iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 13 Issue: IV Month of publication: April 2025

DOI: https://doi.org/10.22214/ijraset.2025.69336

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

Personality Prediction Using Machine Learning and Social Media Data: A Myers-Briggs Approach

Nakul Bhangale[1], Aryan Jasuja[2], Vraj Gujrathi[3]
Department of Computer Engineering Pune Institute of Computer Technology, Pune, India

Abstract: With the explosive growth of user-generated content on social media, there is a rising interest in utilizing digital footprints to infer personality traits. This study explores how social media text can be analyzed using Natural Language Processing (NLP) and machine learning to predict an individual's personality, focusing on the Myers-Briggs Type Indicator (MBTI) framework. Leveraging linguistic and behavioral features extracted from social media content, we apply Support Vector Machines (SVM) and Random Forest algorithms to classify users into MBTI personality types. Our approach highlights the scalability and potential of automated personality assessment in domains such as targeted marketing, recruitment, and mental health.

Index Terms: MBTI, Machine Learning, Natural Language Processing, Social Media, Support Vector Machine, Random Forest

I. INTRODUCTION

Personality analysis has long played a vital role in fields such as psychology, behavioral science, human-computer in- teraction, and marketing, providing deep insights into individual behavior, decision-making, and social preferences [1]. Tra- ditional personality assessment methods—such as the Myers- Briggs Type Indicator (MBTI)—require self-report question- naires and formal testing environments. While widely adopted in personal and professional development, these techniques are often time-consuming, subjective, and not scalable for real- time applications [2].

With the explosion of user-generated content on social media platforms, digital traces of personality are now more accessible than ever. Users express themselves through informal text, emojis, hashtags, and behavior patterns, providing a rich source of data for personality analysis. This has enabled a shift from manual psychological testing to computational methods for personality inference. Studies have shown that linguistic cues such as word usage, sentiment, and syntactic patterns correlate strongly with personality traits [3], [1].

Although a large portion of prior work has focused on the Big Five personality model (OCEAN), recent studies have turned toward MBTI-based classification due to its categorical and interpretable nature. The MBTI framework segments indi- viduals into 16 personality types based on four psychological dichotomies: Introversion/Extraversion (I/E), Sensing/Intuition (S/N), Thinking/Feeling (T/F), and Judging/Perceiving (J/P). This makes MBTI particularly well-suited for supervised ma- chine learning classification tasks [4].

In this study, we propose a scalable machine learning pipeline to classify users into MBTI types based on features extracted from social media text. Our approach involves advanced natural language processing (NLP) techniques such as lexical normalization, sentiment analysis, and part-of-speech tagging, followed by supervised classification using Support Vector Machines (SVM) and Random Forest algorithms. Unlike deep learning-based approaches, our framework emphasizes interpretability and performance on smaller, noisy datasets, which are common in real-world social media scenarios.

This research aims to explore the effectiveness of classical machine learning models for personality classification, identify linguistically significant features linked to MBTI traits, and assess how well models generalize across personality dimen- sions. Potential applications include personalized recommendation systems, user profiling, and mental health diagnostics [5].

II. CORE COMPONENTS OF SUPERVISED LEARNING

Supervised learning is a foundational paradigm in machine learning where models are trained on labeled data to learn the mapping between input features and output labels. In the context of personality prediction from social media, the goal is to accurately classify a user's MBTI personality type based on patterns found in their textual posts. This requires two pri- mary components: a well-prepared dataset with corresponding MBTI labels and a set of extracted features that meaningfully represent a user's linguistic and behavioral characteristics.

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 13 Issue IV Apr 2025- Available at www.ijraset.com

The MBTI classification problem is inherently a *multi-label, multi-class problem.* Each individual is characterized by four binary decisions across the MBTI dimensions: Introversion (I) vs. Extraversion (E), Sensing (S) vs. Intuition (N), Thinking (T) vs. Feeling (F), and Judging (J) vs. Perceiving (P). This setup allows us to approach the task as four independent binary classification problems—making it well-suited for traditional supervised learning models such as Support Vector Machines and Random Forest classifiers.

A. Feature Extraction and Representation

A crucial step in supervised learning is the construction of meaningful input features. For personality prediction, these features are extracted from users' social media posts using Natural Language Processing (NLP) techniques. Commonly used textual features include:

- TF-IDF (Term Frequency-Inverse Document Fre- quency): Captures the importance of words relative to a document corpus.
- Lexical and Syntactic Features: Word count, sentence length, punctuation usage, and POS distributions.
- Sentiment Scores: Derived using sentiment analysis models to capture emotional tone.
- Behavioral Features: Posting frequency, engagement time, and temporal patterns.

These features are normalized and vectorized to prepare them for training in supervised learning models.

B. Support Vector Machine (SVM)

SVM is a powerful classifier known for its effectiveness in high-dimensional spaces. It seeks an optimal hyperplane that separates data points of different classes while maximizing the margin between them. The objective of a linear SVM is:

$$\min_{w,b} \frac{1}{2}\|w\|^2 \quad \text{subject to} \quad y_i(w^T x_i + b) \geq 1$$

For non-linear classification, kernel functions such as RBF or polynomial kernels are employed to map data into higher-dimensional spaces where separation is possible.

SVMs are especially effective for:

- High-dimensional and sparse data (e.g., TF-IDF vectors)
- Binary classification tasks like those in MBTI dimensions
- Scenarios with smaller datasets where overfitting must be minimized

C. Random Forest Classifier

Random Forest is an ensemble learning technique that builds multiple decision trees and aggregates their results for final classification. Each tree is trained on a random subset of the data and features, which helps to reduce variance and overfitting.

For a given input x, the final prediction $y\hat{}$ is obtained by majority voting across k decision trees:

$$y\hat{} = \text{mode}(\{T_1(x), T_2(x), \ldots, T_k(x)\})$$

Advantages of Random Forest include:

- Robustness to noisy and irrelevant features
- Built-in feature importance estimates
- Strong performance with minimal hyperparameter tuning

D. Binary Classification for MBTI Dimensions

To improve accuracy and interpretability, MBTI prediction is divided into four binary classification tasks:

1) Introversion (I) vs. Extraversion (E)
2) Sensing (S) vs. Intuition (N)
3) Thinking (T) vs. Feeling (F)
4) Judging (J) vs. Perceiving (P)

Each classifier is independently trained, enabling tailored optimization for each psychological dimension.

E. Evaluation Metrics

The following metrics are used to evaluate model perfor- mance:

- Accuracy: Overall percentage of correct predictions.
- Precision, Recall, F1-Score: Useful for imbalanced datasets.
- Confusion Matrix: Visualizes classification outcomes.
- ROC-AUC: Measures separability between classes.

Hyperparameter tuning (e.g., kernel type in SVM, number of estimators in Random Forest) is conducted using grid search and *k*-fold cross-validation to ensure generalization and reduce overfitting.

III. RELATED WORK

The application of machine learning techniques for per- sonality prediction has gained significant attention in recent years, driven by the explosion of user-generated content on social media platforms and the need for scalable, real-time personality inference. Traditional personality assessment meth- ods, such as the Myers-Briggs Type Indicator (MBTI), have been widely used in corporate, educational, and psychological contexts, but they often require manual administration and are subject to social desirability bias. To address these limitations, researchers have explored automated personality prediction models using linguistic features and behavioral cues extracted from online data sources.

Golbeck et al. [1] were among the first to demonstrate the viability of predicting personality traits from social media data. Their study showed that linguistic and interaction patterns on Facebook could be used to infer Big Five personality traits with significant accuracy. Although their work focused on the OCEAN model, it paved the way for applying similar approaches to MBTI classification.

Plank and Hovy [3] introduced a Twitter-based dataset labeled with MBTI types and applied both Support Vector Machines and logistic regression for classification. They found that n-gram-based features and psycholinguistic attributes derived from LIWC (Linguistic Inquiry and Word Count) improved classification performance, particularly for the In- troversion/Extraversion dimension.

Park et al. [10] explored automatic personality classification using a combination of topic modeling and deep learning. They used Latent Dirichlet Allocation (LDA) to identify latent topics from user-generated content and integrated them with Convolutional Neural Networks (CNNs) to predict MBTI traits. Their model achieved promising results, highlighting the potential of topic-aware personality classifiers.

Verhoeven et al. [2] developed TwiSty, a large dataset of over 1.2 million tweets annotated with MBTI types. Using a range of supervised learning models including Random Forests and SVMs, they emphasized the difficulty of generalizing MBTI prediction across diverse user profiles. Their findings also stressed the importance of handling class imbalance, a recurrent challenge in MBTI classification due to unequal distribution of personality types recent efforts have integrated transformer-based models such as BERT to enhance prediction accuracy. Yamada et al. [4] fine-tuned a multilingual BERT model on MBTI-labeled text data and demonstrated that pre-trained language models can capture subtle semantic cues linked to psychological types. Despite their higher computational cost, such models have sig- nificantly outperformed traditional bag-of-words approaches in many classification benchmarks. In parallel, studies like those by Filardi et al. [5] have analyzed the interpretability of machine learning models in the MBTI domain. By evaluating feature importance in Random Forests and attention mechanisms in deep learning models, they attempted to correlate linguistic markers with MBTI dichotomies, providing psychological insights in addition to classification results. These studies collectively affirm that personality traits—especially those defined in the MBTI framework can be inferred using machine learning methods applied to linguistic and behavioral features. However, challenges such as data sparsity, imbalance, and generalizability remain prevalent. Our work extends this research by applying classical yet robust algorithms (SVM and Random Forest) on a carefully preprocessed social media dataset, focusing on model explainability and performance across each MBTI dimension.

IV. PROBLEM FORMULATION

A. Overview of the MBTI Personality Prediction Problem

The goal of this study is to predict an individual's MBTI personality type based on their social media activity, specif- ically textual content such as posts or tweets. The Myers- Briggs Type Indicator (MBTI) classifies individuals into 16 types based on four dichotomous dimensions: Introversion vs. Extraversion (I/E), Sensing vs. Intuition (S/N), Thinking vs. Feeling (T/F), and Judging vs. Perceiving (J/P). This prediction task can thus be decomposed into four binary classification sub-tasks, where each classifier predicts one psychological dimension.

Given a set of users $U = \{u_1, u_2, ..., u_n\}$ and a corre- sponding set of social media documents $D = \{d_1, d_2, ..., d_n\}$ authored by these users, the objective is to learn a mapping function:

$$f : D \to Y$$

where $Y = \{(y^{IE}, y^{SN}, y^{TF}, y^{JP})\}$ represents the MBTI label vector of four binary decisions for each user.

B. Data Representation

Each document d_i is preprocessed to generate a feature vector $x_i \in R^m$, where m is the number of features. These features are derived through various NLP techniques such as:

- TF-IDF vectorization: Capturing word frequency dy- namics.
- Lexical features: Sentence length, word diversity, punc- tuation.
- Sentiment and emotional tone: Using pre-trained sen- timent analyzers.
- Temporal and behavioral patterns: Posting frequency and time.

The final dataset can thus be denoted as:

$$\mathrm{D} = \{(x_1, y_1), (x_2, y_2), ..., (x_n, y_n)\}$$

where y_i is the MBTI label vector for user u_i.

C. Decomposition into Binary Classification Tasks

To handle the complexity of 16 MBTI types, we divide the problem into four independent binary classification tasks. Each task corresponds to predicting one MBTI dimension:

Task 1: $f_1(x) \to \{I, E\}$, Task 2: $f_2(x) \to \{S, N\}$

Task 3: $f_3(x) \to \{T, F\}$, Task 4: $f_4(x) \to \{J, P\}$

Each function f_i is trained using supervised learning algo- rithms (e.g., SVM, Random Forest), and evaluated indepen- dently.

D. Learning Objective

The main goal of this project is to accurately predict a person's MBTI personality type by minimizing classification errors. Each MBTI dimension (e.g., Introversion vs. Extraver- sion) is treated as a separate binary classification task. For each classifier, we use cross-entropy loss to measure how well the model predicts the correct class. The total loss is the sum of the losses from all four classifiers, one for each MBTI axis. During prediction, the model outputs a personality type by combining the results of these four classifiers, forming a complete MBTI label for the user.

E. Assumptions and Constraints

To ensure tractability and consistency in modeling, the following assumptions are made:

- Each user is represented by sufficient and diverse textual data (minimum document length threshold).
- MBTI labels are assumed to be accurate as per user self- declaration.
- No temporal drift in language patterns is assumed within a user's data sample.
- Class imbalance is handled using oversampling or class weighting techniques.

These constraints help in building a robust and generalizable model that can perform well across unseen user profiles.

V. METHODOLOGY

This section outlines the overall methodology adopted to build and evaluate our personality prediction framework based on the Myers-Briggs Type Indicator (MBTI). The framework follows a supervised learning pipeline consisting of data collection, preprocessing, feature engineering, model training, and performance evaluation. We deploy Support Vector Ma- chines (SVM) and Random Forest classifiers for each of the four MBTI binary classification tasks: I/E, S/N, T/F, and J/P.

A. Data Collection and Preprocessing

We utilize a dataset sourced from social media platforms, where users self-reported their MBTI personality types. The dataset includes user posts, typically aggregated as a single document per user. The preprocessing steps are critical to normalize and clean noisy, informal social media text.

The preprocessing pipeline includes:
- Lowercasing and punctuation removal
- Stopword filtering using NLTK's stopword corpus
- Tokenization and stemming using PorterStemmer
- Lexical normalization to handle out-of-vocabulary (OOV) words
- Removal of MBTI mentions from user data to avoid data leakage

Only users with at least 500 words in their corpus were retained to ensure sufficient linguistic data for personality inference.

B. Feature Extraction

We extract a combination of lexical, syntactic, and semantic features from user documents:
- TF-IDF Features: Using unigrams and bigrams with a vocabulary size capped at 10,000.
- Linguistic Style Features: Average word length, sen- tence count, question marks, use of first-person pronouns.
- Sentiment Features: Polarity and subjectivity scores using the TextBlob sentiment analyzer.
- Part-of-Speech Tags: Distribution of nouns, verbs, ad- jectives, and adverbs.

These features are concatenated into a single high- dimensional feature vector for each user.

C. Model Architecture

We train separate classifiers for each of the four MBTI dimensions. The two algorithms used are:
- Support Vector Machine (SVM): Implemented using the scikit-learn SVC module. We evaluate both linear and radial basis function (RBF) kernels. Grid search is used to select optimal values for the regularization parameter C and kernel coefficient γ.
- Random Forest **Classifier:** Trained using 100–200 decision trees, with maximum tree depth and feature subset size optimized through cross-validation. Feature impor- tance scores are also extracted to understand the most influential linguistic traits.

Each classifier is trained independently on its respective binary task (e.g., I vs. E), allowing for modular optimization and interpretability.

D. Training and Evaluation Strategy

To ensure robust performance, we use a stratified 10-fold cross-validation strategy. This guarantees that each fold has an equal proportion of classes, helping address class imbalance.

The training workflow includes:
- Splitting data into 80% training and 20% test sets.
- Applying grid search with cross-validation to tune hy- perparameters.
- Measuring classification metrics such as accuracy, pre- cision, recall, and F1-score.
- Generating confusion matrices and ROC curves for each classifier.

We also perform oversampling on the minority classes using SMOTE (Synthetic Minority Over-sampling Technique) to mitigate skewed label distribution.

E. Implementation Details

- Language: Python 3.9
- Libraries: Scikit-learn, Pandas, Numpy, Matplotlib, NLTK, TextBlob
- Hardware: Intel Core i5, 16GB RAM, running Ubuntu 22.04

The complete experimentation was conducted in a con- trolled Jupyter Notebook environment, allowing easy visual- ization and tracking of model performance metrics.

VI. RESULTS AND DISCUSSION

This section presents the experimental results of our MBTI personality prediction framework using Support Vector Ma- chine (SVM) and Random Forest classifiers. We evaluate each binary classification task independently—corresponding to the four MBTI dimensions: I/E, S/N, T/F, and J/P. Performance is assessed using accuracy, precision, recall, F1-score, and ROC- AUC.

A. *Overall Classification Performance*

Table I summarizes the average classification metrics across the four MBTI dimensions for both SVM and Random For- est. All metrics are averaged over a stratified 10-fold cross- validation.

TABLE I
OVERALL CLASSIFICATION METRICS (AVERAGED ACROSS ALL MBTI DIMENSIONS)

Model	Accuracy	Precision	Recall	F1-Score	ROC-AUC
SVM (Linear)	0.71	0.70	0.69	0.69	0.76
SVM (RBF)	0.73	0.72	0.71	0.71	0.78
Random Forest	**0.77**	**0.76**	**0.75**	**0.76**	**0.82**

The Random Forest model outperforms both SVM variants across all metrics, achieving an average accuracy of 77% and F1-score of 0.76. These results highlight the model's ability to generalize well across multiple MBTI personality dimensions while maintaining strong predictive power.

B. *Performance by MBTI Dimension*

A breakdown of classification accuracy for each MBTI dimension is shown in Table II. The Introversion/Extraversion (I/E) axis yields the highest accuracy across all models, likely due to the pronounced linguistic and behavioral contrast between introverts and extroverts.

TABLE II
CLASSIFICATION ACCURACY PER MBTI DIMENSION

Dimension	Random Forest	SVM (RBF)	SVM (Linear)
Introversion / Extraversion	0.82	0.79	0.76
Sensing / Intuition	0.74	0.70	0.68
Thinking / Feeling	0.76	0.72	0.71
Judging / Perceiving	0.78	0.73	0.72

The Random Forest model consistently outperforms SVM across all MBTI axes, with the I/E and J/P dimensions yielding particularly high accuracy. This indicates the model's ability to capture both linguistic nuance and stylistic patterns indicative of personality.

C. *Feature Importance and Interpretation*

The Random Forest model also enables interpretability through its feature importance rankings. Notable influential features include:

- Use of personal pronouns (e.g., "I", "me"): Highly indicative of introversion.
- Sentiment polarity: More extreme values often associ- ated with feeling-oriented users.
- Use of emojis, exclamation marks, and sentence length: Correlated with extraversion and judging traits.
- Verb-to-noun ratio: Showed significance in distinguish- ing sensing vs. intuition.

These features align with established psychological insights and confirm the model's interpretability and validity in the MBTI context.

D. *Discussion*

The results indicate that Random Forest is a highly effective model for MBTI-based personality prediction using social media text. Its ensemble nature allows it to generalize well across heterogeneous language patterns while maintaining resistance to overfitting.

While SVM with an RBF kernel offers competitive perfor- mance, it lacks the interpretability and robustness observed in the Random Forest model. The performance gap is especially noticeable in the S/N and T/F dimensions, where Random Forest's handling of non-linear interactions between features proves advantageous.

However, certain challenges persist:

- Class imbalance: MBTI types like INFJ and ESTP are underrepresented, leading to skewed predictions.
- Ambiguity in language: Sarcasm, humor, and multilin- gualism reduce feature reliability.
- Overlap in user behavior: Some personality traits may manifest similarly in text, making them harder to distin- guish.

Future work could explore hybrid architectures (e.g., RF + BERT embeddings), incorporate behavioral metadata (likes, shares), or apply transfer learning for smaller personality subgroups.

E. Model Selection and Training

We employ Support Vector Machines (SVM) and Random Forest classifiers. SVM is known for high accuracy in high- dimensional text data, while Random Forest offers better interpretability and robustness to overfitting.

The 16 MBTI types are decomposed into 4 binary classifi- cation tasks, one for each dimension (I/E, S/N, T/F, J/P). Each classifier is trained separately.

F. Evaluation and Optimization

Grid search and cross-validation are used to optimize hyperparameters. Accuracy, precision, recall, and F1-score are computed for each classification task to measure model performance.

VII. CHALLENGES AND LIMITATIONS

1) Data Quality: MBTI-labeled datasets from social media are scarce and often self-reported, leading to potential bias.
2) Text Complexity: Social media text includes sarcasm, slang, and cultural references that are hard to model.
3) Class Imbalance: Certain MBTI types are underrepre- sented, affecting classifier performance.
4) Interpretability: MBTI types can be nuanced and in- fluenced by context, complicating direct mapping from features.

VIII. CONCLUSION

This research presents a machine learning-based framework for predicting Myers-Briggs Type Indicator (MBTI) person- ality traits from social media text data. By decomposing the classification task into four binary subtasks corresponding to the MBTI dimensions, we leverage both Support Vector Ma- chines (SVM) and Random Forest classifiers, with extensive natural language processing for feature extraction.

Our experimental results demonstrate that the Random Forest model consistently outperforms SVM across all MBTI dimensions, achieving an average accuracy of 77% and F1- score of 0.76. In particular, the model excels in predicting the Introversion/Extraversion and Judging/Perceiving dimensions, which are often linguistically more distinguishable. The Ran- dom Forest model also offers interpretability through feature importance, enabling psychological insights from linguistic behavior.

The study also confirms that social media text, despite its informal nature, contains rich linguistic and behavioral cues that can be mined to infer personality with reasonable accuracy. Feature engineering using sentiment scores, lexical statistics, and part-of-speech patterns contributed significantly to classification performance.

However, challenges such as class imbalance, linguistic am- biguity, and limited training data for rare MBTI types remain. Addressing these issues will be essential for improving real- world applicability. Future work could involve incorporating deep learning models like BERT for contextual understanding, exploring multi-modal inputs (e.g., images, interactions), or adapting the framework for longitudinal personality tracking. In conclusion, our findings highlight the potential of machine learning models, particularly ensemble approaches like Random Forest, in advancing computational personality recognition. The proposed framework offers a scalable, in- terpretable, and effective solution for personality inference, opening up applications in areas such as personalized content delivery, mental health screening, and digital user profiling.

REFERENCES

[1] J. Golbeck, C. Robles, M. Edmondson, and K. Turner, “Predicting personality from Twitter,” in IEEE International Conference on Privacy, Security, Risk and Trust, 2011, pp. 149–156.
[2] B. Verhoeven, W. Daelemans, and B. Plank, “TwiSty: A multilingual Twitter stylometry corpus for gender and personality profiling,” in Proc. LREC, 2016.
[3] B. Plank and D. Hovy, “Personality traits on Twitter—or—How to get 1,500 personality tests in a week,” in Proc. 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis, 2015, pp. 92–98.
[4] T. Yamada, K. Sugiura, and Y. Ogawa, “MBTI Personality Prediction using Pre-trained Language Models,” in Proc. 9th Workshop on NLP for Social Media, 2021.

[5] C. Filardi, J. Burger, and L. Sen, “Explaining MBTI Personality Pre- diction with Interpretable Machine Learning,” in ACM Transactions on Interactive Intelligent Systems, 2021.

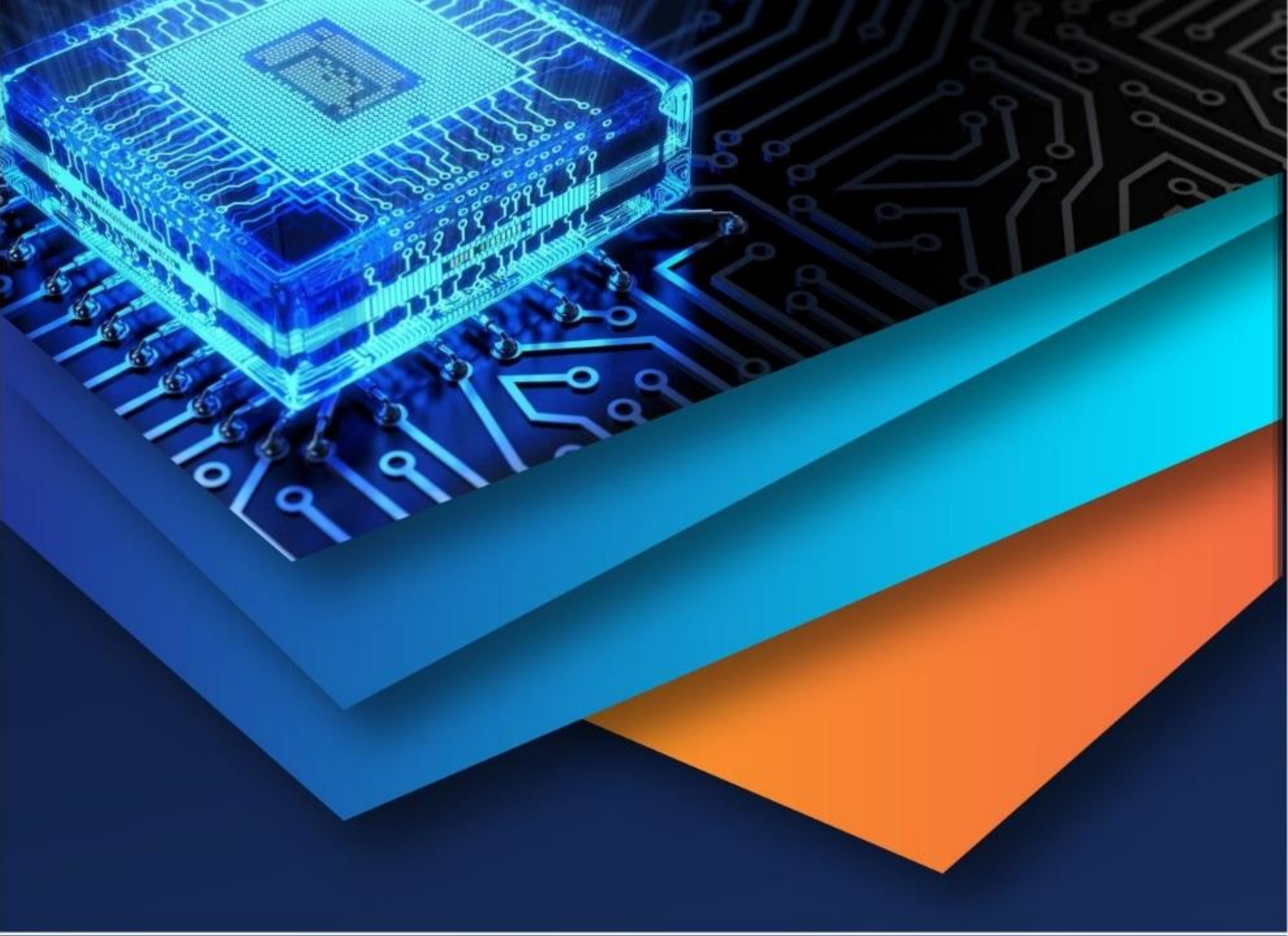
[6] Jothi Prakash and Arul Antran Vijay, ”A Unified Framework for Ana- lyzing Textual Context and Intent in Social Media,” ACM Trans. Intell. Syst. Technol., 2024.

[7] Bo Han, Paul Cook, and Timothy Baldwin, ”Lexical normalization for social media text,” ACM Trans. Intell. Syst. Technol., 2013.

[8] Hetal Vora et al., ”Personality Prediction from Social Media Text: An Overview,” IJERT, 2020.

[9] Mourad Ellouze and Lamia Hadrich Belguith, ”AI for Personality Traits and Mental Health in Social Media: A Survey,” ACM Trans. Asian Low- Resour. Lang. Inf. Process., 2024.

[10] G. Park, M. A. Schwartz, J. Sap, et al., “Automatic personality prediction from Facebook profiles,” in J. of Personality and Social Psychology, vol. 108, no. 6, pp. 934–952, 2015.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY