Extraversion Personality Classification using Machine Learning from Linguistic Features

BY

Tanvir Hasan Rifat
ID: 172-15-9931

Khalad Mossarraf Hossin Mashud
ID: 172-15-10058

Abrar Islam Sajid
ID: 172-15-9694

This Report Presented in Partial Fulfillment of the Requirements for the Degree of Bachelor of Science in Computer Science and Engineering

Supervised By

Mr. Ahmed Al Maruf
Lecturer
Department of CSE
Daffodil International University

Co-Supervised By

Shah Md. Tanvir Siddique
Assistant Professor
Department of CSE
Daffodil International University

DAFFODIL INTERNATIONAL UNIVERSITY

DHAKA, BANGLADESH

JUNE 2021

APPROVAL

This Project titled "**Extraversion Personality Classification using Machine Learning from Linguistic Features**", submitted by **Tanvir Hasan Rifat** 172-15-9931and, **Khalad Mossarraf Hossin Mashud** 172-15-10058and **Abrar Islam Sajid** 172-15-9694 to the Department of Computer Science and Engineering, Daffodil International University, has been accepted as satisfactory for the partial fulfillment of the requirements for the degree of B.Sc. in Computer Science and Engineering and approved as to its style and contents. The presentation has been held on 2nd May,2021.

BOARD OF EXAMINERS

Chairman

Dr. Touhid Bhuiyan
Professor and Head
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Md. Sadekur Rahman
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

Internal Examiner

Dr. Fizar Ahmed
Assistant Professor
Department of Computer Science and Engineering
Faculty of Science & Information Technology
Daffodil International University

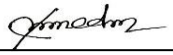
External Examiner

Dr. Shamim H Ripon
Professor
Department of Computer Science and Engineering
East West University

DECLARATION

We hereby declare that, this project has been done by us under the supervision of **Mr. Ahmed Al Maruf, Lecturer, Department of CSE** Daffodil International University. We also declare that neither this project nor any part of this project has been submitted elsewhere for award of any degree or diploma.

Supervised by:



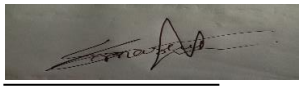
Mr. Ahmed Al Maruf
Lecturer
Department of CSE
Daffodil International University

Co-Supervised by:

Shah Md. Tanvir Siddiquee
Assistant Professor
Department of CSE
Daffodil International University

Submitted by:

Tanvir Hasan Rifat
ID: 172-15-9931
Department of CSE
Daffodil International University



Khalad Mossarraf Hossin Mashud
ID: 172-15-10058
Department of CSE
Daffodil International University

Abrar Islam Sajid
ID: 172-15-9694

Department of CSE

Daffodil International University

ACKNOWLEDGEMENT

First, we express our heartiest thanks and gratefulness to almighty God for His divine blessing makes us possible to complete the final year project/internship successfully.

We really grateful and wish our profound our indebtedness to **Shah Md. Tanvir Siddiquee**, Assistant professor, Department of CSE Daffodil International University and **Mr. Ahmed Al Maruf,** Lecturer, Department of CSE Daffodil International University**,** Dhaka. Deep Knowledge & keen interest of our supervisor in the field of "*Deep Learning*" to carry out this project. Their endless patience, scholarly guidance, continual encouragement, constant and energetic supervision, constructive criticism, valuable advice, reading many inferior drafts and correcting them at all stage have made it possible to complete this project.

We would like to express our heartiest gratitude to -----, -----, and Head**,** Department of CSE, for his kind help to finish our project and also to other faculty member and the staff of CSE department of Daffodil International University.

We would like to thank our entire course mate in Daffodil International University, who took part in this discuss while completing the course work.

Finally, we must acknowledge with due respect the constant support and patients of our parents.

ABSTRACT

Personality is a principal premise of human behavior. At generally essential, personality including examples of thought, feeling, behaviors that make an individual novel. Personality will straight forwardly or in a roundabout way impact the collaboration or inclinations of an individual. This research utilizing different learning algorithms and ideas of data mining to mine on the data features and gain from the example. The point of this analysis is to investigate various choices of the algorithm on altering the personality prediction source code by utilizing logistic regression algorithm and to discover whether the accuracy of the characterization can be improved. There are five characteristic of various individuals that are known as the Big Five characteristic, which is openness, neuroticism, conscientiousness, agreeableness and extraversion that have been put away in the dataset utilized for preparing. At that point, an outline and comparison will be provided on the various measures taken to decrease the issues faced by researchers in this field. Classification methods executed are Support Vector Machine, K Nearest Neighbor, Naïve Bayes, Logistic Regression, Decision Tree, and Random Forest.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 INTRODUCTION

Extraversion is one of the Big Five measurements, mirroring the common change of more explicit attributes, like gregariousness, assertiveness, energy, talkativeness, action level, and excitement-seeking. [9] It shows how friendly and social a person is. A person who scores high in extraversion on a personality test is the energy everyone needs. They appreciate being with people, participating in social gatherings, and are brimming with energy. A person low in extraversion is less preceding and is happier with working without anyone else. [10] Various jobs require various levels of extraversion. An advanced level of extraversion may be helpful for jobs that require a lot of association with others, similar to public relations, instructing, and deals. Places that are more centered around working freely and alone might be more fit to individuals with a lower level of extraversion. The extraversion characteristic is one of five domains in the Big Five personality theory developed by independent researchers over many decades. Numerous enrollment specialists and profession guides utilize this model to test personality. [11] Personality characteristics impact numerous parts of individual behavior, like the mentality toward machines, by and large work execution, just as scholarly inspiration; this last finding proposes that preparation frameworks would be more effective in the event that they could adjust to the students personality. [5]

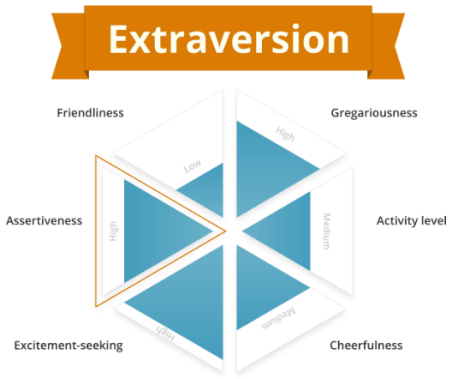


Figure 1: Sub traits of extraversion.

With almost 3 billion people presently utilizing some type of social media, understanding its relationship with personality has become a vital focal point of psychological research .

Here we've focused on the comparison of different algorithms to predict Extraversion Personality Classification from a social media site which is YouTube. We've used YouTube video's subtitle as our dataset to classified the extraversion personality.

1.2 MOTIVATION

We'll try to find the extraversion personality classification of people (could be extrovert or introvert) using machine learning linguistic features from the data sets of YouTube subtitle. As someone's personality is hard to understand just by seeing some videos they share cause for knowing people in real we need to spend time with them for long. Sometimes we can't match people exactly same in between virtual and real life.

As extraversion demonstrates how outgoing and social a person is. High in extraversion implies they appreciate being with people, participating in social gatherings, and are loaded with energy. A person low in extraversion is less outgoing and is more open to working without anyone else.

Prediction is only possible with data mining. And it never gives us the accurate result but with this we can get only probable result. Once we get the prediction of someone's personality, we can at least get a probable idea about them which may help us indeed.

With this methodology if anybody or any companies want to hire someone or invite them as an influencer or for some goods, they can get the minimum result to get the perfect one for them or for the people.

Figure 2: Reasons for research on personality prediction

1.3 RESEARCH QUESTIONNARIES

- Is it possible to predict EPC by the dataset of blogs?
- Is it possible to predict EPC by machine learning algorithms?

1.4 EXPECTED OUTPUT

Data mining is the way toward figuring out enormous informational collections to recognize designs and build up connections to take care of issues through information examination. Data mining instruments allow enterprises to predict future patterns.

At the end of our research after analysis these outcomes, we're expecting an all over score of someone's extraversion personality level.

1.5 REPORT LAYOUT

Chapter 1: “Introduction”. This part is decorated with the basic understanding as Introduction, Motivation, Research questionnaires, Expected outcome, and report layout.

Chapter 2: “Background”. It has Terminologies, Related work and literature review, Challenges.

Chapter 3: “Research Methodology”. It consists of working process, Dataset and attributes, applied mechanism.

Chapter 4: “Experimental Result and discussions”. It consists of Experimental Setup, Experimental Results and Analysis, Discussions.

Chapter 5: “Conclusion”. It consists Summary, Conclusion, Future plan.

CHAPTER 2
BACKGROUND

2.1 TERMINOLOGIES

WEKA:

Weka (Waikato Environment for Knowledge Analysis) is a well-known set-up of machine learning software. It is delivered as an open-source software with an assortment of machine learning algorithms for data mining undertaking. Weka contains apparatuses for data pre-processing, classification, regression, clustering, association rules, and visualization. It gives us an execution of learning algorithms that can be handily utilized with our datasets.

MACHINE LEARNING:

Machine learning is a strategy for data analysis that mechanizes logical model structure. It is a part of artificial intelligence based on the possibility that frameworks can gain from information, distinguish examples and settle on choices with insignificant human intercession. Machine learning is utilized for clinical determination, image processing, forecast, grouping, learning association, regression and so on.

As we're working with data mining and it is one of the rudiments of machine learning. It's feasible to rapidly and naturally produce models that can investigate greater, more perplexing information and convey quicker, more accurate outcomes even on an extremely enormous scope.

2.2 RELATED WORKS AND LITERATURE REVIEW

This section of this paper will present the existing literature conducted on Extraversion Personality Classification. There have already been several studies on Personality Classification with machine learning, artificial neural network, and data mining on different social media like Facebook, Twitter. As we worked on YouTube subtitle,

François Mairesse, Marilyn Walker, [4] developed to predict Personality. The authors have used algorithms such as Naïve-Bayes, Decision Tree. Naïve Bayes algorithm gave them the best result on their paper.

Alireza Souri, Shafigheh Hosseinpour, Amir Masoud Rahmani et al. [8] the authors implemented Boosting-Naïve Bayesian along with other algorithms on their Personality classification based on profiles of social networks research. Naïve Bayes algorithm gave them an accuracy of 46.67%.

Zhenkun Zhou, Ke Xu, Jichang Zhao et al. [9] the authors have used algorithms such as Random Forest, Naïve Bayes, SVM. Random Forest gave them an accuracy of 34.73%, Naïve Bayes gave them an accuracy of 39.17% and SVM gave them an accuracy of 42.31%.

2.3 CHALLENGES

Challenges we faced were many but the pandemic situation of COVID-19 was the most threatening issue while we're researching on this paper. Besides, one of our members was even affected by COVID, though he overcame from this. And working in group from different places just by connecting on social sites was hard to maintain. It was really an unexpected challenging time for all.

Along this, we faced some other challenges too like learning to use new software, applying many algorithms to our datasets, managing all the stuffs we needed staying in home.

But we've successfully overcame from all the challenges.

CHAPTER 3
RESEARCH METHODOLOGY

3.1 WORKING PROCESS:

Firstly, we choose the area to research with. Then we made up our mind that we will research for predicting extraversion personality classification with machine learning from linguistic feature.

Then we collected the data and worked with the data processing for the datasets to make it suitable for applying algorithms on it. After that, we implemented different kinds of machine learning algorithms on our dataset.

3.2 DATASET AND ATTRIBUTES

We collected our dataset from the YouTube subtitle as blogs. We took 57 attributes class in our dataset.
From them, 30 Character based Features, 15 Word based Features, 3 Structural Features, 9 Function Words.
Our classes are Extraversion Personality Classification in a probable range of high and low.

Table 1: Feature dataset attributes.

No	Name	Description	Type	Attribute Information
1	Characters	No. of Character	Numerical	No. of character

2	Punctuations	No. of Punctuations	Numerical	No. of Punctuations
3	Special Characters	No. of Special Characters	Numerical	No. of Special Characters
4	Individual alphabets (a)	No. of Individual alphabets	Numerical	No. of Individual alphabets
5	Individual alphabets (b)	No. of Individual alphabets	Numerical	No. of Individual alphabets
6	Individual alphabets (c)	No. of Individual alphabets	Numerical	No. of Individual alphabets
7	Individual alphabets (d)	No. of Individual alphabets	Numerical	No. of Individual alphabets
8	Individual alphabets (e)	No. of Individual alphabets	Numerical	No. of Individual alphabets
9	Individual alphabets (f)	No. of Individual alphabets	Numerical	No. of Individual alphabets
10	Individual alphabets (g)	No. of Individual alphabets	Numerical	No. of Individual alphabets
11	Individual alphabets (h)	No. of Individual alphabets	Numerical	No. of Individual alphabets
12	Individual alphabets (i)	No. of Individual alphabets	Numerical	No. of Individual alphabets
13	Individual alphabets (j)	No. of Individual alphabets	Numerical	No. of Individual alphabets
14	Individual alphabets (k)	No. of Individual alphabets	Numerical	No. of Individual alphabets
15	Individual alphabets (l)	No. of Individual alphabets	Numerical	No. of Individual alphabets
16	Individual alphabets (m)	No. of Individual alphabets	Numerical	No. of Individual alphabets

17	Individual alphabets (n)	No. of Individual alphabets	Numerical	No. of Individual alphabets
18	Individual alphabets (o)	No. of Individual alphabets	Numerical	No. of Individual alphabets
19	Individual alphabets (p)	No. of Individual alphabets	Numerical	No. of Individual alphabets
20	Individual alphabets (q)	No. of Individual alphabets	Numerical	No. of Individual alphabets
21	Individual alphabets (r)	No. of Individual alphabets	Numerical	No. of Individual alphabets
22	Individual alphabets (s)	No. of Individual alphabets	Numerical	No. of Individual alphabets
23	Individual alphabets (t)	No. of Individual alphabets	Numerical	No. of Individual alphabets
24	Individual alphabets (u)	No. of Individual alphabets	Numerical	No. of Individual alphabets
25	Individual alphabets (v)	No. of Individual alphabets	Numerical	No. of Individual alphabets
26	Individual alphabets (w)	No. of Individual alphabets	Numerical	No. of Individual alphabets
27	Individual alphabets (x)	No. of Individual alphabets	Numerical	No. of Individual alphabets
28	Individual alphabets (y)	No. of Individual alphabets	Numerical	No. of Individual alphabets
29	Individual alphabets (z)	No. of Individual alphabets	Numerical	No. of Individual alphabets
30	Alphabets	Total No. of Alphabets	Numerical	Total No. of Alphabets
31	Words	No. of Words	Numerical	No. of Words

32	Words with 1 Character	No. of words with 1 Character	Numerical	No. of words with 1 Character
33	Words with 2 Character	No. of words with 2 Character	Numerical	No. of words with 2 Character
34	Words with 3 Character	No. of words with 3 Character	Numerical	No. of words with 3 Character
35	Words with 4 Character	No. of words with 4 Character	Numerical	No. of words with 4 Character
36	Words with 5 Character	No. of words with 5 Character	Numerical	No. of words with 5 Character
37	Words with 6 Character	No. of words with 6 Character	Numerical	No. of words with 6 Character
38	Words with 7 Character	No. of words with 7 Character	Numerical	No. of words with 7 Character
39	Words with 8 Character	No. of words with 8 Character	Numerical	No. of words with 8 Character
40	Words with 9 Character	No. of words with 9 Character	Numerical	No. of words with 9 Character
41	Words with 10 Character	No. of words with 10 Character	Numerical	No. of words with 10 Character
42	Words with 11 Character	No. of words with 11 Character	Numerical	No. of words with 11 Character
43	Words with 12 Character	No. of words with 12 Character	Numerical	No. of words with 12 Character
44	Words more than 12 character	No. of words more than 12 Character	Numerical	No. of words more than 12 Character
45	Avg. Word Length	Avg. Word Length	Numerical	Avg. Word Length
46	Sentence	No. of Sentence	Numerical	No. of Sentence

47	Avg. Sentence Length in terms of character	Avg. Sentence Length in terms of character	Numerical	Avg. Sentence Length in terms of character
48	Avg. Sentence Length in terms of word	Avg. Sentence Length in terms of word	Numerical	Avg. Sentence Length in terms of word
49	Function Word	No. of Function Words	Numerical	No. of Function Words
50	Noun	Percentage of Noun	Numerical	Percentage of Noun
51	Pronoun	Percentage of Pronoun	Numerical	Percentage of Pronoun
52	Verb	Percentage of Verb	Numerical	Percentage of Verb
53	Adjective	Percentage of Adjective	Numerical	Percentage of Adjective
54	Adverb	Percentage of Adverb	Numerical	Percentage of Adverb
55	Preposition	Percentage of Preposition	Numerical	Percentage of Preposition
56	Conjunction	Percentage of Conjunction	Numerical	Percentage of Conjunction
57	Interjection	Percentage of Interjection	Numerical	Percentage of Interjection

3.3 APPLIED MECHANISM

We used different types of machine learning techniques from linguistic features in this paper. We have different kinds of results from them. Those algorithms are:

- SVM
- KNN
- Logistic Regression
- Decision Tree
- Random Forest

- Naïve Bayes.

SVM: It is an SML calculation that changes the training data into a higher measurement and fabricates a N-dimensional hyperplane that parts the data into two classes utilizing preparing tuples called "support vectors." SVM classifier means to determine the separators that best discrete classes. SVM is a fantastic classifier for text archives and is broadly utilized in text classification where high dimensionality is the standard. [2] SVM can sum up more precisely on concealed cases relative with classifiers that plan to limit the training blunder to create a superior prediction on multi-class issues. In light of the outcomes given by these strategies, this research will utilize the Support Vector Machine (SVM) method since this strategy can sum up issues well to deliver high accuracy and a lower error rate than the previous existing methods. [1]

We have got a good result with SVM (Support Vector Machine). The accuracy is 53% and the kappa statistics is 0.0585.

Confusion metrics for SVM:

Classified as: a = yes

b = no

Table 2: Confusion metrics for SVM

A	b
212	0
191	1

KNN: The KNN model is used for learning that is driven by instance for classification. To tag text with a particular personality category, KNN considers how the majority of its neighbors voted. Different distance metrics were utilized for neighbor estimation, in particular Euclidean distance, Makowski, Manhattan distance and so forth. Its principal benefits incorporate its straight-forwardness and good accuracy results if the parameters are chosen reasonably. The huge disadvantages of this method are the high computation and memory prerequisites. [2] In this paper, we present a kNN model for predicting an individual's BIG-5 personality traits in light of analysis of the substance of YouTube subtitles. At firstly, we extract features by analyzing the content of the blogs. Then, we discretize the linguistic and emotional features. Finally, we applied the kNN model to predict personality traits based on extraversion. Furthermore, report framework exhibitions as far as precision, recall and F score etc.

We have got a good result with KNN. The accuracy is 49% and the kappa statistics is -0.0352.

Confusion metrics for KNN:

Classified as: a = yes

b = no

Table 3: Confusion metrics for KNN

A	b
124	88
119	73

Logistic Regression: Logistic regression can be an amazing analytical procedure for use when the result variable is dichotomous. The focal numerical concept underlies logistic regression is the logit-the common algorithm of an odds proportion. [4]

We have got a probable result with Logistic Regression. The accuracy is 52% and the kappa statistics is 0.041.

Confusion metrics for Logistic Regression:

Classified as: a = yes

b = no

Table 4: Confusion metrics for Logistic Regression

A	b
118	94
99	93

Decision tree: An ordinary tree incorporates roots, branches and leaves. The similar structure is continued in the Decision Tree. It contains root nodes, components and leaf nodes. A root node is the parent of all things considered, and as the name recommends, it is the highest node in Tree. A decision tree is a tree where each node shows a feature (attribute), each connection (branch) shows a decision (rule), and each leaf shows a result. Decision trees impersonate human-level reasoning, so it's so easy to get the data and make some great and it's so simple to catch the data and make some sensible interpretations. The entire thought is to create a tree like this for the entire data and cycle a solitary result at each leaf. [3]

We have got a probable result with Decision Tree. The accuracy is 52% and the kappa statistics is 0.0329.

Confusion metrics for Decision Tree:

Classified as: a = yes

b = no

Table 5: Confusion metrics for Decision Tree

A	b
135	77
116	76

Random Forest: Random Forest is a combination of a progression of tree structure classifiers. Random Forest has been uncontrollably utilized in classification and prediction and utilized in regression as well. It possesses high classification accuracy, endures variances and commotion well, and never got overfitting. Random Forest has been quite possibly the most well-known exploration strategies in the data mining territory.[5]
We have got a probable result with Random Forest. The accuracy is 54% and the kappa statistics is 0.0693.

Confusion metrics for Random Forest:

Classified as: a = yes

b = no

Table 6: Confusion metrics for Random Forest

A	b
147	65
120	72

Naïve Bayes: The Naive Bayesian classifier depends on Bayes' hypothesis with the freedom suppositions between predictors. A Naive Bayesian model is easy to work, with no convoluted iterative parameter assessment, which makes it particularly helpful for broad datasets. Despite its effortlessness, the Naive Bayesian classifier habitually does surprisingly well and is generally utilized on the grounds that it frequently beats more refined classification methods. [2]

The exploration we have for this paper will add on data mining and machine learning because of crude data classified utilizing the Gaussian Naive Bayes classification. The researched prototype we implemented empowers us to predict a person's personality by investigating every one of the YouTube subtitle datasets.[6] We also demonstrate that Naive Bayes functions admirably for explicit almost functional feature dependencies, in this manner arriving at its best exhibition.

We have got a probable result with Naïve Bayes. The accuracy is 53% and the kappa statistics is 0.044.

Confusion metrics for Naïve Bayes:

Classified as: a = yes

b = no

Table 7: Confusion metrics for Naïve Bayes

A	b
156	56
133	59

CHAPTER 4

EXPERIMENTAL RESULTS AND DISCUSSION

4.1: EXPERIMENTAL SETUP:

In our research paper, we will predict the extraversion personality classification with the help of machine learning classifiers. We have used 6 algorithms in our dataset to find out which gives the best result and compare the algorithm results to predict Extraversion. For the implementation, we have used Weka. We used the CSV file in weka to get the results of algorithms. We have used 6 folds cross-validations for each classifier.

4.2: EXPERIMENTAL RESULTS AND ANALYSIS

We used different types of machine learning techniques in this paper and those algorithms are SVM, KNN, Logistic Regression, Decision Tree, Random Forest, Naïve Bayes.

Table 8: Result of Different Algorithms for Prediction

Classifiers	Precision	F score	recall	Correctly Classified %	Kappa Statistics	ROC	MCC	PRC area
SVM	0.53	0.53	0.527	52.9703	0.0585	0.529	0.059	0.517
KNN	0.483	0.483	0.488	48.7624	0.0352	0.49	0.036	0.496
Logistic Regression	0.522	0.522	0.522	52.2277	0.041	0.542	0.041	0.557
Decision Tree	0.525	0.515	0.522	52.4759	0.0329	0.492	0.004	0.497
Random Forest	0.539	0.53	0.543	54.2079	0.0693	0.547	0.072	0.543
Naïve Bayes	0.527	0.509	0.532	53.2178	0.044	0.523	0.036	0.508

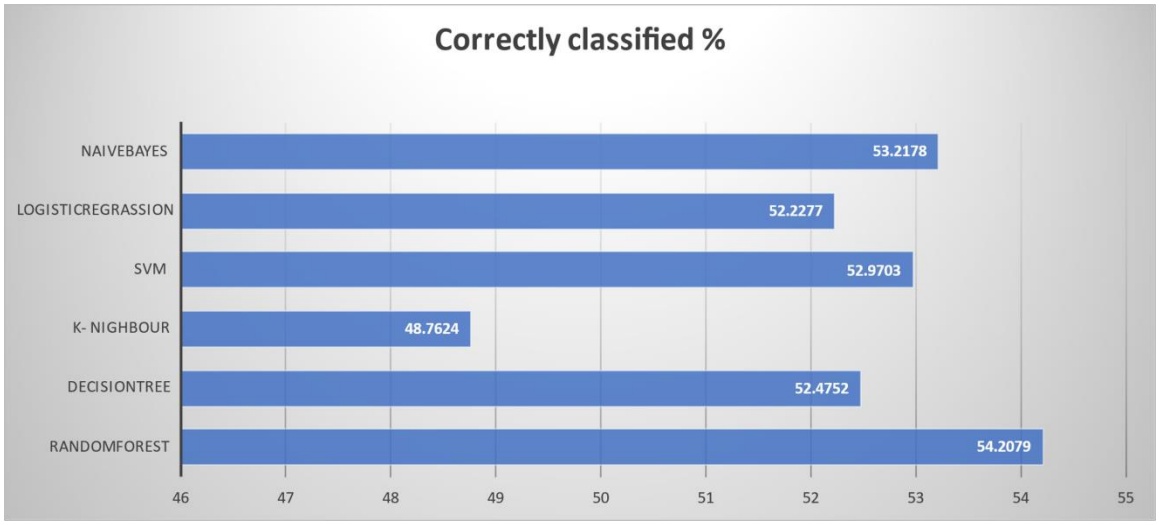


Figure 3: Bar graph Correctly Classified of Different Classifiers.

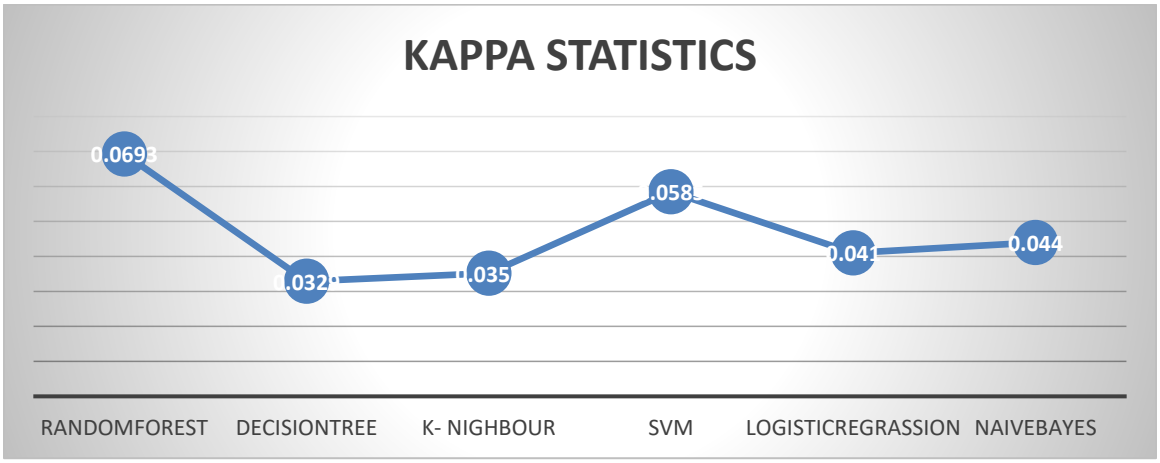


Figure 4: Line graph of Kappa Statistics of Different Classifiers.

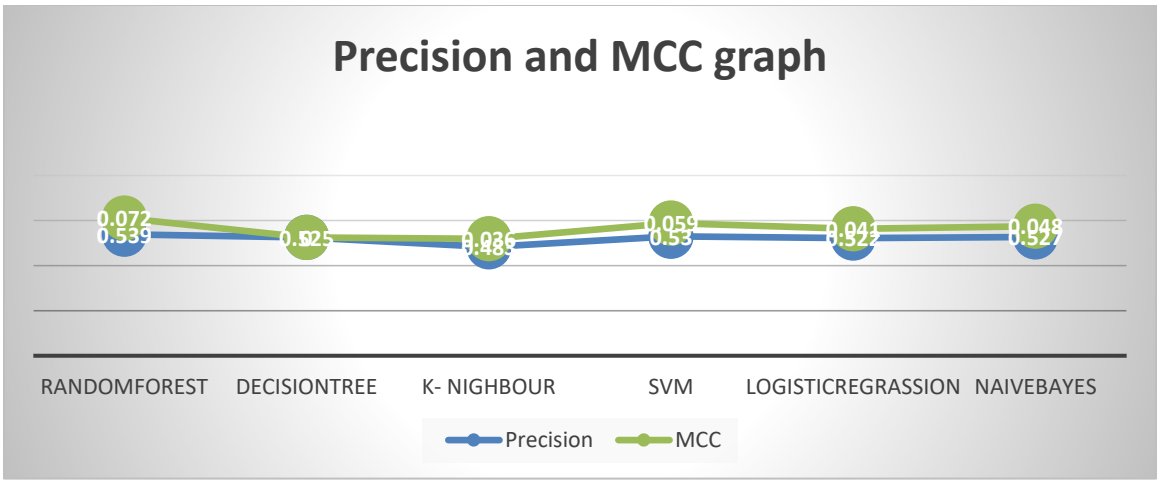


Figure 5: Line graph of Precision and MCC value of different Algorithms.

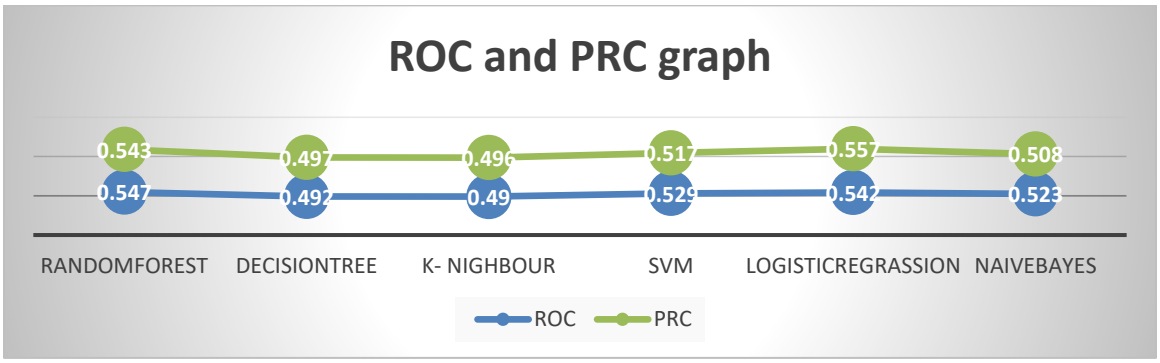


Figure 6: Line graph of ROC and PRC value of different Algorithms.

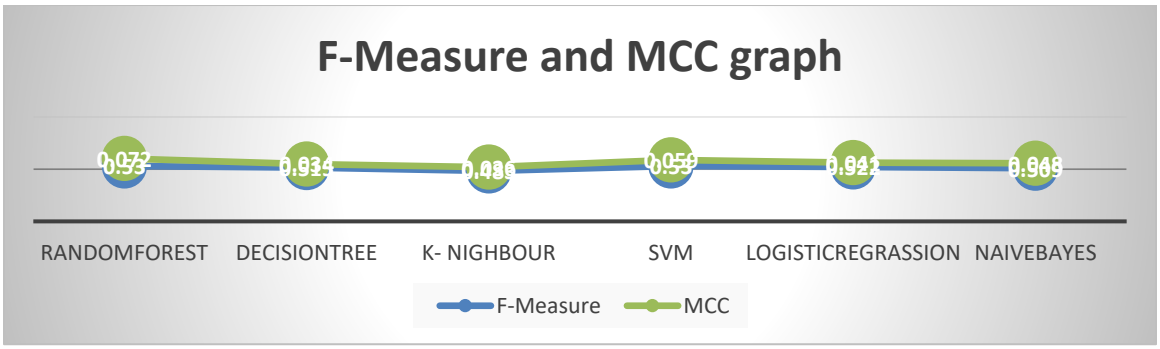


Figure 7: Line graph of F-Measure and MCC of different Algorithms.

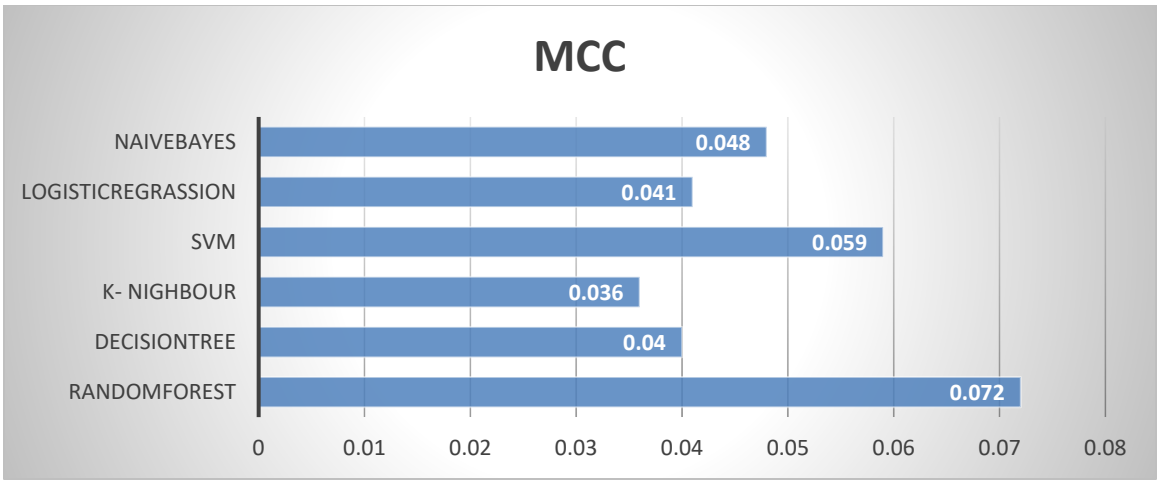


Figure 8: Bar graph of MCC of different Algorithms.

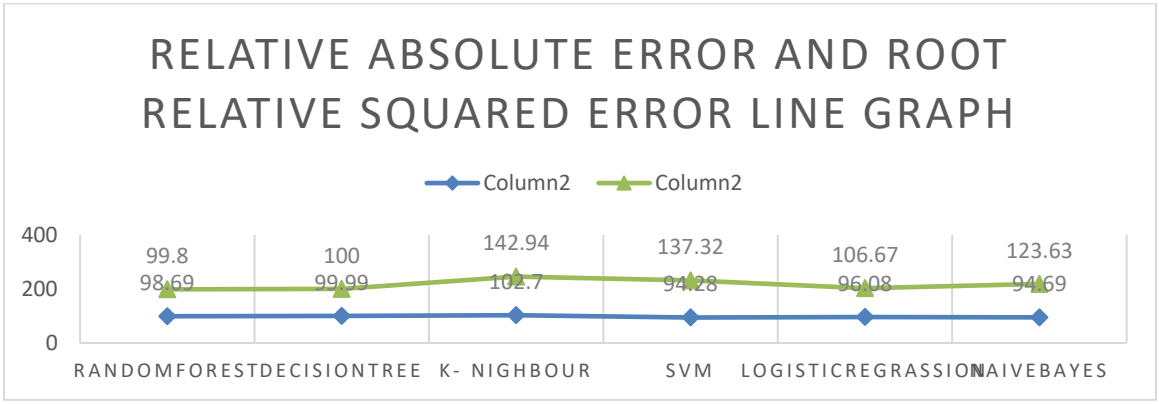


Figure 9: Line Graph of Relative Absolute Error and Root Relative Squared Error of different Algorithms.

Table 9: Incorrectly Classified, Mean absolute error, Root mean squared error, Relative absolute error, Root relative squared error of different Algorithms

Algorithms	Incorrectly Classified %	Mean absolute error	Root mean squared error	Relative absolute error %	Root relative squared error %
SVM	47.03	0.47	0.68	94.21	137.32
KNN	51.24	0.512	0.71	102.7	142.94
Logistic Regression	47.78	0.48	0.53	96.08	106.67
Decision Tree	47.53	0.5	0.49	99.99	100
Random Forest	45.8	0.49	0.5	98.69	99.8
Naïve Bayes	46.79	0.47	0.61	94.69	123.63

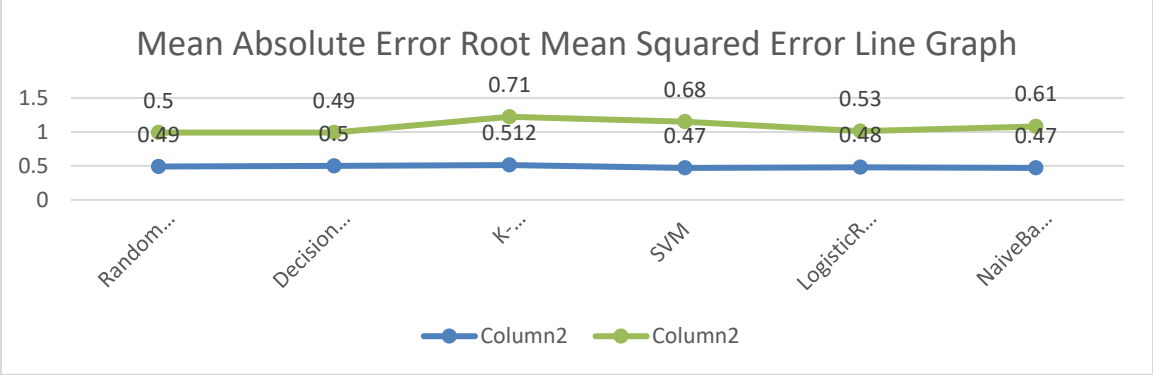


Figure 10: Line graph of Mean Absolute error Root mean squared error of different Algorithms.

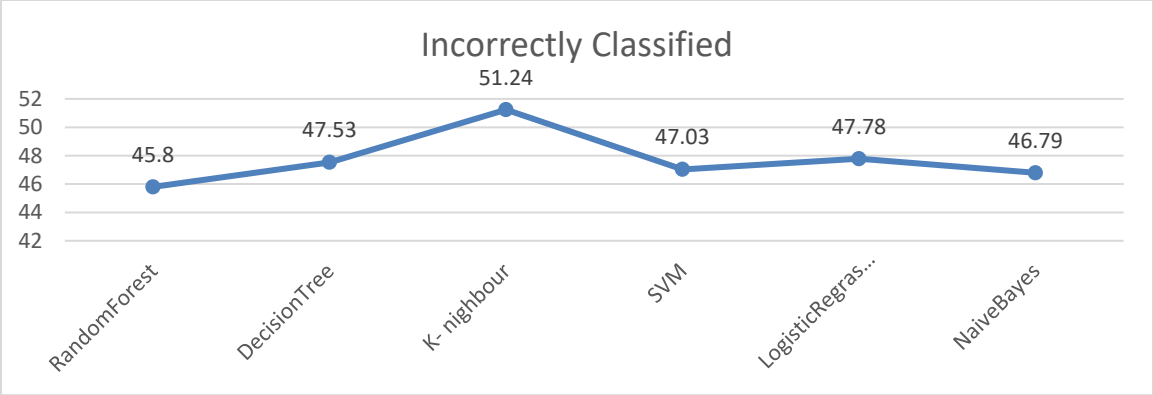


Figure 11: Line graph of Incorrectly Classified of different Algorithms.

4.3: DISCUSSION

So, after implementing all the algorithms and after analysis we have found Random Forest, Naïve Bayes and SVM give us the most accuracy.

Random Forest gives us an accuracy of almost 54.2079%

Naïve Bayes gives us an accuracy of almost 53.2178%

SVM gives us an accuracy of almost 52.9703%

CHAPTER 5

CONCLUSION AND FUTUE PALN

5.1 SUMMARY:

The nature of human behavior is different from person to person. If we provide this with a range of average, we would be able to differ the extraversion personality from person to person. So, our research paper will help to predict Extraversion Personality Classification for people.

We have implemented some important machine learning algorithms in our dataset to find out the best result so that it can predict EPC. All of our algorithms gave us excellent results but Random Forest, Naïve Bayes and SVM have given us the best accuracy result among all classifiers.

5.2 CONCLUSION:

We have implemented 6 algorithms in our dataset.

Random Forest, Naïve Bayes and SVM gave us the most accuracy of almost on an average of 53%. Our purpose is to predict EPC by using machine learning. It will provide us goods indeed. Like, we can get a probable prediction for the people with high EPC as they can influence or lead or motivate others, moreover likely they can be the top priority categorized people. In other hand, people with low EPC can't be fitted in everywhere thus they're. But at least we can have a minimum prediction about them that should we follow/choose them or not.

5.3 FUTURE PLAN:

As we worked here with YouTube subtitle as blogs and classified the extraversion personality. In future we can have more studies on other organizations or companies for providing the best of extraversion personality to them. We used here only 6 algorithms. In future we would use more algorithms so that we can be able to predict more accurate result.

REFERENCE

[1] Ninda Anggoro Utamia , Warih Maharania, Imelda Atastina, “Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method”, Conference/Journal, Vol 179, Pages 177-184, 2021

[2] Garg, Shruti and Kumar, Prince and Garg, Ashwani, “Comparison of Machine Learning Algorithms for Content Based Personality Resolution of Tweets”, Conference/Journal, January 2020

[3] Patel, Harsh and Prajapati, Purvi, “International Journal of Computer Sciences and Engineering”, Conference/Journal, Vol 6, Page 74-78, October 2018.

[4] Peng, Joanne and Lee, Kuk and Ingersoll, Gary, “An Introduction to Logistic Regression Analysis and Reporting”, Conference/Journal, Vol 96, Page 3-14, September 2002.

[5] François Mairesse, Marilyn Walker, “Words mark the nerds: Computational models of personality recognition through language”, April 2021.

[6] Yanli Liu, Yourong Wang, Jian Zhang, “Lecture Notes in Computer Science”, 2012, Vol 7473.

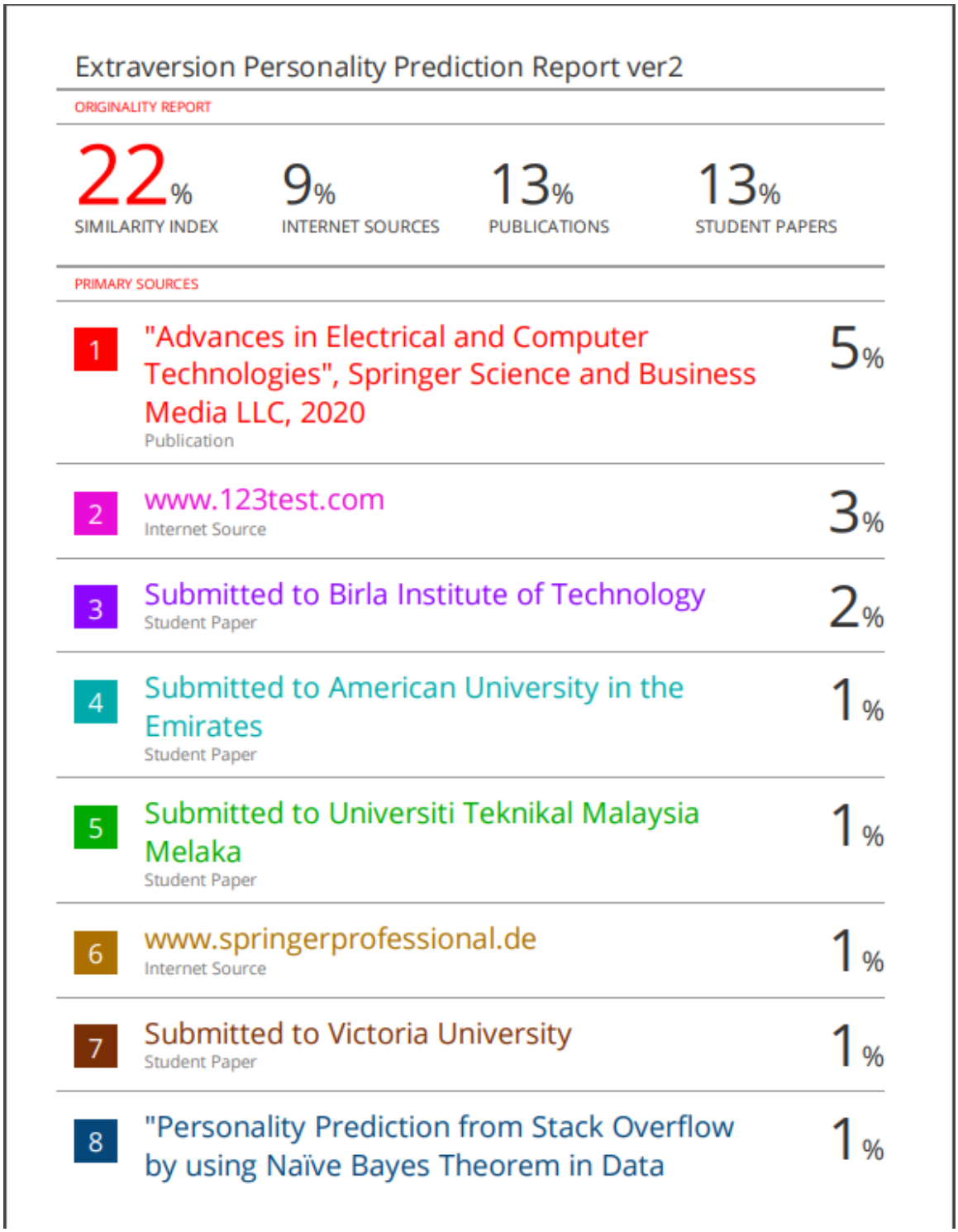
[7] Sandhya Katiyar, Sanjay Kumar, Himdweep Walia, “Personality Prediction from Stack Overflow by using Naïve Bayes Theorem in Data Mining”, Conference/Journal, Volume-9 Issue-3, January 2020.

[8] Alireza Souri, Shafigheh Hosseinpour, Amir Masoud Rahmani, “Personality classification based on profiles of social networks’ users and the five-factor model of personality”, Conference/Journal, *Human-centric Computing and Information Sciences*, Vol 8, Article no 24 , August 2018

[9] Zhenkun Zhou, Ke Xu, Jichang Zhao, “Extroverts tweet differently from introverts in Weibo”, Conference/Journal, *EPJ Data Science*, volume 7, Article number: 18, 2018.

[10] Extraversion, available at << https://www.123test.com>>, last accessed on 30-04-2021 at 08:00 PM.

[11] Extraversion, available at << https://www.sciencedirect.com>>, last accessed on 30-04-2021 at 08:00 PM.



Extraversion Personality Prediction Report ver2

ORIGINALITY REPORT

22%	9%	13%	13%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	"Advances in Electrical and Computer Technologies", Springer Science and Business Media LLC, 2020 Publication	5%
2	www.123test.com Internet Source	3%
3	Submitted to Birla Institute of Technology Student Paper	2%
4	Submitted to American University in the Emirates Student Paper	1%
5	Submitted to Universiti Teknikal Malaysia Melaka Student Paper	1%
6	www.springerprofessional.de Internet Source	1%
7	Submitted to Victoria University Student Paper	1%
8	"Personality Prediction from Stack Overflow by using Naïve Bayes Theorem in Data	1%