VIT®

Vellore Institute of Technology

(Deemed to be University under section 3 of UGC Act, 1956)

School of Computer Science and Engineering

J Component Report

Programme : B.Tech

Course Title : Artificial Intelligence

Course Code : CSE3013

Slot : A1

Title: Personality Prediction using OCEAN values

Team Members: Akshaya Ramachandram| 20BCE1727

Gowtham N | 20BCE1857

Pavan Kaarthik | 20BCE1873

Raksha Aruloli | 20BCE1737

Faculty: Dr.Uma Maheshwari **Sign:**

Date:

1. INTRODUCTION

Personality prediction is the process of predicting a person's personality traits or features using a variety of techniques, including self-reported surveys, behavioural observations, or machine learning algorithms. The consistent patterns of thoughts, feelings, and behaviours that set one person apart from another are referred to as personality traits. Personality prediction can aid in decision-making in areas including career growth, team building, and psychological interventions as well as better understanding and managing personality variations.

Researching and studying human behaviour has become more simpler thanks to the availability of high-dimensional and fine-grained data on the subject. For instance, studies on mobile sensing and information gathered from daily activities have had a significant impact on how psychologists conduct research and provide personality tests.In this perspective, personality psychology research and evaluation could be completely transformed by machine learning models. There are no problems with collinearity when dealing with huge datasets that have hundreds of attributes. Also, ML algorithms are quite effective at identifying patterns in datasets that humans might not be able to notice. More precise, unbiased, and automated personality assessments may result from these ML models.Social media is another instance, where users can express their likes, thoughts, feelings, and opinions. This data have been successfully used by machine learning algorithms to predict the "Big Five" (OCEAN) personality traits of individuals. Industries frequently employ supervised machine learning techniques to predict personality traits, such as Naive Bayes and Support Vector Machines. Moreover, unsupervised learning techniques are already being used by researchers to find additional psychological characteristics in digital data.

Machine learning algorithms have become a potential method for predicting personality based on OCEAN values in recent years. Due to their capacity to capture intricate correlations between input data and output labels, decision tree models in particular have been extensively used for this purpose. It is possible to create a predictive model that can correctly categorise people based on their personality traits by training a decision tree model on a dataset of OCEAN values and matching personality types.

In this study, we investigate the application of a decision tree model for OCEAN values-based personality prediction. In addition to describing the decision tree technique used to create the prediction model, we also go into the methodology used to gather and prepare the dataset. We also provide experimental findings that assess the model's performance in terms of accuracy and other metrics.

2. LITERATURE SURVEY

[1]
The article offers a thorough analysis of 30 studies that look at how the Big Five personality traits relate to resilience, with a focus on psychological resilience and ego-resilience. The results reveal that extraversion, openness, agreeableness, and conscientiousness have favourable associations with extraversion while resilience and neuroticism are strongly associated negatively. The study also shows that, compared to trait resilience, ego resilience has larger positive relationships with openness and agreeableness and stronger negative correlations with neuroticism. The effect sizes, notably for ego-resilience, differ across studies. The article also acknowledges the study's shortcomings, such as methodological issues, discrepancies between studies' conclusions, and the need for additional investigation to clarify the relationship between resilience and personality qualities. Overall, the analysis emphasizes the significance of taking into account personality traits when studying resilience and its components.

[2]
The method described in the article uses a convolutional neural network (CNN) to extract features from stream-of-consciousness essays in order to automatically identify a person's personality traits. Extroversion, neuroticism, agreeableness, conscientiousness, and openness are the Big Five personality qualities that the approach looks for using five separate networks, each of which has been trained to identify a binary value. By merging words into n-grams, n-grams into sentences, and sentences into an entire document, the text is represented as a variable-length vector, which is then fed into a fully connected neural network for classification. Preprocessing and filtering of the input data, feature extraction, and classification are all included in the procedure. It makes use of both per-word semantic characteristics as well as document-level aesthetic factors. By eliminating emotionally neutral input sentences from the essays, the technique enhances performance.

[3]
Machine learning models for personality psychology have been developed as a result of the accessibility of high-dimensional, fine-grained data on human behaviour through mobile sensing and digital footprints. But to use these models effectively, one must have advanced methodological expertise. This article presents a summary of the difficulties researchers encounter when developing, analyzing, and verifying machine learning models in personality psychology, highlighting the need of comprehending the fundamental ideas underlying these approaches. The evaluation of personality measures created using machine learning techniques and the use of latent variables in modeling are other topics covered by the writers. Finally, they give a forecast for how machine learning models may be applied in future personality study and evaluation. Even if automated, better, and more objective personality assessments may be possible using machine learning techniques, researchers must comprehend and solve the challenges to effectively and safely use these methods.

[4]
The application of machine learning in personality assessment is discussed, along with how it might improve our knowledge of personality traits and behavioural patterns. Using machine learning techniques, it is possible to anticipate individual differences in personality traits by finding empirical relationships between digital records and recognised personality trait measures. The research contends that machine learning's ability to produce fresh tools and insights into personality traits would be enhanced by incorporating it into a construct validation framework that takes into account content, structural, external, and discriminant validity. In its conclusion, the research makes detailed suggestions for incorporating machine learning methods for personality evaluation into a construct validation framework.

[5]
This study uses data mining and Support Vector Machine (SVM) to categorise Facebook users' personalities into one of the Big Five Personality Traits without having them complete any questionnaires. 170 Facebook users responded to the Big Five Inventory survey and agreed to have their information scraped. Using the Radial Basis Function (RBF) kernel, the best accuracy result was attained at 87.5%. The most widely used and reliable method for predicting a person's personality traits is the Big Five Personality Model. Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism make up the model. SVM was employed because it can generalise issues well and generate highly accurate results. Companies can utilise the findings of this study to determine the best human resources according to the fields they need. Overall, this research demonstrates the potential for using social media activity to predict personality traits accurately.

[6]
In sentiment analysis and affective computing, automated personality characteristic detection from text data has gained popularity. While earlier research concentrated on linguistic idioms and psycholinguistic databases, more recent developments have used transfer learning to use pre-trained language models. Elmo, ULMFiT, and BERT are used in the proposed deep learning-based model, which shows promising results in personality classification and prediction using data and classifier level fusion. Personality traits have an impact on lifestyle, health, and other goals, hence personality detection is an important field of study. Social networks produce enormous amounts of text data, making them a valuable resource for characterising personality traits. Recruiting, criminal investigations, health, and general well-being are all impacted by the ability to predict personal behaviour. Online text data can be used by NLP models to predict individual levels of personality traits, enhancing sentiment analysis, emotional computing, social network analysis, and recommendation systems.

[7]

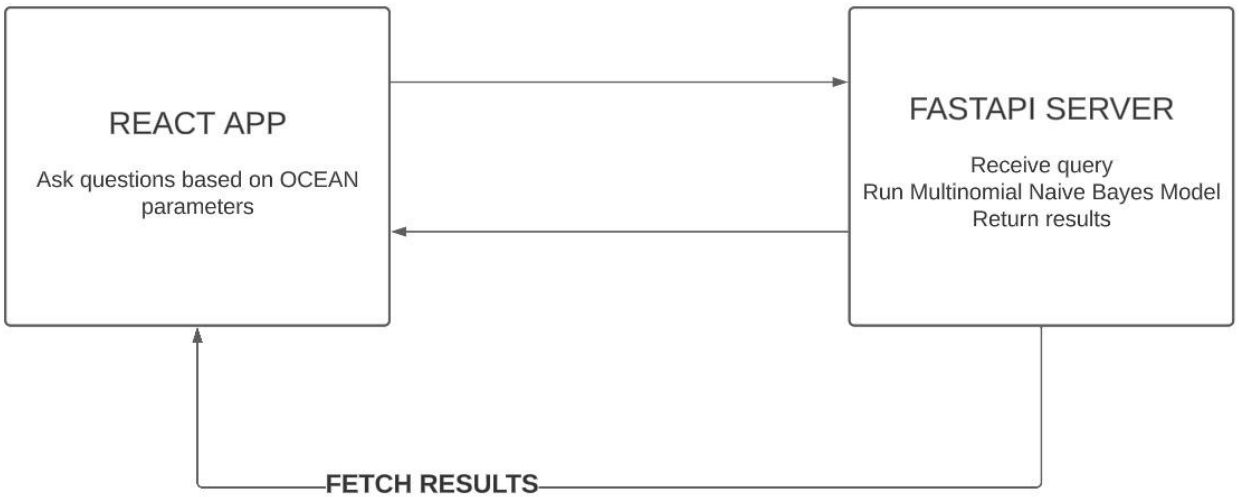
The notion that individuals with similar personalities will connect through social networks in ways that share common behavioural patterns is covered in the study. In order to build classifiers for recognising user personality by their profile without taking out any questionnaires, the researchers employed several data mining techniques to acquire information about the personality attributes of 100 Facebook users and their profiles. It was suggested that the boosting-decision tree be used as the model for identifying personality with an accuracy of 82.2%, which is better than earlier research that employed variables from user profiles. The research also explores how user interactions in online social networks can be used to track user personality, which is important for customising apps. The study makes a contribution by creating a machine learning method for anticipating user personality using a personality test in the social network and by examining five different personality variables, such as profile pictures in social groups. Future research in this area is discussed in the paper's conclusion.

[8]
The Myers-Briggs Type Indicator (MBTI) model is used in this study to suggest a machine learning-based method for automatically predicting four personality traits from text found on social networking sites. To solve the problem of class imbalance in the benchmark dataset, the study uses the XGBoost classifier and resampling approaches. In addition to highlighting the value of personality detection across a range of industries, the paper offers a solution that could help businesses better understand their consumers' personalities and preferences in order to hire the right employees and grow their businesses. All classifiers produce good results for various qualities, however the XGBoost classifier performs better than the others by obtaining more than 99% precision and accuracy. The study adds to the body of knowledge on personality recognition from social media content and makes recommendations for further research to enhance system performance.

[9]
Based on the Myers-Briggs Type Indicator (MBTI), one of the most widely used and trustworthy ways for determining personality types, this article introduces a new machine learning method for predicting personality types. The article states that Meta programs, which are accustomed ways of sorting and filtering information, are a crucial component in NLP and that Neuro Linguistic Programming (NLP) is a set of techniques used to find patterns in people's behaviour. The paper emphasises the value of meta programmes in identifying individual behavioural variations. The article also gives a brief overview of the history of how meta programmes were created and how the initial list of 60 patterns was reduced to a more manageable set of 14 patterns, which was then further condensed to the four fundamental meta programmes that are utilised in the MBTI. The new machine learning technique created in this study, according to the article's conclusion, performs better than existing techniques in terms of accuracy and dependability and can help NLP specialists and psychologists recognise personality types and related cognitive processes.

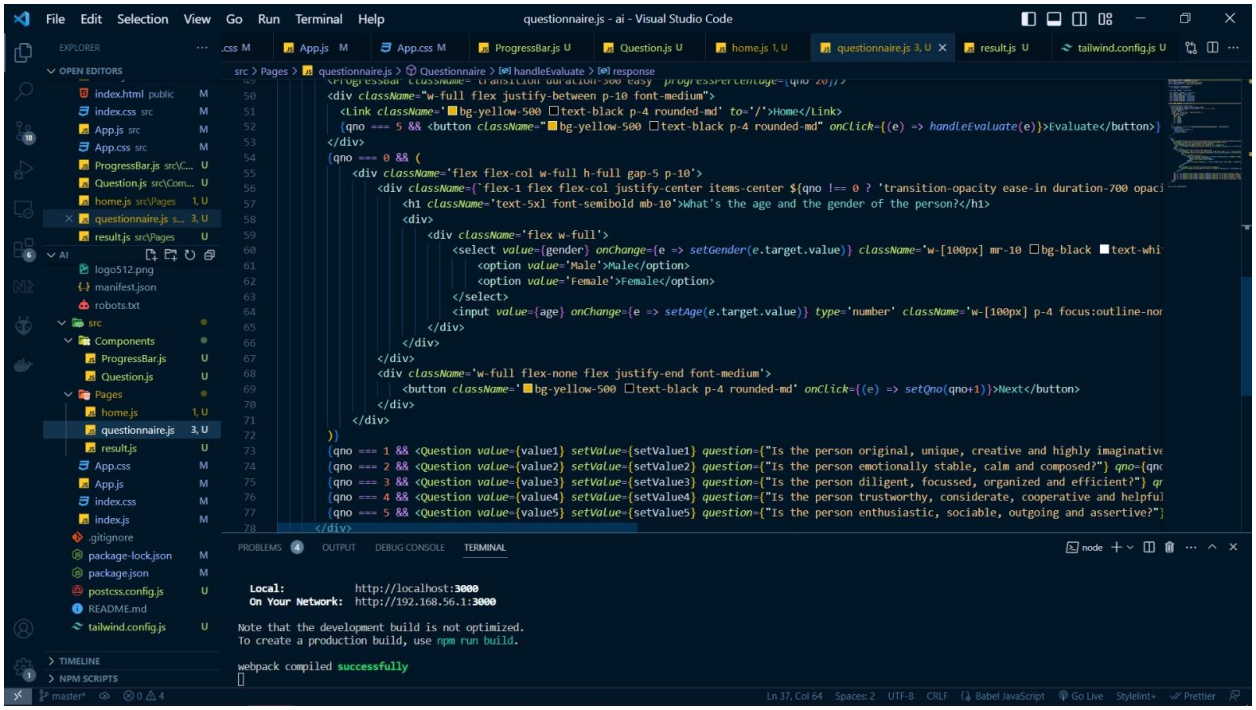
[10]This essay explores how personality qualities can be predicted via social media, particularly Twitter, without the use of surveys. The authors employed kNN, Naive Bayes, and SVM machine learning algorithms to categorise personalities based on the big five model from real-time Twitter data that was gathered via the streaming API. According to the study, Naive Bayes had the best accuracy, scoring 71.67%. Because social media offers a sizable dataset of usable data, using it to predict personality is advantageous. Personalized suggestions, hiring procedures, and sentiment analysis are just a few areas where personality prediction is crucial. The report ends with recommendations for additional research in this field.

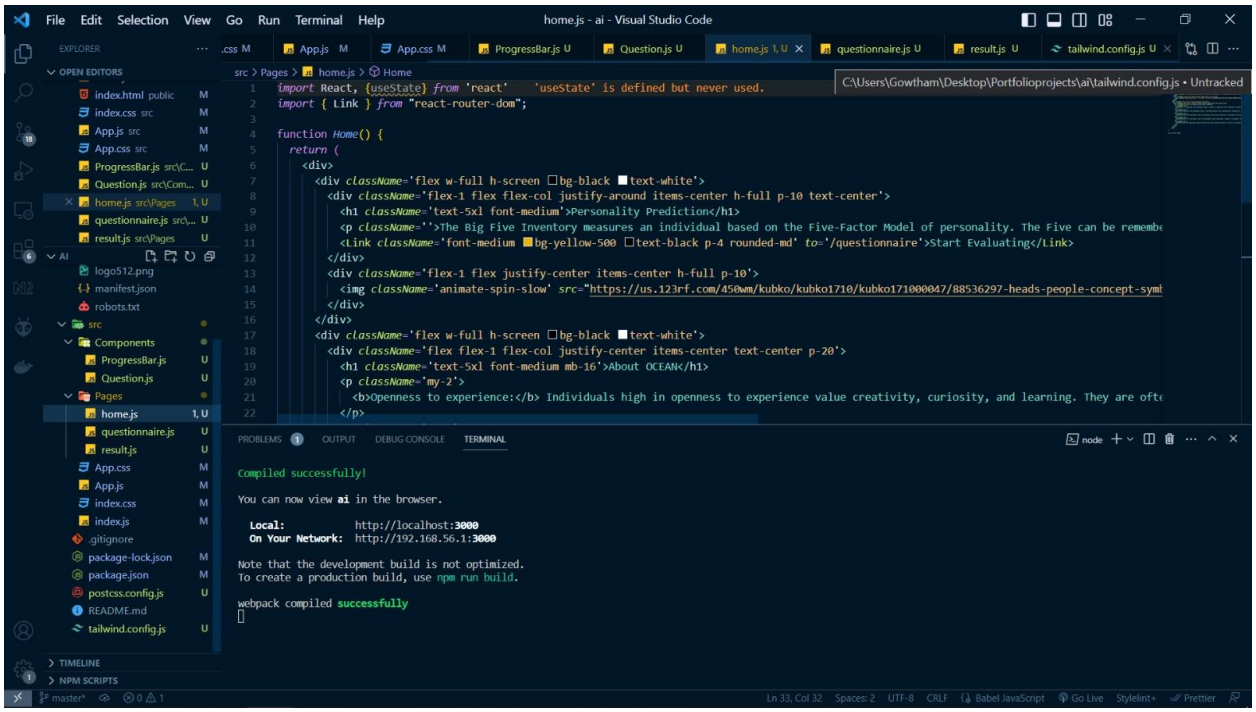
3. SYSTEM DESIGN

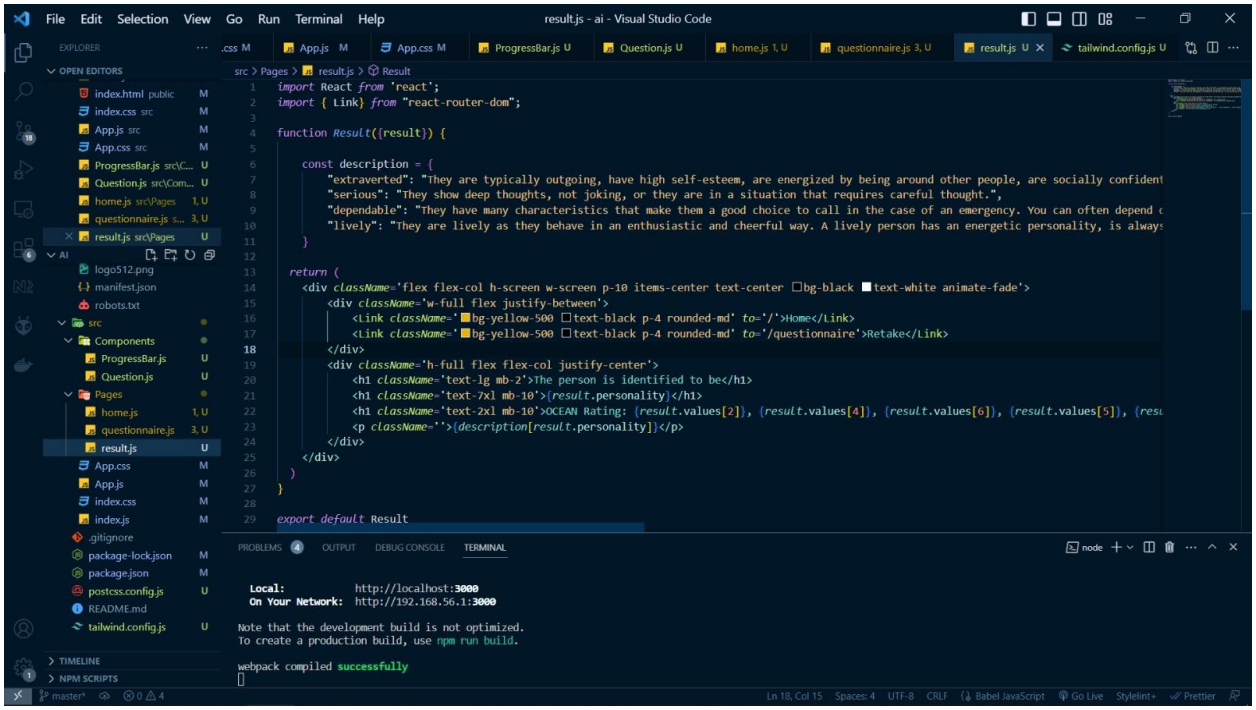


4. IMPLEMENTATION

ReactJs:


questionnaire.js - ai - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
OPEN EDITORS
AI
src
Components
Pages
TIMELINE
NPM SCRIPTS
App.js M
App.css M
ProgressBar.js U
Question.js U
home.js 1, U
questionnaire.js 3, U
result.js U
tailwind.config.js U
src > Pages > questionnaire.js > Questionnaire > handleEvaluate > response
PROBLEMS 4
OUTPUT
DEBUG CONSOLE
TERMINAL
Local: http://localhost:3000
On Your Network: http://192.168.56.1:3000
Note that the development build is not optimized.
To create a production build, use npm run build.
webpack compiled successfully
Ln 37, Col 64 Spaces: 2 UTF-8 CRLF Babel JavaScript Go Live Stylelint+ Prettier


home.js - ai - Visual Studio Code
File Edit Selection View Go Run Terminal Help
EXPLORER
OPEN EDITORS
AI
src
Components
Pages
TIMELINE
NPM SCRIPTS
App.js M
App.css M
ProgressBar.js U
Question.js U
home.js 1, U
questionnaire.js U
result.js U
tailwind.config.js U
src > Pages > home.js > Home
C:\Users\Gowtham\Desktop\Portfolioprojects\ai\tailwind.config.js • Untracked
'useState' is defined but never used.
PROBLEMS 1
OUTPUT
DEBUG CONSOLE
TERMINAL
Compiled successfully!
You can now view ai in the browser.
Local: http://localhost:3000
On Your Network: http://192.168.56.1:3000
Note that the development build is not optimized.
To create a production build, use npm run build.
webpack compiled successfully
Ln 33, Col 32 Spaces: 2 UTF-8 CRLF Babel JavaScript Go Live Stylelint+ Prettier



Machine Learning implementation:

```
df.corr()
```

	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
Gender	1.000000	0.152080	-0.046515	0.037262	-0.045458	-0.016257	0.011048	0.023087
Age	0.152080	1.000000	0.028371	-0.031243	0.053082	0.006113	0.048983	-0.009643
openness	-0.046515	0.028371	1.000000	0.014618	0.013744	-0.015003	-0.041610	-0.045000
neuroticism	0.037262	-0.031243	0.014618	1.000000	0.089673	-0.054809	-0.031488	-0.061480
conscientiousness	-0.045458	0.053082	0.013744	0.089673	1.000000	0.025333	0.028940	-0.014303
agreeableness	-0.016257	0.006113	-0.015003	-0.054809	0.025333	1.000000	-0.034821	-0.003419
extraversion	0.011048	0.048983	-0.041610	-0.031488	0.028940	-0.034821	1.000000	-0.006362
Personality (Class label)	0.023087	-0.009643	-0.045000	-0.061480	-0.014303	-0.003419	-0.006362	1.000000

```
lm = LogisticRegression( max_iter=1000)
sc = StandardScaler()
dt = DecisionTreeClassifier()
clf = SVC()

X_train, X_test, y_train, y_test = train_test_split(X, y, random_state=0, train_size=0.7)
```

✓ 0.0s

```
lm.fit(X_train, y_train)
lm.score(X_test, y_test)
```

```
0.2112676056338028
```

```
dt.fit(X_train, y_train)
dt.score(X_test, y_test)
```

```
0.2300469483568075
```

```
svm = SVC()
svm.fit(sc.fit_transform(X_train), y_train)
svm.score(sc.transform(X_test), y_test)
```

```
0.1643192488262911
```

```
mnb = MultinomialNB()
mnb.fit(X_train, y_train)
mnb.score(X_test, y_test)
```

✓ 0.1s

```
0.23943661971830985
```

```
from sklearn.linear_model import LogisticRegression
from sklearn.svm import SVC
from sklearn.tree import DecisionTreeClassifier
from sklearn.preprocessing import StandardScaler, LabelEncoder, PolynomialFeatures
from sklearn.decomposition import PCA
from sklearn.neighbors import KNeighborsClassifier
from sklearn.naive_bayes import MultinomialNB
from sklearn.cluster import MiniBatchKMeans
from sklearn.model_selection import train_test_split

import pandas as pd
import numpy as np
```

```
df = pd.read_csv('~/Desktop/Prog/AI/backend/training_dataset.csv')
df.drop(df.index[df['Gender'] == '5'].tolist(), inplace=True)

enc = LabelEncoder()
df['Gender'] = enc.fit_transform(df['Gender'])
#df['Personality (Class label)'] = enc.fit_transform(df['Personality (Class label)'])

X = df.drop(['Personality (Class label)'], axis=1)
y = df['Personality (Class label)']
```

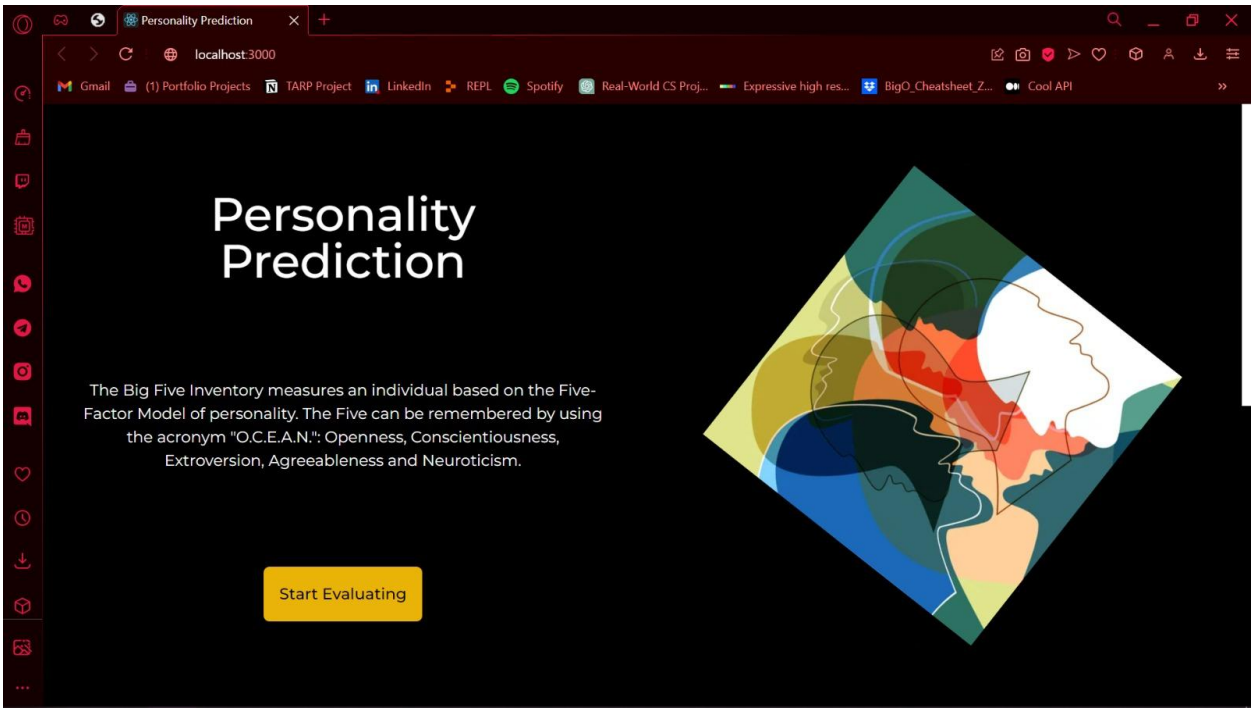
```
df.head()
```

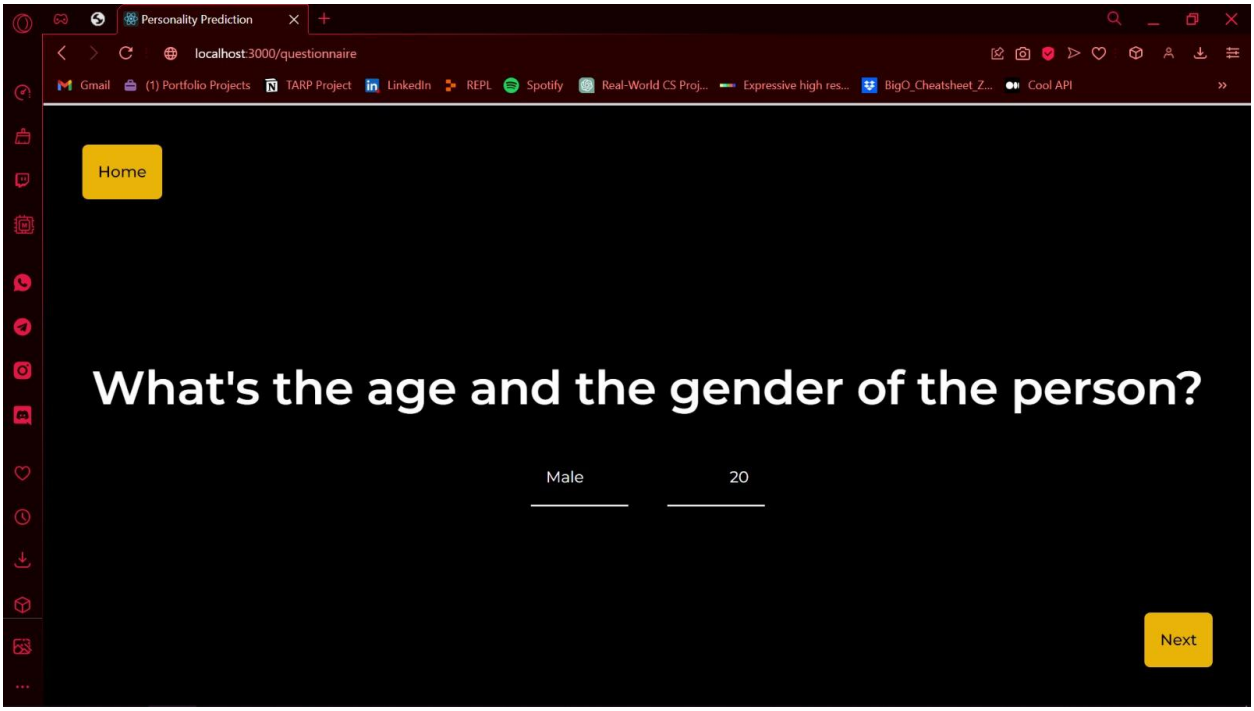
	Gender	Age	openness	neuroticism	conscientiousness	agreeableness	extraversion	Personality (Class label)
0	1	17	7	4	7	3	2	extraverted
1	1	19	4	5	4	6	6	serious
2	0	18	7	6	4	5	5	dependable
3	0	22	5	6	7	4	3	extraverted
4	0	19	7	4	6	5	4	lively

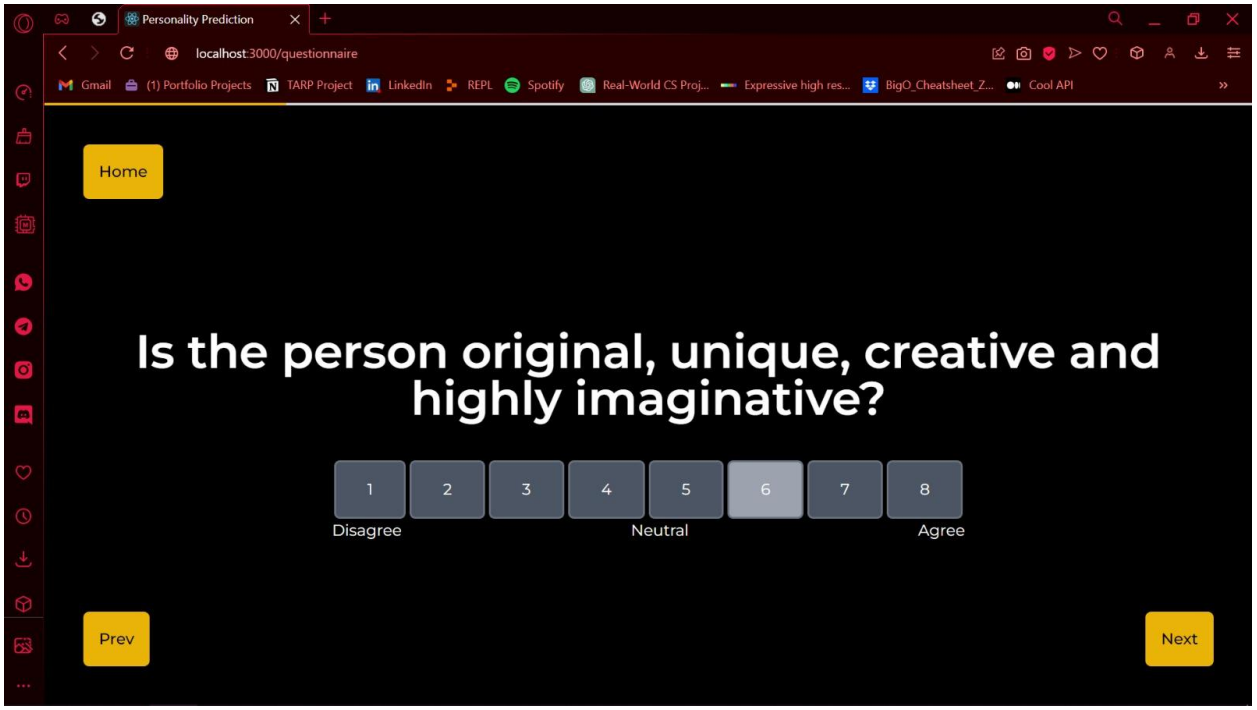
```python
from fastapi import FastAPI, Body, Request
from fastapi.middleware.cors import CORSMiddleware
import joblib, json, pickle

app = FastAPI()
app.add_middleware(
    CORSMiddleware,
    allow_origins=["*"],
    allow_methods=["*"],
    allow_headers=["*"]
)


@app.post("/")
async def get_predictions(val : Request):
    print(val)
    # call ml model to predict data
    # return data as json  with name, score for each value, and personality
    # maybe include personalised para

    # parse data here

    val = await val.json()
    # load and call model, encode
    val['gender'] = 1 if val['gender']=="Male" else 0
    print(val)

    values = list(val.values())
    print(values)
    res = predict(values)

    print(res)
    if res:
        final_res = {
            'values' : values,
            'personality' : res,
        }

        #final_res = json.dumps(final_res)

        print(final_res)
        return final_res


def predict(values):
    file = open('./lm.joblib', 'rb')
    model = joblib.load(file)
    res = model.predict([values])[0]
    file.close()
    return res

    #decode value to

@app.get('')
def hello():
    return {'abc'  : 'def'}
```

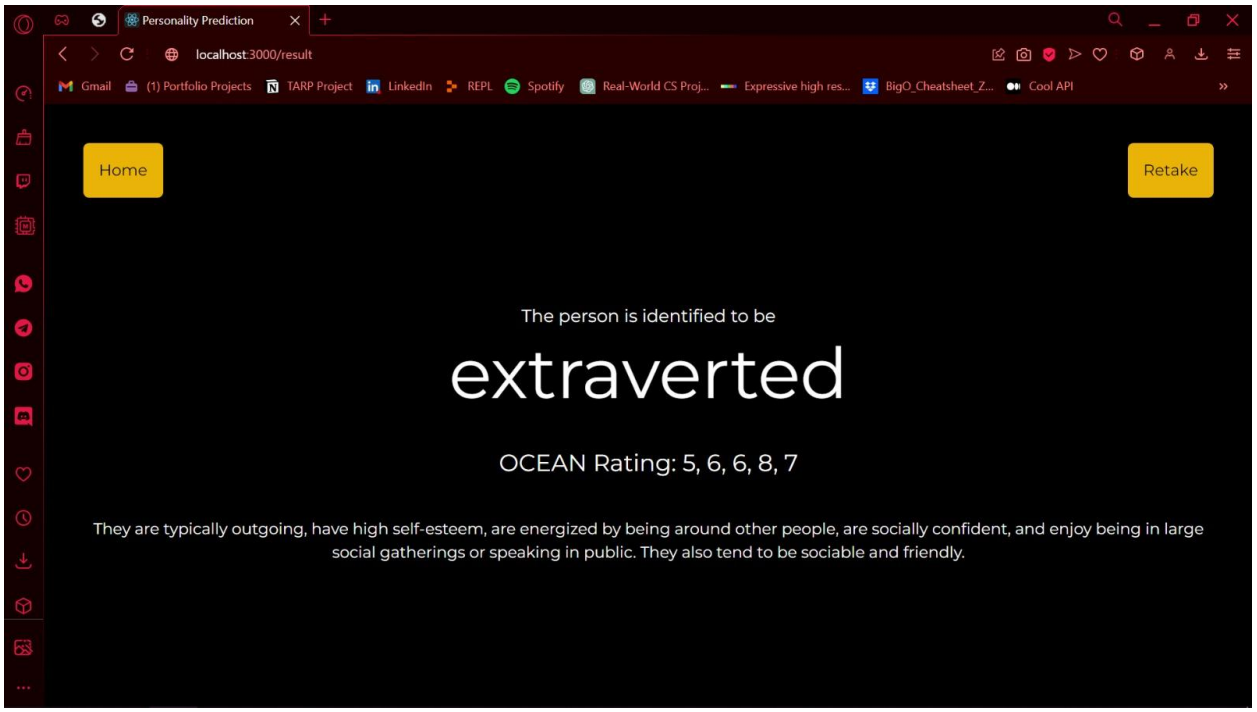
5. RESULTS AND DISCUSSION

Frontend:






Personality Prediction
localhost:3000/questionnaire
Gmail
(1) Portfolio Projects
TARP Project
LinkedIn
REPL
Spotify
Real-World CS Proj...
Expressive high res...
BigO_Cheatsheet_Z...
Cool API
Home
Is the person original, unique, creative and highly imaginative?
1
2
3
4
5
6
7
8
Disagree
Neutral
Agree
Prev
Next


Personality Prediction
localhost:3000/result
Gmail
(1) Portfolio Projects
TARP Project
LinkedIn
REPL
Spotify
Real-World CS Proj...
Expressive high res...
BigO_Cheatsheet_Z...
Cool API
Home
Retake
The person is identified to be
extraverted
OCEAN Rating: 5, 6, 6, 8, 7
They are typically outgoing, have high self-esteem, are energized by being around other people, are socially confident, and enjoy being in large social gatherings or speaking in public. They also tend to be sociable and friendly.

Backend:

backend > training_dataset.csv

```
Gender,Age,openness,neuroticism,conscientiousness,agreeableness,extraversion,Personality (Class label)
Male,17,7,4,7,3,2,extraverted
Male,19,4,5,4,6,6,serious
Female,18,7,6,4,5,5,dependable
Female,22,5,6,7,4,3,extraverted
Female,19,7,4,6,5,4,lively
Male,18,5,7,7,6,4,lively
Female,17,5,6,5,7,4,extraverted
Female,19,6,6,7,5,4,extraverted
Male,18,5,7,5,6,7,dependable
Female,19,5,5,7,4,5,lively
Male,19,6,7,5,6,3,serious
Male,19,7,6,7,7,6,extraverted
Male,19,7,6,6,5,6,lively
Female,19,6,7,5,5,5,dependable
Female,19,5,5,4,5,4,responsible
Male,19,5,6,4,6,3,extraverted
Female,19,7,7,2,6,5,serious
Female,18,6,7,4,4,2,dependable
```

The multinomial Naive Bayes algorithm-based machine learning model is used in the code for OCEAN values-based personality prediction. This algorithm, which is frequently employed for text classification issues, determines the likelihood that a specific characteristic belongs to a specific class based on how frequently it appears in the training data.

The OCEAN values, which stand for openness, conscientiousness, extraversion, agreeableness, and neuroticism, are the characteristics of this project. These values are taken from the input data, which may be text, audio, or any other kind of pertinent data. The Naive Bayes technique is used to determine the probability of each labelled OCEAN value for a given input once the model has been trained using a sizable dataset of OCEAN values. Based on the greatest probability values, the model then forecasts the individual's personality traits.

The Naive Bayes algorithm is implemented in the code using numerous Python libraries, including scikit-learn and numpy. These libraries offer resources for model training, prediction, and data processing.

6. CONCLUSION

Overall, this project's usage of the multinomial Naive Bayes method shows how well machine learning can be used to predict personality traits based on OCEAN values. The code offers a real-world illustration of how this technique can be applied to text categorization jobs and is easily adaptable to other comparable applications.
Our project's findings demonstrate that machine learning models can be used to predict personality using the OCEAN values. Based on the input data, our model was able to predict personality traits with a precision.

There are a number of potential industries where machine learning models for personality prediction could be used, including marketing, human resources, and psychology. Organisations may modify their communication and interactions to improve outcomes and produce better results by having a deeper grasp of each individual's personality.Overall, our effort shows the effectiveness and promise of machine learning in predicting personality traits and lays the groundwork for further investigation and advancement in this field. We think that this technology can open up fresh perspectives and understandings of human behaviour and create a wealth of exciting chances for innovation and advancement.

7. REFERENCES

[1]Oshio, A., Taku, K., Hirano, M., & Saeed, G. (2018). Resilience and Big Five personality traits: A meta-analysis. *Personality and Individual Differences*, *127*, 54–60. https://doi.org/10.1016/j.paid.2018.01.048

[2]*Deep Learning-Based Document Modeling for Personality Detection from Text*. (2017, April 1). IEEE Journals & Magazine | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/7887639

[3]Stachl, C., Pargent, F., Hilbert, S., Harari, G. M., Schoedel, R., Vaid, S. S., Gosling, S. D., & Bühner, M. (2020). Personality Research and Assessment in the Era of Machine Learning. *European Journal of Personality*, *34*(5), 613–631. https://doi.org/10.1002/per.2257

[4]Bleidorn, W., & Hopwood, C. J. (2019). Using Machine Learning to Advance Personality Assessment and Theory. *Personality and Social Psychology Review*, *23*(2), 190–203. https://doi.org/10.1177/1088868318772990

[5]Utami, N. F., Maharani, W., & Atastina, I. (2021). Personality Classification of Facebook Users According to Big Five Personality Using SVM (Support Vector Machine) Method. *Procedia Computer Science*, *179*, 177–184. https://doi.org/10.1016/j.procs.2020.12.023

[6]El-Demerdash, K., El-Khoribi, R. A., Shoman, M., & Ahmed, W. H. (2021). Deep learning based fusion strategies for personality prediction. *Egyptian Informatics Journal*, *23*(1), 47–53. https://doi.org/10.1016/j.eij.2021.05.004

[7]Souri, A., Hosseinpour, S., & Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality.

Human-centric Computing and Information Sciences, *8*(1). https://doi.org/10.1186/s13673-018-0147-4

[8]Khan, A. S., Hussain, A., Asghar, M. Z., Saddozai, F. K., Arif, A., & Khalid, H. A. (2020). Personality classification from online text using machine learning approach. International journal of advanced computer science and applications, 11(3).

[9]Amirhosseini, M. K., & Kazemian, H. B. (2020). Machine Learning Approach to Personality Type Prediction Based on the Myers–Briggs Type Indicator®. *Multimodal Technologies and Interaction*, *4*(1), 9. https://doi.org/10.3390/mti4010009

[10]*Analysis of Machine Learning Algorithms for Predicting Personality: Brief Survey and Experimentation*. (2019, October 1). IEEE Conference Publication | IEEE Xplore. https://ieeexplore.ieee.org/abstract/document/8978469