|  |  |
| --- | --- |
|  | **DAYANANDA SAGAR UNIVERSITY**  **KUDLU GATE, BANGALORE – 560068** |



**Bachelor of Technology**

**in**

**COMPUTER SCIENCE AND ENGINEERING**

**Special Topic- 2 Report**

**PERSONALITY CLASSIFICATION**

By

**Amogh M Cheluvaraj (ENG20CS0026)**

**Ankitha Shree B (ENG20CS0034)**

**Arun V (ENG20CS0040)**

**C Bharat Reddy (ENG20CS0066)**

**Under the supervision of**

**Dr.BonduVenkateswarlu**

**Associate Professor in CS**

**DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING,**

**SCHOOL OF ENGINEERING**

**DAYANANDA SAGAR UNIVERSITY,**

**(2022-2023)**



School of Engineering

Department of Computer Science& Engineering

Kudlu Gate, Bangalore –560068

Karnataka, India

**CERTIFICATE**

This is to certify that the Special Topic 2 titled **“TITLE”**is carried out by **Name1 (USN), Name2 (USN),Name2 (USN),Name3 (USN),Name4 (USN),** bonafide students of Bachelor of Technology in Computer Science and Engineering at the School of Engineering, Dayananda Sagar University, Bangalore in partial fulfillment for the award of degree in Bachelor of Technology in Computer Science and Engineering, during the year **2022-2023**.

|  |  |  |
| --- | --- | --- |
| **Guide Name** | **Dr. Girisha G S** | **Dr. Uday Kumar Reddy**  **K R** |
| Assistant/Associate/ Professor  Dept. of CSE,  School of Engineering  Dayananda Sagar University  Date: | Chairman, CSE  School of Engineering  Dayananda Sagar University  Date: | Dean  School of Engineering  Dayananda Sagar University  Date: |

**Name of the Examiner** **Signature of Examiner**

1.

2.

<

**DECLARATION**

We, **Name1 (USN),Name2 (USN),Name2 (USN),Name3 (USN),Name4 (USN),** are student’s of the fourth semester B.Tech in **Computer Science and Engineering**, at School of Engineering, **Dayananda Sagar University**, hereby declare that the Special Topic 2 titled **“Title”** has been carried out by us and submitted in partial fulfillment for the award of degree in **Bachelor of Technology in Computer Science and Engineering**during the academic year **2022-2023**.

|  |  |
| --- | --- |
| **Student** | **Signature** |
| **Name1:**  **USN :** |  |
| **Name2**  **USN :** |  |
| **Name3:**  **USN :** |  |
| **Name4:**  **USN :** |  |
| **Place : Bangalore**  **Date :** |  |

**ACKNOWLEDGEMENT**

It is a great pleasure for us to acknowledge the assistance and support of many individuals who have been responsible for the successful completion of this Special Topic 2.

First, we take this opportunity to express our sincere gratitude to School of Engineering & Technology, Dayananda Sagar University for providing us with a great opportunity to pursue our Bachelor’sdegree in this institution.

We would like to thank **Dr. Uday Kumar Reddy K R, Dean**, **School of Engineering & Technology**, **Dayananda Sagar University**for his constant encouragement and expert advice. It is a matter of immense pleasure to express our sincere thanks to **Dr. Girisha G S, Chairman**, **Department of Computer Science, and Engineering**, **Dayananda Sagar University,** for providing the right academic guidance that made our task possible.

We would like to thank our guide **……………………..**, Associate / **Assistant/ Professor**, **Dept. of Computer Science and Engineering**, **Dayananda Sagar University**, for sparing his/her valuable time to extend help in every step of our Special Topic 2, which paved the way for smooth progress and the fruitful culmination of the project.

We would like to thank our Special Topic 2 Coordinators, Dr. Savitha Hiremath, Dr. Bondu Venkateswarlu, and all the staff members of Computer Science and Engineering for their support.

We are also grateful to our family and friends who provided us with everyrequirement throughout the course.We would like to thank one and all who directly or indirectly helped us in the Special Topic 2.

****

**TABLE OF CONTENTS**

Page

LIST OF ABBREVIATIONS .................................................................................v

LIST OF FIGURES ………....................................................................................vi

LIST OF TABLES …………..................................................................................vii

ABSTRACT ............................................................................................................ viii

CHAPTER 1 INTRODUCTION………................................................................. 1

1.1. ..................................................................................................................... 2

1.2. ...................................................................................................................... 3

CHAPTER 2 PROBLEM DEFINITION………..................................................... 5

CHAPTER 3 LITERATURE SURVEY................................................................... 5

CHAPTER 4PROJECT DESCRIPTION.................................................................... 5

4.1. PROPOSED DESIGN ................................................................................. 5

4.2. ASSUMPTIONS AND DEPENDENCIES................................................... 5

CHAPTER 5REQUIREMENTS ................................................................................ 6

5.1. FUNCTIONAL REQUIREMENTS ............................................................ 6

5.2. ...................................................................................................................... 6

CHAPTER 6METHODOLOGY................................................................................ 7

CHAPTER 7EXPERIMENTATION.......................................................................... 8

CHAPTER 8TESTING AND RESULTS ….............................................................. 9

RESEARCH COMPONENT/ FUNDING/Paper Publication Details(IF ANY)………

REFERENCES... ....................................................................................................... 10

APPENDIX A ............................................................................................................ 11

**NOMENCLATURE USED**

|  |  |
| --- | --- |
| AI | Artificial Intelligence |
| DL | Deep Learning |
| GUI | Graphical User Interface |
| PHP | Pre-Processor Hyper text |
| MySQL | My Structured Query Language |

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| Fig. No. | Description of the figure | Page No. |
| 3.1 | Proposed Block Diagram | 5 |
| 4.1(a) |  | 7 |
| 4.1(b) |  | 8 |
| 4.2(a) |  | 12 |
| 4.2(b) |  | 12 |
| 4.2(c) |  | 13 |

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| Table No. | Description of the Table | Page No. |
|  |  | 9 |
|  |  | 22 |
|  |  | 22 |
|  |  | 31 |

**Abstract**

Personality is defined as the combination of traits and qualities of an individual which describes his own unique character. This project is done to automatically analyze a person’s personality based on certain criteria. The scope of this project can spread to a large area for instance in making the right career choices despite the confusion prevailing within yourself. Here we concentrate on an online survey taken up by any individual based on which the model predicts his personality. To build ones career is the first and foremost intention of many people. Every job has different requirements of personality features in order to take up the vacancy. To help them understand their personality in depth in turn helping them to choose the right field is a tedious task for at least three out ten of them. Well this project could solve their problem. Not only does this help in picking up a right career but also help in knowing more about a person. To proceed in anyway ahead it is important for us to know our strengths and weakness. This could again be determined be easy through our project. Our key focus for this project is using Machine learning to build a classifier capable of sorting people into their Myers-Briggs Type Index(MBTI) personality type based on text samples from their social media posts and real time dataset

**CHAPTER 1 INTRODUCTION**

Personality classification has been one of the most researched topics in the recent past. Personality is a combination of an individual’s behaviour and characteristics features that determines how he/she reacts under different circumstances. Individual Behaviour can influence by all choices like a person observes regarding various things like books, clothes, music and films. Personality can also affect his/her interaction with the outside world and his/her environment. Personality can also be used as an additional feature during recruitment process, career counselling, health counselling, etc. Predicting personality by analysing the behaviour of the person is an old technique. This manual method of personality prediction required a lot of time and resources. Analyzing personality based on one’s nature was a tedious task and a lot of human effort would be required to do such analysis. This traditional method of predicting personality would require a lot of time and was very limited in scale. Also, this manual analysis did not give accurate results while analyzing the personality of a user from their nature and behaviour. Since analysis was done manually, it affects the accuracy of the results as humans prone to be prejudice and generally see the things accordingly. The recent developments in technology have created a number of online platforms in which people can share information about themselves or in general and interact with each other. In today’s World Personality detection is one of the growing fields of research and social media data is a useful source of data for investigators because using this data we may assess the personality of a person like if he is an introvert or extrovert, his/her strength and weakness etc. Thus having an idea about one’s personality may in turn give us an idea about how a person may react during different situations, about their various types of preferences, about their mental health, whether the person is suitable for the job and so on. The Personality Classification project builds an understanding of strengths and blind spots. It also helps people understand how they might differ from one another. It is used to help people understand themselves more completely by understanding what their own preferences and tendencies are, and to help them work better with others through understanding that different people have different personalities. It is valuable for individuals and teams as they tackle such challenges as communication, handling conflict, managing change, making decisions, being a leader, or changing careers. It could be used for psychological science and understand performance expectations for real-world tasks. Finding correlations between a person’s MBTI personality type and writing style.

So machine learning models may be used to identify the personality of a person based on their social media data by building various machine learning models. Here the MBTI data set has been used to predict the online user personality. This data set has the data posted by a person in various social media platforms like Facebook, twitter etc. For our project we have gathered our own dataset along with the twitter posts collected as dataset for our training model. We sent out a google form for the college students to take up the survey based on their input for the given question we extract the features and classify their personalities. Most people believe that there are only two types of personalities: introverts and extroverts. MBTI evaluation helps us understand that personality is much more than that. With over 3.5 million assessments conducted each year, MBTI is the most widely used personality indicator globally. The Myers Briggs Type Indicator (MBTI) is a personality type system that divides everyone into 16 distinct personalities based on 4 dimensions. Researchers widely use machine learning to predict personality and psychological traits from digital records. But as a field of research, personality prediction is at a relatively early stage. It is important to understand if the predicted personalities retain characteristics from psychological science and understand performance expectations for real-world tasks. We're developing an MBTI personality classifier that uses machine learning models to predict a person’s personality based on the social media posts and real-time data as input. We find correlations between a person’s MBTI personality type and writing style. The classifier also demonstrates the validity of the MBTI test. We have used a decent amount of mined personality annotated data from social media.

Data Mining techniques are therefore used study and analyses data and then identify any hidden patterns or information from a large data set. These techniques are used to mine user characteristics and then train the model accordingly to predict the personality of other users in the future. Using these techniques, the personality of an applicant is analyzed who is applying for a job in an organization which gives priority to one’s behavior and personality rather than technical knowledge. Also, the applicant gets to know what all personality traits are in him/her and what all traits are missing. Thus, he/she can then be guided to develop those traits or to strengthen the other traits accordingly. The major purpose of this paper is to overview the data mining algorithms which are used to predict the personality of the user. We have used 3 algorithms in this project they are Support Vector Machine (SVM), Naive Bayes Algorithm and Logistic Regression. We perform comparative analysis for these algorithms and in turn test the accuracy of each which is then displayed in form of pie chart.

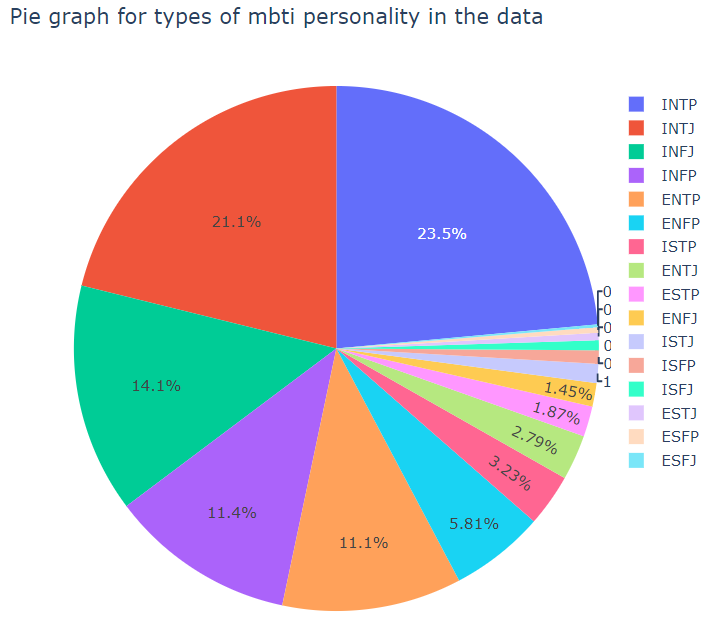
1. **figures and Tables**



Figure 1.1A chart with descriptions of each Myers–Briggs personality type

**Graphical user interface, application

Description automatically generated**

Pie chart displaying the number of samples of each personality present in the dataset

**CHAPTER 2 PROBLEM DEFINITION**

Predicting personality by analyzing the behavior of the person is an old technique. Analyzing personality based on one’s nature was a tedious task, and a lot of human effort would be required to do such analysis.Precise definition of the problemThere is significant interest in this area within the academic realm of psychology as well as the private sector. For example, many employers wish to know more about the personality of potential hires, to better manage the culture of their firm. In the world of work today, there is a need to empower individuals to become autonomous decision-makers. For successful career development, this empowerment should aim to help individuals to acquire decision-making skills. Lack of empowerment has led many people to 3 realise that the career they are in is not the best fit for them while already in the marketplace.

**CHAPTER 3**

**LITERATURE REVIEW**

Novel approaches to automated personality classification: Ideas and their potentials: This paper proposes several new research directions regarding the problem of Automated Personality Classification (APC). Firstly, we investigate possible improvements of the existing solutions to the problem of APC, for which we use different combinations of the APC corpora, psychological trait measurements, and learning algorithms. Afterwards, we consider extensions of the APC problem and the related tasks, such as dynamical APC and detecting personality inconsistency in a text. This entire research was performed in the context of social networks and the related data mining mechanisms.

Aleksandar Kartelj et. al. [2] said that reliable approaches can be used to classify the personality in various new researches by applying the concept of Automated Personality Classification. Firstly, we examined all the possible solutions and what all improvements can be made to the existing problems of Automated Personality Classification. Then we considered the extension of the Automated Personality Classification [APC] problem such as the Dynamic APC and how to remove inconsistency in textual data. This entire research was carried out in the context of social networks and related data mining mechanisms.

An ever-growing number of users share their thoughts and experiences using the Twitter micro logging service. Although sometimes dismissed as containing too little content to convey significant information, these messages can be combined to build a larger picture of the user posting them. One particularly notable personality trait which can be discovered this way is psychopathy: the tendency for disregarding others and the rule of society. In this paper, we explore techniques to apply data mining towards the goal of identifying those who score in the top 1.4% of a well-known psychopathy metric using information available from their Twitter accounts. We apply a newly proposed form of ensemble learning, Select RUSBoost (which adds feature selection to our earlier imbalance-aware ensemble in order to resolve high dimensionality), employ four classification learners, and use four feature selection techniques. The results show that when using the optimal choices of techniques, we can achieve an AUC value of 0.736. Furthermore, these results were only achieved when using the Select RUSBoost technique, demonstrating the importance of feature selection, data sampling, and ensemble learning. Overall, we show that data mining can be a valuable tool for law enforcement and others interested in identifying abnormal psychiatric states from Twitter data.

R. Wald have used social media like twitter contents to identify human psychology. They said Twitter, a micro blogging site, is used by several users to share their experiences and thoughts about their day-to-day life. Although researchers have often discarded the method of predicting personality by analysing the tweets because they are of the view that it contains very little content to predict significant information, but these tweets can be combined to make a larger picture of the user who is posting them. Select RUSBoost, a new form of ensemble learning has been used to predict psychopathy using Twitter, which uses four classification learners and four feature selection techniques.

Yago Saez et. al. [5] developed a system for analyzing the behavioral traits and cheerfulness of a user. Eysenck's theory defines too the human personality, the authors built a system wherein they would collect the text messages from different media resources such as twitter and face book and then classify them into various personality types. Although a clear link between behavioral traits and cheerfulness cannot be established, however some correlations do exist between them which could be found out in the coming future. J. Golbeck et. al. [6] said that Social network is a platform where the users tend to reveal themselves to the outside world, sharing their behavioral information and giving imminent to other people into their lives. Personality plays an important part in many types of interactions among people; it can be used to predict the job satisfaction, professional as well as romantic relationship success. Until now, in order to accurately predict user’s personality, they surveyed among different individuals with help of a survey test. However, this was highly unfeasible while collecting data from social media platforms and hence correct analysis of personality was a problem.

This paper[4], the author says that an output variable must be predicted or classified in the training dataset. It is expected that an algorithm would learn patterns from training data and then apply them to test data to predict or classify. “K-Nearest Neighbor (KNN)”, “Logistic Regression”, and “Stochastic Gradient Descent” are examples of classification algorithms.

“Myers–Briggs Type Indicator (MBTI)” Research in the personality field has long piqued the interest of psychologists, and one such study was conducted on the “Myers–Briggs Type Indicator” by a psychiatrist named “Carl Jung”. Then, “Katharine Briggs” and “Isabel Myers Briggs” created the “Myers-Briggs Type Indicator” for testing personality in the 1920s, based on “Jung’s theory of psychological types” . This model instrument has 16 personality types represented by a “personality types key” . In the “MBTI” system, for example, people classified as “INTPs” prefer “Introversion (I)”, “Intuition (N)”, “Thinking (T)”, and “Perception (P)” personality traits. We can classify the needs or behaviour of individuals according to labels, and then the machine can learn the patterns.

Further our research gave us insight about Manasi Ombhase[5] paper where she stresses on personality classification considered by personality psychology. According to that study[5], personality can be defined as a dynamic and organized set of characteristics of a person, which have a unique influence on cognition, motivation and behavior of that person. In this paper the problem of automated personality classification is considered based on information from the following content: textual content that the person wrote and meta information about a person received on request, through social networks or other means.

Predicting personality with Social Media, Social media [6] is a place where the users tend to present themselves to the outside world, sharing their personal details and giving insights to other people into their lives. Personality plays an important part in many types of interactions among people; it can be used to predict the job satisfaction, professional as well as romantic relationship success. Until now, in order to accurately predict user’s personality they needed to take a personality test. However, this was highly impractical while collecting data from social media platforms and hence correct analysis of personality was a problem

**CHAPTER 4: PROJECT DESCRIPTION**

**Project Design**

Dataset

Classification Algorithms

Evaluation of MBTI Personality Traits

Pre-Processing of Data

Personality Classification

Personality Type

**This is the proposed design for our project**

**CHAPTER 5: REQUIREMENTS**

**Hardware Requirements Specifications**

* **6 GB RAM**
* **Windows 7 or above**
* **10 GB Storage Space**
* **2 GB Graphics Card**

**Software Requirements Specifications**

* **Programming Language – Python**
* **Display Driver 11.0**
* **Environment - VS Code or any IDE**

**Python Libraries**

* **Pandas**
* **Pickle**
* **os. path**
* **plotly**
* **spacy**
* **IPython**
* **Matplotlib**
* **Nltk**
* **Sklearn**
* **TfidfVectorizer**
* **LinearSVC**
* **Pipeline**
* **Metrics**
* **classification\_report, confusion\_matrix, accuracy\_score, f1\_score**
* **LogisticReression**
* **MultinomialNB**

**CHAPTER 6: METHODOLOGY**

First, we collected the MBTI dataset from Kaggle and real-time dataset and converted all the text/posts to lower case, removed the URL links, removed special characters and numbers from texts, removed the extra spaces, removed stop words, removed MBTI personality words from the text, then Lemmatization, and at last, reconstructed samples to be equal-sized chunks (Around 500 words per sample). We then found features using CountVectorizer by converting the posts into DTM matrix of word count, but we used TfidfVectorizer to weigh the importance of words (features) across all posts and select more relevant features. Then, we created a pipelined model in order to facilitate the entire process of TfidfVectorizer and different classification models. Later, we trained the model and saved it to the disk and started making predictions. We then used the test data, which contains data from both MBTI dataset and collected data, to make predictions and analyze the accuracy of the different models. At last, we take an input from the user, and we predict his personality.

**CHAPTER 7 EXPERIMENTATION**

**The three algorithms used in our project are:**

**Support Vector Machine:** This is a type of machine which is basically used for analysis the data which receive from supervised learning and identify the patterns for classification [8]. Training data set is taken and checked that whether the test data belongs to existing class or not for personality classification and classification. Data is represented by Support Vector Machine model in the form of a point commonly in space which further classified in a line or in a hyper plane. The main idea behind the support vector machine algorithm is that if a classifier performs well at the most challenging comparisons, then it will definitely perform even better at the most easy comparisons. The steps involved in Support Vector Machine are:

1. Create vectors for given question answers.

2. Then calculate the weights of the vectors.

3. Get the vectors with highest value and find value of personality

4. Finally predict personality type.

**Naive Bayes Algorithm:** The performance of Naïve Bayes Algorithm in classifying data is quite accurate because the conditional independence assumption on which the entire algorithm is set up is rarely true for the real-world applications. The application of Bayes theorem forms the basis of Naïve Bayes Algorithm. A variation of Naïve Bayes Algorithm is Multinomial Naïve Bayes Algorithm which is also designed for classification purposes. The steps involved in Naive Bayes Algorithm are:

Step 1: Convert dataset into a frequency table.

Step 2: Create likelihood table based on probabilities of similar types of personality traits.

Step 3: Use Naïve Bayesian equations to calculate the posterior probability of each class.

Step 4: Highest probability class is the outcome of the prediction.

**Logistic Regression:** The system incorporates the working principle of parameter recognition in context to conditional posterior of data. The working model of our approach provides significant advantages due to dynamic flexibility and improved prediction. Model used to predict test dataset is “Logistic Regression” as Logistic regression is an effective model to predict output class labels for dependent categorical data.

Critical Code of the project

# Creating an instance to vectorizer:

    vectorizer = TfidfVectorizer()

    # Training the vectorizer:

    X\_train\_tfidf = vectorizer.fit\_transform(X\_train)

    # Training the classifier:

    clf = LinearSVC()

    clf.fit(X\_train\_tfidf, y\_train)

    # Pipelining the vectorizer and the classifier

    text\_clf = Pipeline([('tfidf',TfidfVectorizer()),('clf',LinearSVC())])

    text\_clf.fit(X\_train, y\_train)

    # saving the model to disk

    pickle.dump(text\_clf, open(filename, 'wb'))

predictions = text\_clf.predict(X\_test)

print(confusion\_matrix(y\_test,predictions))

A picture containing calendar

Description automatically generated

print(classification\_report(y\_test,predictions))

Table

Description automatically generated

print(accuracy\_score(y\_test, predictions))

>>>>0.8354388611294428

print(f"Overall accuracy of the model: {round(metrics.accuracy\_score(y\_test, predictions),2)}")

Overall accuracy of the model: 0.84

**CHAPTER 8: TESTING AND RESULTS**

**TESTING TECHNIQUES**

1. **Unit testing:**  It is a testing technique where it tests the smallest testable unit ofanapplication. It is done during the coding phase by the developers. To perform unit testing, a developer writes a piece of code (unit tests) to verify the code to be tested (unit) is correct. This section encompasses a range of checks that may be adopted generally to the data which is entered into an application system it checks for valid input data which includes mandatory fields cannot be empty, unique field value testing, null values testing, negative value testing, garbage value testing etc.

Let us explain two test cases. If we observe the first test case it takes an text input which says “when the whole world is silent, even one voice becomes powerful”. When testing was been performed on this test case the expected output was supposed to be ENJS( Extroverts, Intuitives, Judgers, Sensors), which should be right output and after testing the model the actual output produced was also ENJS. Since the actual output and the expected output is the same this test case was concludes to be ‘Pass’.

1. **Interface testing:** Interface testing is a type of software testing method which acts as an interacting or communicating medium between tworemote softwaresystems. ' Usually, an interface is a medium that helps two components to communicate with each other. It can be in any form like API, web design, graphics, etc.

When we run the program it automatically directs to the GUI/frontend of the project. Then they have the home tab and the login tab on the left. New users have to first signup create an account and the login to test their personality. Testing is performed for maintenance of security. Login credentials must be valid.

1. **Integration testing:** Integration testing is the stage in the software testing process in which individual software components are brought together and tested as a group The purpose of integration testing is to check whether individual modules, which are already expected to have been unit tested, work together as expected.

We integrate both the front end and back end of the project. Suppose a person logins in and his credentials match only then he is allowed to take up the test and personality prediction can happen.

1. **System testing:** In system testing, integration testing passed components are taken as input. The goal of integration testing is to detect any irregularity between the units that are integrated together. System testing detects defects within both the integrated units and the whole system. The result of system testing is the observed behavior of a component or a system when it is tested.
2. **Acceptance testing:** Acceptance Testing is a method of software testing where a system is tested for acceptability. The major aim of this test is to evaluate the compliance of the system with the business requirements and assess whether it is acceptable for delivery or not. The last phase of software testing after performing system testing.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Serial No** | **Test Id** | | | | **Input** | | **Actual Output** | **Expected**  **Output** | **Tested Output** | |
| 1 | 121 | | | | When the whole world is silent, even one voice becomes powerful. | | ENJF | ENJF | Pass | |
| 2 | 122 | | | | The important thing is not to stop questioning. Curiosity has its own reason for existence. | | INTP | INTP | Pass | |
| 3 | 123 | | | | When the whole world is silent, even one voice becomes powerful - Malala | | INJP | INJP | Pass | |
| 4 | | | 124 | | They are attentive and people-focused, and they enjoy taking part in their social community. Encourage, lift, and strengthen one another. For the positive energy spread to one will be felt by us all. | | INFJ | | ESFJ | | Fail | |
| 5 | | | | 125 | Many of life’s failures are people who did not realize how close they were to success when they gave up - Thomas.A.Edison | | INTJ | | INTJ | | Pass | |

**CHAPTER 8.2: RESULTS**

**Test Case**

Enter some text or any post:

Everybody isn't your friend just because they hang around and laugh with you doesn't mean they are your friends. People pretend well... at the end of the day only yourself and family will be with u....

**Result of test case**

["Everybody isn't your friend just because they hang around and laugh with you doesn't mean they are your friends. People pretend well... at the end of the day only yourself and family will be with u...."] ['INTJ']

INTJ

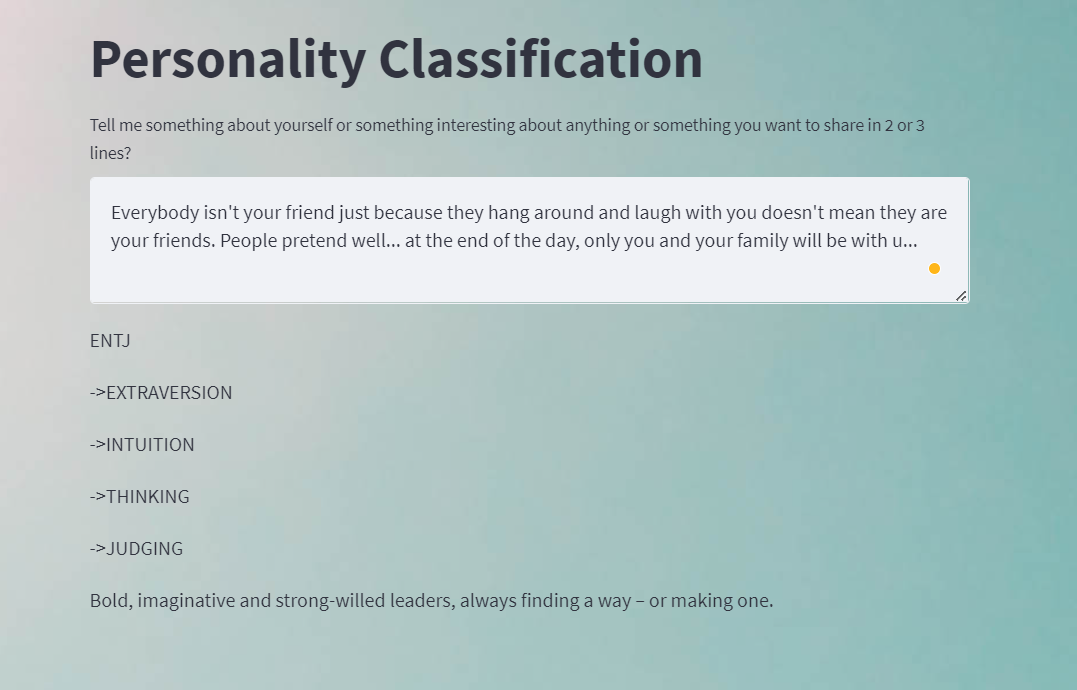
INTROVERSION

INTUITION

THINKING

JUDGING

Imaginative and strategic thinkers, with a plan for everything.



Evaluation of results in form of tables and Graphs

|  |  |  |  |
| --- | --- | --- | --- |
| **Support Vector Classifier** | | | |
|  | **Precision** | **Recall** | **F1\_Score** | |
|  |  |  |  | |
| **ENFJ** | 0.84 | 0.58 | 0.69 | |
| **ENFP** | 0.82 | 0.78 | 0.8 | |
| **ENTJ** | 0.9 | 0.8 | 0.84 | |
| **ENTP** | 0.86 | 0.83 | 0.84 | |
| **ESFJ** | 0.83 | 0.45 | 0.59 | |
| **ESFP** | 0.88 | 0.48 | 0.62 | |
| **ESTJ** | 0.9 | 0.84 | 0.87 | |
| **ESTP** | 0.95 | 0.9 | 0.92 | |
| **INFJ** | 0.81 | 0.84 | 0.83 | |
| **INFP** | 0.8 | 0.82 | 0.81 | |
| **INTJ** | 0.83 | 0.87 | 0.85 | |
| **INTP** | 0.84 | 0.87 | 0.86 | |
| **ISFJ** | 0.8 | 0.61 | 0.69 | |
| **ISFP** | 0.81 | 0.6 | 0.69 | |
| **ISTJ** | 0.86 | 0.68 | 0.76 | |
| **ISTP** | 0.89 | 0.79 | 0.84 | |
| **Average** | 0.781044 | 0.850826 | 0.734408 | |
| **Accuracy** | 0.835439 |  |  | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Multinomial Naïve Bayes** | | | |
|  | **Precision** | **Recall** | **F1\_Score** |
|  |  |  |  |
| **ENFJ** | 0.38 | 0.62 | 0.47 |
| **ENFP** | 0.55 | 0.54 | 0.46 |
| **ENTJ** | 1 | 0.29 | 0.02 |
| **ENTP** | 0.6 | 0.66 | 0.61 |
| **ESFJ** | 0.38 | 0.5 | 0.43 |
| **ESFP** | 0.43 | 0.5 | 0.46 |
| **ESTJ** | 0.76 | 0.61 | 0.46 |
| **ESTP** | 0.76 | 0.61 | 0.46 |
| **INFJ** | 0.38 | 0.5 | 0.43 |
| **INFP** | 0.29 | 0.54 | 0.38 |
| **INTJ** | 0.67 | 0.24 | 0.36 |
| **INTP** | 0.26 | 0.38 | 0.41 |
| **ISFJ** | 0.29 | 0.54 | 0.38 |
| **ISFP** | 0.55 | 0.61 | 0.58 |
| **ISTJ** | 0.63 | 0.45 | 0.59 |
| **ISTP** | 0.58 | 0.61 | 0.69 |
| **Average** | 0.058116 | 0.305478 | 0.082994 |
| **Accuracy** | 0.580751 |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Logistic Regression** | | | |
|  | **Precision** | **Recall** | **F1\_Score** |
|  |  |  |  |
| **ENFJ** | 0.82 | 0.53 | 0.64 |
| **ENFP** | 0.86 | 0.75 | 0.8 |
| **ENTJ** | 0.9 | 0.74 | 0.81 |
| **ENTP** | 0.84 | 0.8 | 0.82 |
| **ESFJ** | 0.78 | 0.21 | 0.33 |
| **ESFP** | 0.86 | 0.33 | 0.48 |
| **ESTJ** | 0.96 | 0.7 | 0.81 |
| **ESTP** | 0.95 | 0.85 | 0.9 |
| **INFJ** | 0.8 | 0.84 | 0.82 |
| **INFP** | 0.79 | 0.83 | 0.81 |
| **INTJ** | 0.81 | 0.88 | 0.84 |
| **INTP** | 0.83 | 0.88 | 0.86 |
| **ISFJ** | 0.89 | 0.47 | 0.61 |
| **ISFP** | 0.79 | 0.52 | 0.63 |
| **ISTJ** | 0.85 | 0.51 | 0.64 |
| **ISTP** | 0.9 | 0.71 | 0.79 |
| **Average** | 0.724331 | 0.850952 | 0.659541 |
| **Accuracy** | 0.824833 |  |  |

**REFERENCES**

[1]. S Katiyar, H Walia, S Kumar – Personality Classification System using Data Mining 2020

[2]. BayuYudhaPratama, RiyanartoSarno - Personality Classification Based on Twitter Text

[3]. Bhawna Singh, Swasti Singhal - Automated Personality Classification Using Data Mining Techniques

[4]. SakdipatOntoum, Jonathan H. Chan - Personality Type Based on MBTI with Text Posting Style by using ML and DL

[5]. Huansheng Ning, Sahraoui Dhelim, Nyothiri Aung – Friend Recommendation System Based on Big-Five Personality Traits and Hybrid Filtering

[6]. Automatic Personality Recognition of Authors Using Big Five Factor Model - KasulaChaithanya Pramodh

[7]. Bottom-Up and Top-Down:Predicting Personality with Psycholinguistic and Language Model Features - Yash Mehta, SaminFatehi

[8]. Personality Classification with Data Mining - Gaurav R. Savant (M.E. P.T. C.S.E.), Dr. G. R. Bamnote (Professor)