel like street life fucked head there some how
Next week lil"ll" be flying for our family vacation. The flight	next week ll flying our family vacation flight t very
Everything sets me off and i'm not having a panic attack	everything set off almost panic attack going bad soon
lil"ll"ve been taking 12.5 mg zolof for about 6 months i've	ve taking 12.5 mg zolof about 6 months year super

In Fig. 5, after applying the NLP preprocessing techniques, we can see the difference between raw text and processed text. The texts are in lowercase, and stopwords and special characters are removed. Stemming has been applied to the texts.

Feature extraction and vectorization are fundamental steps in Natural Language Processing (NLP) that involve converting preprocessed data into numerical representations that machine learning algorithms can understand and process. These techniques play a crucial role in transforming raw text data into a numerical format. So, we pass the data to a learning model. The following are some of the common techniques for feature extraction and vectorization.

- 3) **Bag-of-Words (BoW)**: Fig. 6 explains the Bag-of-Words work principle. This technique was used to understand and analyze the text. It works by counting the frequency of words in a document while ignoring their order of appearance. For example, BoW would count how many times each word (the cat, sat, on, mat) appears in the sentence "The cat sat on the mat" without taking into account its order. This would result in a numerical representation of the text that machine learning algorithms can understand. BoW is simple and versatile, making it useful for tasks like sentiment analysis and text classification. However, it has some limitations, like losing the word sequence and making high-dimensional data. This method shows each document as a vector, with each element

corresponding to the frequency of a word in it. It ignores the order of the words in the text [9].

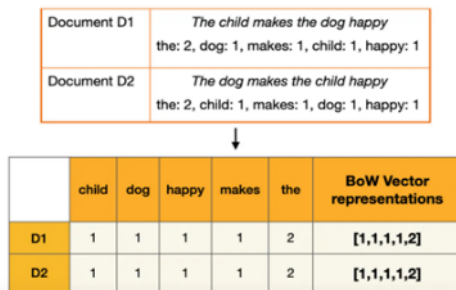


Fig. 6. Bag-of-Words work principle

3
2) *Term Frequency-Inverse Document Frequency (TF-IDF)*: TFG-IDF analyses two things. How often the word appears in a text (TF) and how different it is in all texts (IDF). If a word appears a lot in one text but not a lot in others, it is probably significant TF_IDF. Every word is scored based on these factors. In a text, high scores indicate that the word is significant, while low scores indicate that it is less significant. By concentrating on the important words and avoiding the common words, we can better understand what texts are about [10].

E. Modelling

We use the classification method to predict the chances of stress in a text to determine whether there is stress or not. We analysed the text using the Random Forest model to determine whether or not.

The Random Forest is a very powerful model that can provide precise predictions and is not vulnerable to overfitting, that is the model will absorb too much from the training data. It is easy to work with and even handles missing data thoroughly, moreover, it works well for many features, interpretability, and high accuracy. Though it is not so easy to comprehend how it makes its own decisions, most of the time it reveals which features are more important. In addition, it is really quick to train, which 25 convenient for those situations where there is a lot of data. The Random Forest classifier predicts the sentiments of the test data, and then the predicted labels are measured against the ground truth labels, utilizing accuracy levels and other performance metrics.

V. RESULTS/FINDINGS 13

We evaluate model performance metrics such as accuracy, precision, recall, F1-score, and ROC-AUC. This evaluation helps in determining how effectively the model performs.

For the hypothesis test, we conducted the Mann-Whitney test, which was used to analyse the relationship between the length of words and the accuracy of the model to predict the stress level of an individual, which resulted in a static score of 1098656.00 and a p-value of 0.0000153, which was much

less than the threshold of 0.05.

We got an accuracy of 89% on the test set by applying the TF/IDF feature extraction method with the Random Forest model, which was a good accuracy, a precision of 97% on class 0 and 83% on class 1, and a recall value of 78% on class 0 and 98% on class 1. These metrics show the classifier's ability to accurately categorize all instances of every sentiment class as we can see in Fig. 7

	precision	recall	f1-score	support
0	0.97	0.78	0.87	338
1	0.83	0.98	0.90	372
accuracy			0.89	710
macro avg	0.90	0.88	0.89	710
weighted avg	0.90	0.89	0.89	710

Fig. 7. Classification Report

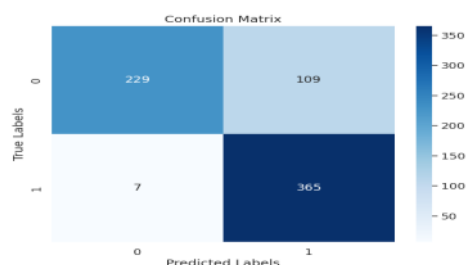


Fig. 8. Heatmap for Confusion Matrix

In Fig. 8, the confusion matrix shows that out of 338 instances where stress was not present (class 0), the model correctly predicted 265 instances (true negatives) but incorrectly classified 73 instances as stressed (false positives). On the other hand, a total of 372 times when stress was present (class 1), our model correctly identified 365 instances (true positives) and made 7 instances non-stressed (false negatives).

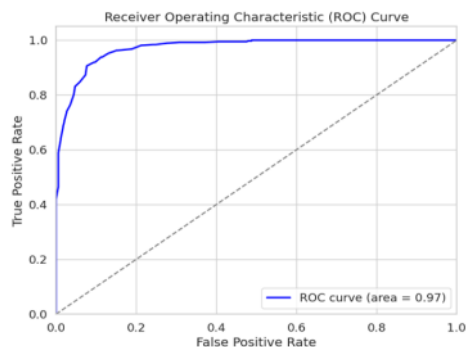


Fig. 9. ROC Curve

Fig. 9 shows the true positive rate (sensitivity) against the false positive rate (1-specificity) at various thresholds. The classifier's general performance is measured by the area under the ROC curve (AUC), where higher values indicate better class discrimination between classes. Finally, according to the Mann-Whitney U test, the p-value is less than 0.05. We reject the null hypothesis. There is a significant relationship between the number of words in the text and the model prediction of 1. Individuals tend to use many words in their posts if they are expressing stress.

VI. INTERPRETATION OF THE RESULTS

In the above section, we have informed about the results of our model's performance by considering various performance metrics such as accuracy, precision and recall. In this section, we will interpret the significance of the results in terms of a higher accuracy of 89%, the precision of 90% and recall of 88% which ensures that this model can improve customer engagement by effectively and accurately detecting whether the individual is stressed or not.

Additionally, in an organization, the model can analyse the employees' sentiments on their company's platform to enable the department to they can provide mental support on time to address workplace stressors. However, this model can be used to investigate which segments of the age group were suffering from stress-related issues by analysing the posts, comments and tweets and based on that potential target marketing can be done to solve the problems and requirements so that the organizations can improve their effectiveness which results in increasing customer acquisition and engagement.

Finally, we can say based on the results of the hypothesis test and reject the null hypothesis that there is no relationship between the length of the text and the accuracy of the model as we can say that we have got p-value less than the threshold of 0.05. With this hypothesis, we can say that the length of the text can influence the model's performance metrics. Businesses can consider the length of the text as the most important factor and this hypothesis also validates this so that businesses can make their approaches and find more insights from social media.

VII. CONCLUSION

For sentiment analysis on text data, the Random Forest classifier achieves 89% accuracy. The classification report, ROC curve, and AUC scores further support the classifier's performance. They show effective discrimination between sentiment classes and balanced performance across precision, recall, and F1-score metrics. These performance metrics show that the Random Forest model works well in differentiating stressed from non-stressed text. The Random Forest classifier suggests the possibility of accurate sentiment analysis in real-world applications. From the hypothesis test, we could further understand how people communicate their mental health concerns and for future work, we will investigate linguistic patterns associated with stress expression in text data.

REFERENCES

- [1] T. Nijhawan, G. Attigeri, and T. Ananthakrishna, "Stress detection using natural language processing and machine learning over social interactions," *Journal of Big Data*, Mar. 20, 2022. [Online]. Available: <https://link.springer.com/article/10.1186/s40537-022-00575-6#citeas>
- [2] S. Sahu, E. Kithani, M. Motwani, S. Motwani, and A. Ahuja, "Stress Detection of Office Employees Using Sentiment Analysis," *Advances in intelligent systems and computing*, Jan. 01, 2021. [Online]. Available: https://link.springer.com/chapter/10.1007/978-981-33-4367-2_15
- [3] "Twitter Analysis for Depression on Social Networks based on Sentiment and Stress," *IEEE Conference Publication — IEEE Xplore*, Oct. 01, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/8963550>
- [4] S. M. Chaware*, C. Makashir, C. Athavale, M. Athavale, and T. Baraskar, "Stress Detection Methodology based on Social Media Network: A Proposed Design," *International Journal of Innovative Technology and Exploring Engineering*, vol. 9, no. 3, pp. 3489–3492, Jan. 2020, doi: 10.35940/ijitee.b7537.019320. [Online]. Available: <http://dx.doi.org/10.35940/ijitee.b7537.019320>
- [5] "Detection of Stress Levels in Students using Social Media Feed," *IEEE Conference Publication — IEEE Xplore*, May 01, 2019. [Online]. Available: <https://ieeexplore.ieee.org/abstract/document/9065720>
- [6] L. Nemes and A. Kiss, "Social media sentiment analysis based on COVID-19," *Journal of Information and Telecommunication*, vol. 5, no. 1, pp. 1–15, Jul. 2020, doi: 10.1080/24751839.2020.1790793. [Online]. Available: <http://dx.doi.org/10.1080/24751839.2020.1790793>
- [7] C. J. Hutto and É. Gilbert, "VADER: A Parsimonious Rule-Based Model for Sentiment Analysis of Social Media Text," *Proceedings of the ... International AAAI Conference on Weblogs and Social Media/Proceedings of the International AAAI Conference on Weblogs and Social Media*, May 16, 2014. [Online]. Available: <https://ojs.aaai.org/index.php/ICWSM/article/view/14550>
- [8] S. J. Putra, I. Khalil, M. N. Gunawan, R. Amin, and T. Sutabri, "A Hybrid Model for Social Media Sentiment Analysis for Indonesian Text," Nov. 19, 2018. [Online]. Available: <https://dl.acm.org/doi/abs/10.1145/3282373.3282850>
casa_token=IQAwA5EetbIAAAA:gKeldKdE7VY8KC9-9KEjnn2h-F-m5ZfNEv78LfL3CBcA1p7fCx8A6a8SBAX0H9NP1f_6J7X-ZAWjw
- [9] Medium, Aysel Aydin, "4 — Bag of Words Model in NLP" [Online]. Available: <https://ayseaydin.medium.com/4-bag-of-words-model-in-nlp-434cb38cdd1b> [Accessed on: Apr 4, 2024]
- [10] Towards Data science, Raymond Cheng, "Understanding TF-IDF: A Traditional Approach to Feature Extraction in NLP" [Online]. Available: <https://towardsdatascience.com/understanding-tf-idf-a-traditional-approach-to-feature-extraction-in-nlp-a5bfbe04723f> [Accessed on: Apr 4, 2024]

DAPA CA-2 Report

ORIGINALITY REPORT

26%

SIMILARITY INDEX

14%

INTERNET SOURCES

6%

PUBLICATIONS

21%

STUDENT PAPERS

PRIMARY SOURCES

1

Submitted to National College of Ireland

Student Paper

18%

2

www.coursehero.com

Internet Source

1%

3

www.ijraset.com

Internet Source

1%

4

Submitted to Maastricht University

Student Paper

1%

5

M. Skourtos, D. Damigos, C. Tourkolias, A. Kontogianni. "Efficient energy labelling: the impact of information content and style on product choice", Energy Efficiency, 2021

Publication

<1%

6

Submitted to Shanghai High School International Division

Student Paper

<1%

7

riunet.upv.es

Internet Source

<1%

8	Submitted to National University of Ireland, Galway Student Paper	<1 %
9	hdl.handle.net Internet Source	<1 %
10	www.researchgate.net Internet Source	<1 %
11	Khalid Hussain, Khaldoon Nusair, Muhammad Junaid, Waqas Aman. "A two-actor model for understanding user engagement with content creators: Applying social capital theory", Computers in Human Behavior, 2024 Publication	<1 %
12	Pons, Eugene H.. "Twitter Activity of Urban and Rural Colleges: A Sentiment Analysis Using the Dialogic Loop", Florida International University, 2023 Publication	<1 %
13	www.mdpi.com Internet Source	<1 %
14	www.ncbi.nlm.nih.gov Internet Source	<1 %
15	edoc.unibas.ch Internet Source	<1 %
16	research.usq.edu.au Internet Source	<1 %

17	rgu-repository.worktribe.com Internet Source	<1 %
18	www.hindawi.com Internet Source	<1 %
19	McCarter, Kortny. "Leveraging Sentiment, Bibliometric and Trend Analysis to Evaluate and Predict Quantum Cybersecurity Research Investments", The George Washington University, 2024 Publication	<1 %
20	fastercapital.com Internet Source	<1 %
21	F. M. Javed Mehedi Shamrat, Sovon Chakraborty, M. M. Imran, Jannatun Naeem Muna et al. "Sentiment analysis on twitter tweets about COVID-19 vaccines using NLP and supervised KNN classification algorithm", Indonesian Journal of Electrical Engineering and Computer Science, 2021 Publication	<1 %
22	Nagpal, Meghan Shyama. "Managing Type 2 Diabetes during a Pandemic: An Investigation of Reddit Data", University of Toronto (Canada), 2021 Publication	<1 %
23	Shivakumar R. Goniwada. "Introduction to Datafication", Springer Science and Business	<1 %

24

ebin.pub

Internet Source

<1 %

25

vdoc.pub

Internet Source

<1 %

26

Alfredo Daza, Nemias Saboya, Jorge Isaac Necochea-Chamorro, Karoline Zavaleta Ramos, Yesenia del Rosario Vásquez Valencia. "Systematic review of machine learning techniques to predict anxiety and stress in college students", Informatics in Medicine Unlocked, 2023

Publication

<1 %

27

Lecture Notes in Computer Science, 2015.

Publication

<1 %

28

Maite Taboada, Julian Brooke, Milan Tofiloski, Kimberly Voll, Manfred Stede. "Lexicon-Based Methods for Sentiment Analysis", Computational Linguistics, 2011

Publication

<1 %

Exclude quotes Off

Exclude matches Off

Exclude bibliography On