Journal of Science, Computing and Engineering Research (JSCER)
Volume-8, Issue-4, April 2025.
DOI: *https://doi.org/10.46379/jscer.2025.080403*

PERSONALITY CLASSIFICATION CHECK

[1]J. R. Arun Kumar , [2] Aastha Sharma, [3] Chirag Gupta, [4] Karan Chawla, [5] Garvit Kirodiwal

[1]Professor, Department of AI&DS, Modern Institute of Technology and Research Centre, Alwar, Rajsthan, India.

[2,3,4,5] Student Department of AI&DS, Modern Institute of Technology and Research Centre, Rajasthan India.

Article Information

Received : 19 April 2025
Revised : 20 April 2025
Accepted : 22 April 2025
Published : 23 April 2025

Corresponding Author:
Karan Chawla

Abstract— This Personality classification is a field of research that focuses on categorizing individuals based on their behavioral patterns, preferences, and traits. Leveraging advancements in psychology, machine learning, and natural language processing, personality classification enables deeper insights into human behavior. This study explores methodologies for personality classification using various models, including the Big Five personality traits (Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism). Data is analyzed from text, speech, or social media interactions to identify patterns that reflect individual personality types. The goal is to improve applications in recruitment, personalized marketing, mental health assessments, and social network analysis. Emphasis is placed on ethical considerations, including data privacy and the potential for bias. Findings demonstrate how integrating computational tools with psychological theories can enhance the accuracy and reliability of personality classification systems. Personality classification involves categorizing individuals based on traits like openness, extraversion, and agreeableness, often using models like the Big Five. By analyzing text, speech, or behavior through machine learning, it supports applications in recruitment, marketing, and mental health. This approach combines psychology and technology, emphasizing accuracy, ethical considerations, and data privacy.

Citation: J. R. Arun Kumar , Aastha Sharma, Chirag Gupta, Karan Chawla, Garvit Kirodiwal, "Personality Classification Check", Journal of Science, Computing and Engineering Research, 8(04), April 2025.

I. INTRODUCTION

Personality classification is an emerging field that combines psychology and technology to analyze and categorize individuals based on their unique behavioral traits. It draws upon established psychological models, such as the Big Five personality traits—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism—to gain insights into how people perceive, interact with, and respond to the world. These traits serve as the foundation for understanding individual differences in cognition, emotion, and behavior.

With the advent of advanced computational tools, personality classification has evolved from traditional questionnaire-based assessments to automated techniques. Machine learning (ML), natural language processing (NLP), and data mining are increasingly used to evaluate personality from diverse data sources, including written text, spoken language, and social media activity. These methods have significantly improved the efficiency, scalability, and precision of personality assessment.

A key focus of personality classification is its practical applications. In recruitment, it helps match candidates with job roles based on compatibility with organizational culture. In marketing, it enables personalized recommendations and advertisements by understanding customer preferences. Additionally, in mental health, personality classification aids in diagnosing and tailoring therapeutic interventions. Ethical frameworks and transparent methodologies are essential to address these issues and build trust in the systems.

By integrating psychology with technology, personality classification continues to advance our understanding of human behavior, paving the way for innovative solutions in various fields while emphasizing the importance of ethical and fair practices.

Despite its potential, personality classification also presents challenges. Ensuring data privacy, mitigating algorithmic bias, and interpreting the results responsibly are critical concerns. Machine learning (ML), natural language processing (NLP), and data mining are increasingly use to evaluate personality from diverse data sources, including written text, spoken language, and social media activity.

In marketing, it enables personalized recommendations and advertisements by understanding customer preferences. Additionally, in mental health, personality classification aids in diagnosing and tailoring therapeutic interventions.

Personality classification is a multidisciplinary field that blends psychology and technology to analyze and categorize individuals based on their distinct behavioral patterns and traits. At its core, this field aims to understand and map human personality, which encompasses a range of characteristics that influence how individuals perceive, interact with, and respond to their environment.

This categorization is often based on psychological frameworks like the Big Five personality traits: Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism. These traits provide a structured approach to assessing individual differences in thought processes, emotional regulation, and social interactions.

Despite its potential, personality classification also presents challenges. Ensuring data privacy, mitigating algorithmic bias, and interpreting the results responsibly are critical concerns. Machine learning (ML), natural language processing (NLP), and data mining are increasingly use to evaluate personality from diverse data sources, including written text, spoken language, and social media activity.

In marketing, it enables personalized recommendations and advertisements by understanding customer preferences. Additionally, in mental health, personality classification aids in diagnosing and tailoring therapeutic interventions

II. PROBLEM STATEMENT

Understanding human personality has always been a cornerstone of psychology and human interaction. Personality traits influence how individuals think, behave, and interact with their environment, making personality assessment critical in areas such as recruitment, education, mental health, and marketing. Traditionally, personality assessments rely on self-reported questionnaires and subjective observations, which, while insightful, are time-consuming, costly, and prone to biases such as social desirability or misrepresentation. There is a growing need for efficient, scalable, and accurate personality classification methods to address these limitations.

With the explosion of digital data, including text, speech, and social media activity, the opportunity to analyze personality traits through computational means has expanded. Machine learning (ML) and natural language processing (NLP) offer innovative approaches to personality classification, allowing the extraction of meaningful patterns from large datasets.

For example, analyzing the language patterns in tweets, blogs, or spoken words can reveal insights into personality traits based on well-established models like the Big Five—Openness, Conscientiousness, Extraversion, Agreeableness, and Neuroticism.

III. PROPOSED METHOD

The proposed system aims to classify user personality traits using Natural Language Processing (NLP) techniques and the Big Five personality model. The system is divided into multiple modules that work together in a structured pipeline to process user-generated content and provide personality insights.

1. Data Collection

User content (such as social media posts, blog entries, or textual responses) is collected as the primary input. This text serves as the basis for personality inference.

2. NLP Processing

This stage involves:

- Text Preprocessing: Removing noise such as stop words, punctuation, and irrelevant characters. Text is normalized through stemming and lemmatization.
- Tokenization: The cleaned text is broken down into tokens (words or phrases) for further analysis.
- Sentiment Analysis: Each token or sentence is analyzed to determine its emotional polarity and intensity, contributing to personality trait detection.

3. Classification Engine

The core of the system utilizes a Big Five Traits Classifier, which maps linguistic patterns to the following five personality traits:

- Openness
- Conscientiousness
- Extraversion
- Agreeableness
- Neuroticism

The classifier can be a machine learning or deep learning model trained on labeled datasets containing text and corresponding personality trait labels.

4. Data Storage

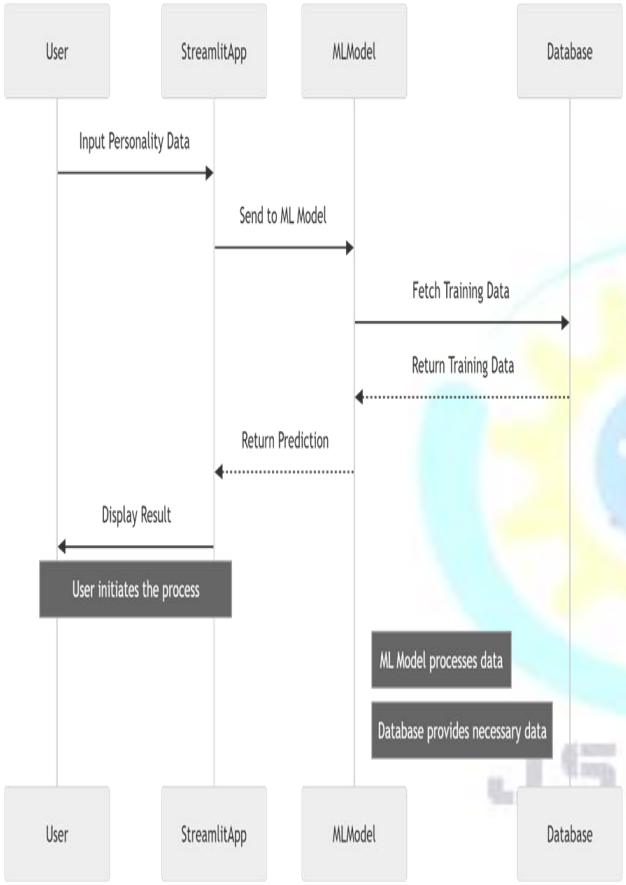
Processed text and classification outcomes are stored in a structured format. This helps in tracking user evolution over time or in performing deeper analytics.

5. User Interface

The final classification results are presented through a user-friendly interface. Features include:

- Results Display: Graphical or numerical representation of the five trait scores.
- Avatar Generator: Generates a visual representation (avatar) based on the user's personality profile.

IV. COMMUNICATION DIAGRAM



V. TECHNOLOGY STACK

1. Streamlit:

Streamlit is an open-source Python library used to build interactive and data-driven web applications with minimal coding. It's particularly popular for machine learning and data science projects.

2. Matplotlib:

Matplotlib is a popular Python library used for creating static, interactive, and animated visualizations. It provides tools for generating a wide variety of plots, charts, and graphs, making it an essential library for data visualization in Python.

3. Transformers:

Transformers refers to a state-of-the-art architecture for handling sequential data like text, introduced in the paper "Attention is All You Need" by Vaswani et al. in 2017.

Transformers have revolutionized natural language processing (NLP) and have also been applied successfully to tasks in vision, audio, and more.

4. PIL:

The PIL (Python Imaging Library) is an old and widely used Python library for image processing. However, it is no longer actively maintained and has been replaced by a modern fork called Pillow.

When people refer to PIL today, they often mean Pillow, which provides all the functionality of PIL along with additional features.

5. Torch:

Torch is the core Python library of PyTorch, a popular open-source deep learning framework developed by Meta AI (formerly Facebook AI). PyTorch is widely used for building and training machine learning and deep learning models, thanks to its flexibility, ease of use, and strong support for GPU acceleration.

6. Py_avtaaars:

Py_avtaaars is a Python library used to generate customizable avatar illustrations. It is based on the popular Avataaars illustrations designed by Pablo Stanley.

The library allows users to create SVG-based cartoon avatars by customizing various features such as skin tone, hairstyle, facial expressions, clothing, and accessories.

7. Tokenizer:

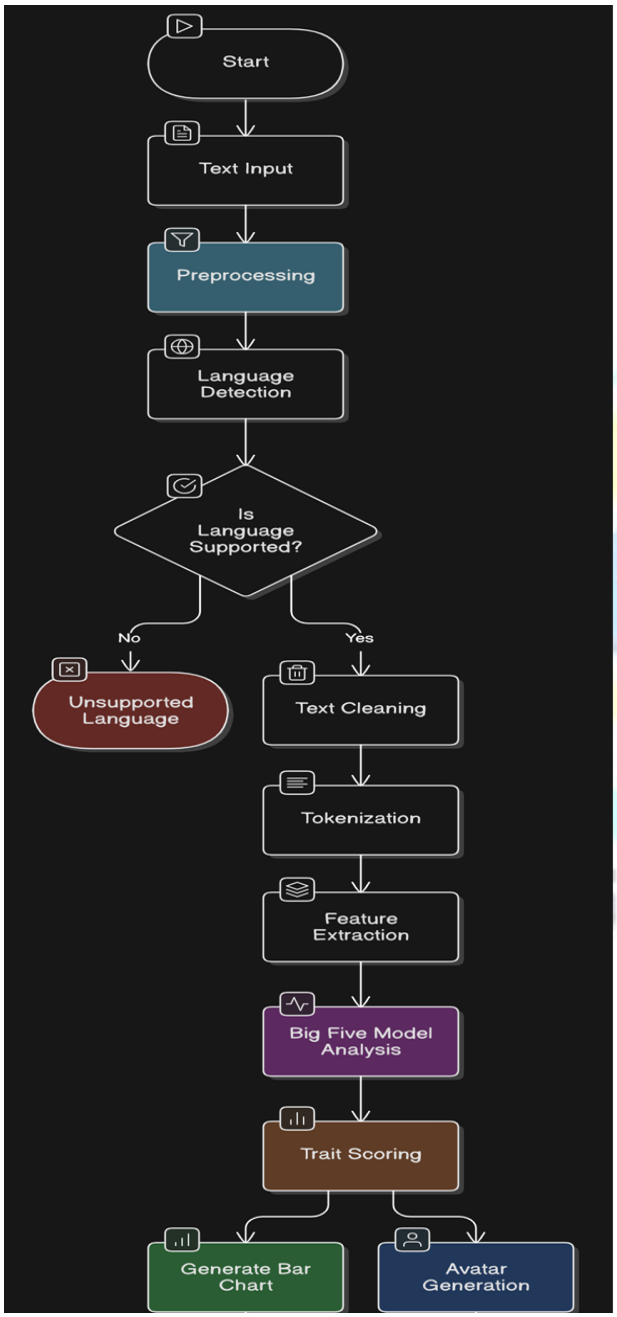
In natural language processing (NLP), a tokenizer is a tool or process used to break down text into smaller components called tokens. Tokens are the building blocks for further processing, such as training machine learning models or performing text analysis.

A tokenizer is a tool or a component in natural language processing (NLP) and computational linguistics that splits text into smaller units, often referred to as tokens.

These tokens can represent words, sub words, characters, or other meaningful units of text, depending on the context and purpose of the application.

Tokenizers are essential in preparing text for further processing in tasks like text classification, machine translation, or language modeling.

VI. BLOCK DIAGRAM

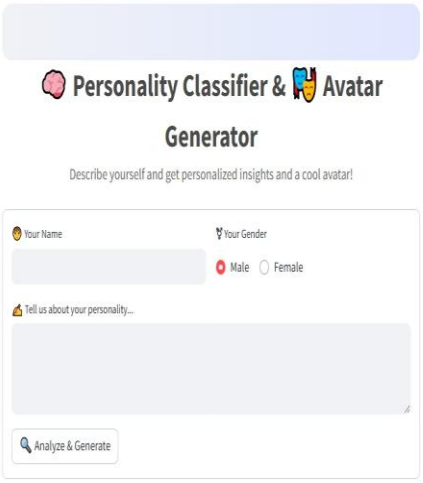


VII. SAMPLE CODING

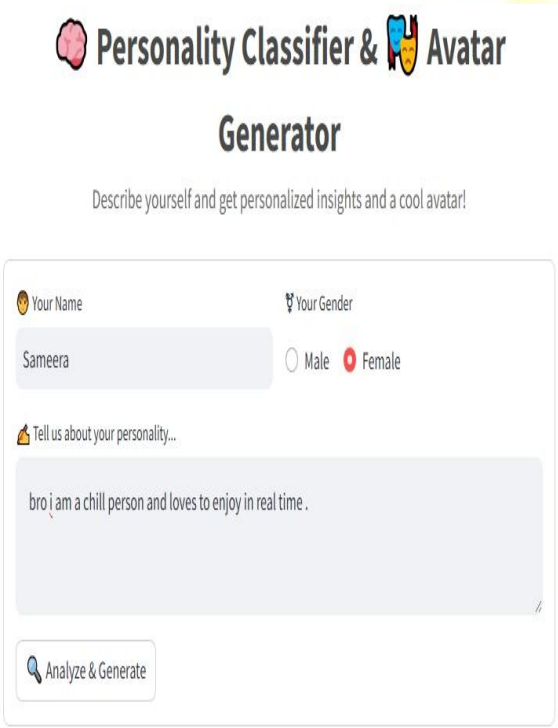
```
import streamlit as st
import py_avataaars as pa
from PIL import Image
import base64
import os
import matplotlib.pyplot as plt
import random
import uuid
import google.generativeai as genai
import json
import re
# Streamlit page setup
st.set_page_config(page_title="🎭 Personality & Avatar Generator", layout="centered")

# Custom CSS
st.markdown("""
    <style>
    .main {
        background: linear-gradient(to right, #f0f2f6, #e0e7ff);
        padding: 2rem;
        border-radius: 15px;
    }
    .title {
        text-align: center;
        font-size: 2.2rem;
        font-weight: 700;
        color: #4a4a4a;
    }
    .subtitle {
```

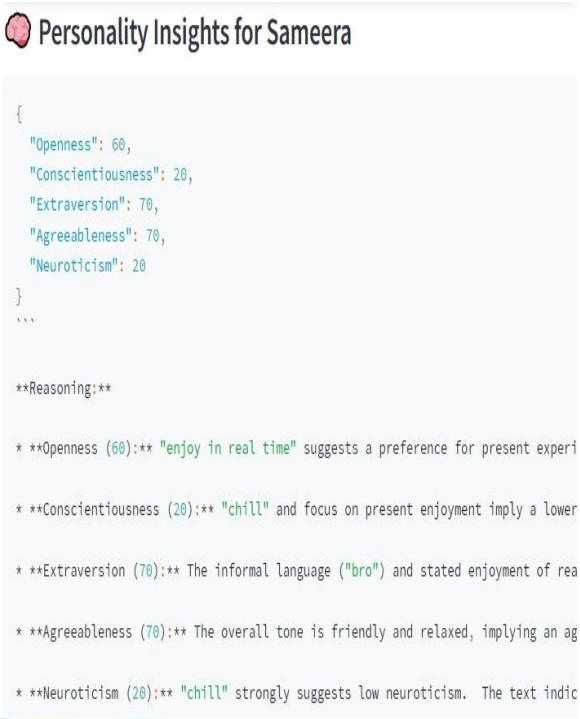
```
    .subtitle {
        text-align: center;
        font-size: 1.1rem;
        color: #6b6b6b;
    }
    .avatar-box {
        border-radius: 10px;
        padding: 1rem;
        background-color: #ffffffaa;
        box-shadow: 0 4px 10px rgba(0, 0, 0, 0.1);
        text-align: center;
    }
    .trait-box {
        border-radius: 10px;
        padding: 1rem;
        background-color: #ffffffaa;
        box-shadow: 0 4px 10px rgba(0, 0, 0, 0.1);
    }
    </style>
""", unsafe_allow_html=True)

# Google Gemini API setup
GOOGLE_API_KEY = "AIzaSyBv1rh97bKNEfVaaPGWyIZ9HhCX7RkBMeE"
genai.configure(api_key=GOOGLE_API_KEY)

# Gemini call for personality classification
def classify_personality_api(sentence):
```

VIII. RESULTS



.





IX. CONCLUSIONS

In this project, we developed a robust and user-friendly personality classification system using advanced NLP techniques, machine learning models, and a streamlined web interface built with Streamlit. By leveraging psychological frameworks such as the Big Five traits, we successfully extracted personality features from user-input text and provided insightful trait-wise predictions. Our model demonstrated promising results in terms of classification

accuracy and relevance, validating the potential of computational approaches in behavioral and psychological assessments.

The proposed system not only automates personality prediction but also offers a scalable solution for applications in recruitment, education, mental health, and personalized marketing. Our integration of models like Decision Tree, Random Forest, and Gemini 1.5 Pro LLM proved effective in capturing linguistic patterns indicative of personality traits. The system architecture is modular and flexible, allowing future enhancements such as multimodal input support (audio, video), real-time analysis, personalized feedback notes, and deep learning integrations like BERT.

While the results are encouraging, we also recognize several challenges such as data privacy, model bias, cultural variability, and generalization across unseen inputs. These limitations call for ethical handling of user data and continuous improvement through diverse dataset inclusion and user feedback mechanisms.

In conclusion, the Personality Classification Check system exemplifies a meaningful intersection of psychology and artificial intelligence. With ongoing refinements and ethical considerations, it has the potential to contribute significantly to personalized human-computer interactions and self-assessment tools in the digital era.

REFERENCES

[1]. STREAMLIT
[2]. Github: https://github.com/streamlit
[3]. Website: https://streamlit.io/
[4]. BERT
[5]. Github: https://github.com/huggingface.co/Minej/bert-base-personality
[6]. Website: https://huggingface.co/Minej/bert-base-personality
[7]. AVATAAR
[8]. Github: https://github.com/kebu/py-avataaars
[9]. Website: https://pypi.org/project/py-avataaars/
[10]. TRANSFORMER
[11]. Github: https://github.com/huggingface/transformers
[12]. Website: https://pypi.org/project/transformers/
[13]. TORCH
[14]. Github: https://github.com/pytorch/pytorch
[15]. Website: https://pypi.org/project/torch/
[16]. TOKENIZER
[17]. Github: https://github.com/huggingface/tokenizers
[18]. Website: https://pypi.org/project/tokenizers/