JCBI
Journal of Computing & Biomedical Informatics
ISSN: 2710 - 1606
Volume 08 Issue 02
2025
Research Article
https://doi.org/10.56979/802/2025

Personality Prediction of the Users Based on Tweets through Machine Learning Techniques

Shiza Aslam[1], Muhammad Usman Javeed[2], Shafqat Maria Aslam[3], Munawar Iqbal[4], Hasnat Ahmad[5], Anees Tariq[6,*]

[1,2]Department of Computer Science, COMSATS University of Islamabad, Sahiwal Campus,5700, Pakistan
[3]School of Computer Science, Shaanxi Normal University, Xi'an, Shaanxi, 710062, China
[4,5]Department of Computer Science, University of Engineering and Technology, Taxila,47050, Pakistan
[6]Department of Computer Science, Szabist University Islamabad, Pakistan
[*]Corresponding Author: Anees Tariq. Email: anees.tariq@szabist-isb.edu.pk

Received: January 02, 2025 **Accepted:** January 24, 2025

Abstract: With the advancement of interpersonal organizations, the massive rich data from Social platforms such as Facebook, YouTube, Instagram, and Twitter supply determining information about Social communications and human manners. Various approaches have been created to characterize clients' characters depending on their social exercises and propensities for language use. Depending on AI calculations, information sources, and capabilities, specific methodologies vary. The Informal community application records the enormous measure of users' conduct communicated in different exercises like preferences, notices, posts, remarks, photographs, labels, historical textual features, tweets, user profiles, and offers. This examination attempts to execute a few extreme learning designs to see the correlation by exhaustive investigation strategy through the exact results. We inspect the presence of plans of interpersonal organizations and semantic attributes comparative with character connections utilizing the myPersonality project dataset. MyPersonality dataset consists of tweet data with 20 attributes and 9917 records from 250 users. The investigation also looks at two AI models and plays out the link between every one of the feature sets and personality traits. The outcomes for the forecast exactness show that regardless of whether it is tried under a similar dataset, the character expectation framework based on the Logistic Regression classifier outflanks the standard for all the included sets, with a prediction accuracy of 98.9%. Different analysts use numerous old Machine Learning calculations to establish their models. The best prediction accuracy of 99.8% is gained by using the Random Forest classifier. This research will be beneficial for recruitment, mental health assessment, behavior analysis, and personalized marketing by using traits from tweets.
Keywords: Machine Learning, Natural Language Processing, Text Mining, Semantic Analysis, Social media Networking, Enormous Five-character Quality, Personality prediction, Social networking, Features Set, Interpersonal Organization, Personality Traits, Social Behaviour, Features Analysis.

1. Introduction

Personality is a person's characteristic thinking, feeling, and behavior patterns. It also refers to how a person responds to specific circumstances. It is a mixture of features that distinguishes a person. Facebook, Twitter, and Weibo, among other social media platforms, are the most popular destinations for internet users. Usage of social communication sites is increasing day by day. Researchers can examine and recognize people's online behaviors, tastes, and personalities in their activity on social media platforms. Individuals utilize web-based media to concentrate on specific issues connected with their lives and family well creatures, brain research, monetary issues, association with social orders, climate, and legislative mess. Character alludes to the single changes in an individual's trademark examples of reasoning, feeling, and acting. The Materials and Methods moreover insinuate how a person responds to well-defined conditions. The number of people who utilize these interpersonal interaction goals is increasing daily. Analysts can analyze and perceive individual online practices, tastes, and characters in their movement via web-based media stages. Researchers have demonstrated this and can assess personality traits using Social communication platforms like Twitter. Researchers can analyze interconnection patterns, profile data, language, or multimedia content throughout discussion or tweet updates to extract raw data to determine personality traits. On status profiles or inclinations, distinctive identities arise in different social relations and interaction behaviors. Researchers can examine their interaction patterns, profile information, text, or multimedia assets used throughout a discussion or tweet update to determine personality traits. Our personality prediction on the Social media network is based on users' social behavior and language usage tendencies.

The current study point of convergence on Twitter as Statistics shows that in June 2016, there were three hundred and thirteen million Twitter users [1]. Even though Twitter is now more extensively used to upload photographs and videos, this study focuses on the linguistic element of users' tweet updates. Social behavior toward friends and structural data such as the number of friends, groups joined, and Tweets. Analysts can investigate their association design, profile information, language, sight, and sound substance utilized during conversation or declarations to remove unrefined data to choose character characteristics. On Twitter, different characters depend on the rise of distinctive social relations and communication hones. Researchers can analyze their collaboration plan, profile information, content, or blended-media assets utilized amid a discussion or declaration to choose character qualities. Users often express their feelings and considerations through notices and remarks. Despite the process on social media, which is currently more broadly used to transfer photos and recordings, this review centres around the phonetic component of user notices. Given social conduct toward companions and underlying information like the number of companions, bunches joined, and preferences, in addition to other things. This study's character forecast framework aims to characterize a user's character in light of McCrae and Costa's enormous five-character qualities, a model that divides a single character into five attributes: agreeableness, conscientiousness, neuroticism, extraversion, and openness. Collecting conditions improves accuracy, and the forecast model incorporates the results of two machine-learning calculations.

The reason for this review is to make a framework that can naturally decide a user's character dependent on their Social Network propensities. To start, we select the most helpful elements for every character aspect and adequately anticipate the characteristics of the user. Based on the myPersonality dataset, a proposed strategy for planning and carrying out one class of Social Network Analysis involves just a pair of classifications of semantic elements like Linguistic Inquiry and Word Count (LIWC) and Structured Programming for Linguistic Cue Extraction (SPLICE). We examine the connection between each list of capabilities and its character credits. Various organizational properties are developed and their links to explicit features. At that point, we test the characteristics' expectation potential by foreseeing every character's highlight. There are four distinctive commitments: first, to exhibit the connection among users' characters and collaborate with them in social media networks. Secondly, it shows the higher capability of free interpersonal organization highlights for character expectation utilizing the Boost AI approach; third, it exhibits the connection between LIWC word references and the SNA highlights set; and fourth, it presents situations when a higher consistency is required. We use it to check out the highlights that have the most grounded relationship. Finally, Artificial Intelligence calculations are altering, placing them in the expectation model, and considering the most elevated calculation legitimacy for the enormous five-character quality characteristics to see how effectively we can predict character attributes from Twitter. We utilize three distinctive AI procedures as standard techniques for the forecast model to the applicable highlights information for order correlations. We use data assembled by the web-based media organization. The enormous five-character qualities programmed distinguishing proof are still designated. We investigate the relationship between each feature set and its personality attributes.

Regarding social elements, we're interested in network size, density, brokerage, transitivity measures, and their relationships with specific attributes. Then, we predict the traits to test each personality trait. We use it to look at the features that are most closely related. We utilize it to look at the features that have the strongest correlation.

Finally, machine learning methods are changed, adapted into a predictor model, and the highest algorithm validity for the enormous five-character personality qualities is compared to see how well personality attributes can be forecasted from Twitter. Three different machine learning methods are used as standard procedures for the predictor model to match the relevant attributes data for classification comparisons. We use information gathered by the Twitter social media network. The enormous five-character qualities of automatic identification are being targeted. There are four distinct contributions: firstly, demonstrate the relationship between the user's personalities and their interaction behavior in social media networking; secondly, show the higher perspective of independent Social media attributes for personality prediction using the Boost machine learning technique; thirdly, show the connection linking LIWC dictionaries and SNA characters set; fourthly, when excessive predictability is required to initiate the cases.

The organization will predict character qualities, and analysts will pick various methodologies. Few Specialists have moved toward APR as a grouping issue and have applied order strategies like Support Vector Machine (SVM) [2], K Nearest Neighbor (KNN), Naive Bayes (NB), Decision Trees (DT), Sequential Minimal Optimization for Support Vector Machine (SMO), Bayesian Logistic Regression (BLR), Multinomial Naive Bayes (MNB), and Rule Learning, among others [3,4]. Yet, different scientists have utilized multiple/multivariate linear regression (MLR) to anticipate the character attribute scores of clients in interpersonal organizations. M. Tadesse et al. [5], the character attribute of the client depends on various highlights and proportions of the enormous five-character model. It analyzes the presence of constructions of informal community and phonetic elements from the dataset. A framework based on the XGBoost classifier is selected to classify the user's features.

AI calculations are in use. Laleh, Asadzadeh, and Shahram et al. [6], the proposed model uses LASSO calculation to choose the leading highlights and to foresee the user's enormous five-character qualities. Oberlander and Nowson (2006) used the Naive Bayes prediction model as a learning method using distinct sets of n-grams as features to classify bloggers' extraversion, consistency, agreeableness, and conscientiousness. Golbeck et al. [7] used linguistic features of 279 users' personalities predicted from Facebook, like word count and interactive media attributes. Ross et al. [8] According to the study, Shyness is strongly associated with overall online time and inversely related to the circle of followers. Sumner et al. [9] create a link between a user's personality and their Facebook activity, content, and emotion. Kalish and Robins [10] inspect the effect of individual personality differentiation intrinsic instant network surroundings through an experiment. There are numerous drawbacks, such as techniques requiring a prolonged training time for the system due to consecutive inputs, and algorithms cannot capture the syntactic meaning of words; thus, the text is lost. Human asset chiefs should be able to recognize individual practices and qualities as they can grasp laborers' various characteristics. These characters should be seen with the goal that the organization can use their discernible attributes and abilities in the proper work.

The rest of the paper is formatted as follows. Part II discusses related work on personality prediction. The features of the myPersonality datasets are instigated and described in section III. The methods and data presentment are covered in section IV, followed by feature extrication and feature selection. The dataset is divided into two sections: contents feature extrication and Inspection of social interaction patterns, using a feature selection step that describes the traditional personality characteristic prediction method. The correlation findings are examined in part V and section VI. Multiple prediction models based on the SNA, LIWC, and SPLICE sets of features are used to anticipate the findings accurately. Section VII summarizes our findings and proposes suggestions for future research.

2. Related Work

A growing number of study items relating to a user's interaction with social media have recently gained increased traction in the research community. This area is very challenging for academic researchers and software vendors. For example, IBM offers a service called IBM Watson Personality Insights that provides an API for programmers to extract information from users' social media accounts and predict users' personality traits. IBM has listed some potential applications of personality prediction. Most temperament prediction studies use social media data to predict personality traits. The enormous five-character qualities Model principally supports the results of the projections. Two principal disciplines examine personality: Computing interpretation and Social Network Analysis.

According to the computing interpretation field, Pennebaker et al. [11] organized a pioneering effort to extract characters from text. They examined terms in various contexts, such as personality components with language signals in journal articles, school writing assignments, and social neuroscience research compositions. Their findings suggest that happy people use more articles, whereas loners and Those with a low level of conscientiousness employ more qualifier phrases. Psychotics include more words with a pessimistic tone. Argamon et al. [12] used phonetic elements such as job words, judgmental and assessment articulations, and modular action words to group neuroticism and extraversion. Their findings revealed that neuroticism is linked to using relevant linguistic high points for occurrence assessment linguistic scholarly classification. However, the results of extraversion were less obvious.

Neuroticism has been connected to irrational views in research or powerless adjusting achievement on the prosperity character [13]. Oberlander and Nowson (2006) used the Naive Bayes forecast model as a learning calculation and various arrangements of n-grams as highlights to group bloggers' extraversion, strength, appropriateness, and principles. Karney and Bradbury [14] investigated the connections between the Enormous five-character traits. Rather than network design and other extralinguistic ideas, character recognition is used in Network Analysis and has a far shorter history. Gosling et al. [15] considered the impact of a user's social collaboration.

Behavior on the character. Inspected persons' characteristics through Twitter usage and noticeable data of profile. Each individual's highlights were based on factual facts rather than conceptual properties. Davis et al. [16] demonstrated that people can judge others' personalities based on their Twitter profiles. Golbeck et al. linguistic variables like word count and social networking attributes were used to predict the personality of 279 Facebook members. Ross et al. [17] exposed that Shyness is inversely associated with the number of companions and is strongly linked to the amount of time spent on the internet. Sumner et al.'s findings revealed that phrases conveying pessimism, outrage, taboo themes, money, religion, or death are strongly associated with transparency. Designated promoting, user procurement, user care, and special interactions. Notwithstanding the Facebook and Twitter information used for the character forecast, a few scientists created character expectation models given cell phone logs. The author [18] tracked down the connection between the user's nature and Twitter use, tweets content, and emotions. Another examination bunch [19] showed that a few visual examples are associated with the character attributes of clients, and a client's character can be anticipated from the most loved pictures on the Flickr online photograph-sharing stage. Information from a blog web page writer called Live Journal was utilized to gauge an individual's impact and character attributes in an exploration study [20]. In another review, YouTube information was utilized as the source space to fabricate the preparation model, and an individual's extraversion attribute was anticipated in the little gathering meeting [21]. Moreover, scientists fostered a character forecast model using discourse information and Support Vector Machine (SVM) Regression calculation [22]. As displayed in the investigations clarified above, an individual's character can be anticipated given literary information, yet additionally, client profile, most loved pictures, Facebook likes, video, discourse, and contributing to a blog webpage information. In our multifaceted character expectation project, we will initially explore the utilization of text-based information. To execute character forecast on Twitter. [23] and [24] assembled a character forecast framework for the enormous five-character qualities Model. The Dark Triad character model was also given a character forecast framework. [23], [24], and [25] MRC Psycho-linguistic Database and LIWC (Linguistic Inquiry and Word Count) were used to create the character expectation framework for English. One more concentration in good time for [26], and LIWC was utilized to assemble a character forecast framework on Twitter. In our review, the myPersonality dataset is utilized as an example of character outcomes on Twitter profile information. Information was gathered by Schwartz et al. [27] through a Twitter application that carried out the enormous five-character qualities test included in other mental tests. The implementation incorporates getting the assent of users to track their informational data to make it usable for different examination grounds. Bachrach et al. [28] discovered a link between a user's mobility and personality using the myPersonality dataset. The outcomes exhibit that suitability is emphatically connected with the number of labels, though neuroticism has a bad relationship with the number of companions. Farhadi et al. [29] concentrated on the connection between the feelings communicated through tweet refreshes and users' age, orientation, and character. He discovered that open people are more excited in their tweets than people who are neurotic. Cantador et al. [30] utilize the data to examine the association between character kinds and user liking in various distractions, including songs, movies, episodes, and literature. In brain science, the hypothesis given the enormous five-character qualities is the generally broadly acknowledged model to depict the fundamental design of a human character. The hypothesis given these elements is called the five-character model (the enormous five-character qualities), and it is the most generally acknowledged model of character. It provides language and a mathematical set-up that pulls at the same instant a significant portion of the study findings in the brain science of a single person's differences and mentality. It lessens the enormous number of individual descriptive words into five primary character characteristics that structure the Abbreviation Sea [31]]. It was first examined during the 1990s when five-character attributes were laid out and have been utilized until now. As per Table 1, people in the enormous five-character qualities model differ as far as the OCEAN. It addresses arrangement based on qualities that could catch character contrasts.

Table 1. Characteristics Of The Enormous Five-Character Qualities

Values	Openness(O)	Conscientiousness(C)	Extraversion(E)	Agreeableness(A)	Neuroticism(N)
NO	74	120	154	116	151
YES	176	130	96	134	99

Table 2. Summary Of Previous Personality Prediction Studies

Metadata	Author Name	Method	Attribute	Outcome
Twitter	Golbeck et al.	ZeroR	LIWC	MAE 0.118
	I.F. Iatan	ANN	LIWC	NRMSE 0.079
	Ong et al.	Xgoost	LIWC	Accuracy 97.9%
Facebook	Farnadi et al	SVM	LIWC SNA Time-Related features	Precision 0.71
	Tandera et al.	SVM	SNA SPLICE LIWC	Accuracy 70.4%
	Schwartz	R	n-grams Extracted topics	R 0.42
Blogger	T. Yarkoni	Spearman's Rank correlation Coefficient	LIWC n-grams	P 0.32

In Table 2, a summary of studies done in the past with the methods and metadata used are shown concerning the accuracy achieved. The table compares various studies that analyze social media data (specifically from Twitter and Facebook) using different methodologies and attributes, focusing on the outcomes of their analyses.

2.1. Twitter

Golbeck et al. used the ZeroR method focusing on LIWC (Linguistic Inquiry and Word Count) and reported a Mean Absolute Error (MAE) of 0.118. It suggests a baseline performance for predicting outcomes based on the provided linguistic features. I.F. Iatan employed an Artificial Neural Network (ANN) also using LIWC metrics, achieving a Normalized Root Mean Square Error (NRMSE) of 0.079, indicating a better predictive performance compared to the previous study, as lower NRMSE values reflect better model fit. Ong et al. utilized Xgoost with

LIWC and reported an Accuracy of 97.9%, which is exceptionally high and indicates strong model performance in classification tasks.

2.2. Facebook

Farnadi et al. applied Support Vector Machines (SVM) with LIWC and Social Network Analysis (SNA) features, obtaining a Precision of 0.71. It suggests a moderately effective model for identifying relevant outcomes in the data. Tandera et al. also used SVM, combining SNA, SPLICE, and LIWC, reporting an Accuracy of 70.4%. It reflects a solid but lower performance compared to other studies on Twitter. Schwartz implemented a method using R and n-grams based on Extracted topics, achieving a score of 0.42. It indicates a correlation measure, suggesting a moderate relationship in the data analyzed.

2.3. Blogger

Yarkoni reported a Spearman's Rank Correlation Coefficient (P) of 0.32 when analyzing LIWC and n-grams. It indicates a weak to moderate correlation between the variables analyzed in their study. The data reflects a range of methodologies and performance outcomes across different studies on social media platforms. Notably, the high accuracy of Ong et al. on Twitter stands out, while other studies present varying levels of predictive performance. Using different techniques (ANN, SVM, etc.) and attributes (LIWC, SNA) showcases diverse approaches to analyzing social media data.

3. Dataset

We used the myPersonality dataset as a case study to investigate personality traits from social networks. The research was based on using 250 users and 9917 status updates from the myPersonality sample. The Twitter user dataset was labeled using the enormous five-character qualities. Each user had numerous postings aggregated in one file in the dataset, according to the personality categories in Table 3. The datasets were used to choose information about the individual, like the person's social system and tweets.

Table 3. Personality Traits Distribution

Sr. No.	Personality Traits	Characteristics
1.	**Openness (O)**	**High score:** Very creative, imaginative, likes to explore new things, is open to trying ideas and experiences, is adventurous, interested in artistic endeavors, and knowledgeable. **Low score:** Resist new ideas, not imaginative, dislike changes, abstract or theoretical concepts.
2.	**Conscientiousness (C)**	**High score:** Orderliness, dutifulness, self-discipline, having a set schedule, and cautiousness. **Low score: Dislike**'s schedule or structures fails to complete assigned tasks and make messes.
3.	**Extraversion (E)**	**High score:** Enjoys being the center of attention and meeting new people, has a wide social circle with friends, is active, affectionate, and fun-loving. **Low score:** Prefers solitude, is not socialized, dislikes being the center of attention, is reserved, sober and independent.
4.	**Agreeableness (A)**	**High score:** Good-natured, helpful, forgiving, Enjoys helping, cares, and contributing to the happiness of other people. **Low score:** Aggressive, competitive, rude, manipulated, Insults and belittles others.
5.	**Neuroticism (N)**	**High score:** Insecure, worried about little things, feel anxious, get upset easily, and Experience dramatic shifts in mood. **Low score:** Emotionally stable, calm, self-satisfied, secure, relaxed, and deals well with stress.

A detailed description of five personality traits, including Neuroticism, Agreeableness, Extraversion, Conscientiousness, and Openness, is provided in Table 3. These traits are used to categorize and label Twitter users in the dataset used in this study. On the basis of these personality traits, the dataset consists of data from 250 people and their 9917 status updates, which comply with all postings of a single user as of his profile. Table 3. also explains the psychological and behavioral characteristics linked to low and high scores for each trait.

#AUTHID	STATUS	sEXT	sNEU	sAGR	sCON	sOPN	cEXT	cNEU	cAGR	cCON	cOPN	DATE	ETWORKSIZ	ETWEENNES	ETWEENNE	DENSITY	BROKERAGE	BROKERAG	TRANSITIVITY
523e4e93a	e sound of t	2.65	3	3.15	3.25	4.4	n	y	n	n	y	6/19/2009 15:21	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	funny that	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/2/2009 8:41	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	r neck to st	2.65	3	3.15	3.25	4.4	n	y	n	n	y	6/15/2009 13:15	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	y sounds ir	2.65	3	3.15	3.25	4.4	n	y	n	n	y	6/22/2009 4:48	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	is home. <3	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/20/2009 2:31	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	hejokerblog	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/16/2009 15:21	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	NAME* + Te	2.65	3	3.15	3.25	4.4	n	y	n	n	y	6/27/2009 5:41	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	s into her 1	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/18/2009 6:34	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	w contempl	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/9/2009 14:58	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	to swinger	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/7/2009 23:41	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	on the Gre	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/15/2009 19:48	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	agic on the	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/5/2009 4:27	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	r One this	2.65	3	3.15	3.25	4.4	n	y	n	n	y	6/25/2009 4:36	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	on schedule	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/11/2009 3:38	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	of inspirati	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/7/2009 21:35	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	o bed at 9:	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/18/2009 1:29	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	casionally g	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/21/2009 22:18	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	uld be an o	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/31/2009 2:27	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	ME*, let m	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/11/2009 5:44	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	e many flav	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/27/2009 6:59	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	r game on	2.65	3	3.15	3.25	4.4	n	y	n	n	y	6/16/2009 4:52	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	to draw wh	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/26/2009 0:16	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	ssage could	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/30/2009 3:39	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	er that prev	2.65	3	3.15	3.25	4.4	n	y	n	n	y	9/1/2009 2:05	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	HAT TO DR	2.65	3	3.15	3.25	4.4	n	y	n	n	y	8/2/2009 9:14	180	14861.6	93.29	0.03	15661	0.49	0.1
523e4e93a	lied, buried	2.65	3	3.15	3.25	4.4	n	y	n	n	y	7/26/2009 19:53	180	14861.6	93.29	0.03	15661	0.49	0.1

Figure 1. Attributes of Dataset

4. Methodology

A growing number of approaches are utilized on language and social networking components of profiles and tweets to conclude personality traits. It illustrates the higher potential of individual social network features for personality prediction, two different approaches were used: logistic regression and random forest. The personality prediction framework includes the Pre-processing of data and feature extrication, and The machine learning process is followed by selecting features and prediction findings. The machine learning method and prediction results are shown with the pre-processing of data, extrication of features, and selection of features in Fig.2

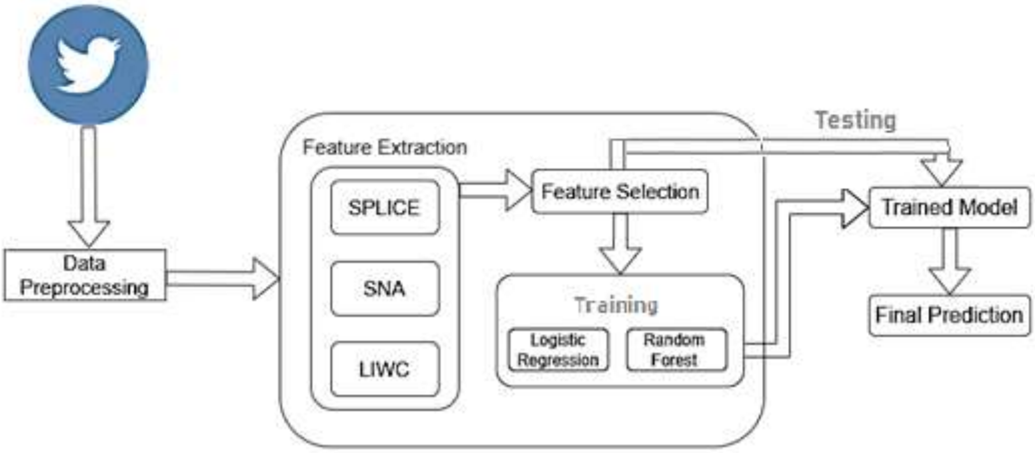

Figure 2. Proposed Methodology

4.1. Pre-Processing of Data

Before proceeding to the feature selection and training stage, the dataset obtained from myPersonality was pre-processed. As shown in Fig 1, we used OpenNLP [32] to pre-process the dataset. We employ the tokenization method to distinguish the last word of each phrase from punctuation and a group of the same terms. The terms in the LIWC and SPLICE linguistic qualities have the same stem. Symbols, names, spaces, URLs, and lowercase letters are removed, and the association between personality and linguistic features is examined. Words that have been stemmed may be injured [37]. For example, in the present or past tense, verb stemming would force distinguishing between them unfeasible. As a result, stemming is not used in the correlation analysis portion of our experiment, and all words are left unstemmed.

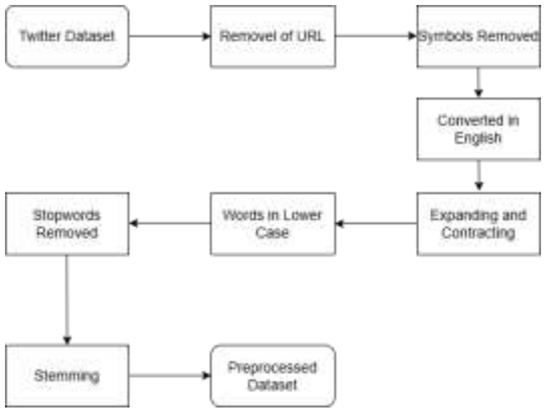

Figure 3. Pre-Processing of Raw Dataset

4.2. Feature Extraction

The existence and activity of other users on social networks have an impact on a user's behavior. These interactions may have an impact on how new information or behaviors are passed down through the groupings. Understanding how such behaviors emerge and spread has a wide range of possible uses. All of the data in the dataset may be divided into two categories. The first category is affiliated with text features extraction, which includes expression count and a subject's count and indicates a user's linguistic patterns on Twitter. We employ two dictionaries, LIWC and SPLICE, to assess the content of Twitter posts. The second category is affiliated with social association behavior interpretation, which includes network size, density, brokerage, and transitivity. These data indicate a user's primary Twitter social media network activities.

We chose LIWC2015 for the text analysis since it is designed to swiftly and efficiently evaluate individual or multiple language files. It aspires to be better than LIWC 2007 and LIWC 2001. Its functionality is transparent and adaptable, allowing the user to experiment with word usage in various ways. The Linguistic Inquiry and Word Count dictionary, sometimes known as LIWC, is commonly used in psychological research. It is used to extricate 85 linguistic attributes from the texts in our study, which includes five subcategories: standard counts (for example, wordage, word size greater than six letters, total prepositions), mental process (for example, behavioral, related to sensation, societal, and emotional affairs), dependency (for example, time information), self-interest (for example, profession, money problems, or physical conditions), and speechlore measurements (for example, counting of several kinds of punctuation marks).

SPLICE stands for Structured Programming for Linguistic Cue Extraction, a more recent dictionary. It is currently being updated but is expected to be widely employed in personality prediction task investigations [33]. It is used to extract 74 linguistic elements, including cues related to the speaker's self-evaluation, which can be positive or negative, as well as convolution and clarity scores.

SNA is an abbreviation for Social Network Analysis, which is a method for evaluating social systems based on the combination of links within particular people or chains of connections that are interlinked with the nodes (symbolizing "actors") and ties (symbolizing the connection among these actors). It is a process for analyzing and evaluating the structure of connections that come within social aspects, particularly persons. This strategy openly indicates that ambiguous associations in societal groupings (for example, friends of friends) are significant. As reported by James and Christakis [34], Social relations appear to be associated with pleasure. Friends are 25 percent more likely to be pleased when a person is happy.

Furthermore, those appearing in the center of a social media network are happier in the long run than those on the perimeter. Within the analyzed networks, a collection of happy and unhappy users was detected with a three-degree separation. The amount of happiness of a person's friends' friends was linked to their happiness. Our research employed variables linked to a user's social network inference with personality attributes, such as network size, betweenness, density, brokerage, and transitivity. The number of nodes in a network, which indicates the number of connections, is called network size. Betweenness is the number of shortest connected paths between two individuals who are not directly connected. For example, an individual with a high level of betweenness is crucial for the flow of details between every user who is unknown to each other personally. Brokerage is the number of coming ties received by every user from others. A person is linked to other users or groups of users who are active in the network and keeps many connections to be a well-organized and significant intermediary towards other network vertices. Transitivity depends on the premise that "friends of my friends are also my friends," which means that 2 or 3 persons are interlinked without using an intermediary via a shared neighbor. One of them is only available through the connections of another individual and displays the number of connections between the network's nodes or depicts the nodes' social relationships.

4.3. Feature Selection

Feature selection is essential to enhance a model for two reasons [38]. Firstly, it reduces the dataset's high dimensionality by removing features that aren't needed for training and improving forecast, and the equation was utilized, such as the quality of selection of feature approach.

Table 4. Values between LIWC and personality trait by Pearson correlation

Openness	2nd Person Singular	Social Procedures Terms	Affective Procdure Terms	Prepositions	Demise
	0.16	0.12	0.14	0.06	0.07
Conscientiousness	1st Person Singular	Netspeak	Dictionary Terms	Perceptual Procedures Related Terms	Common Verbs
	-0.15	-0.12	0.11	0.14	-0.13
Extraversion	2nd Person Singular	Favorable Emotions Terms	Agreement Terms	Past Tense Verbs	Achievements Terms
	0.15	0.1	0.13	0.14	0.19
Agreeableness	Interrogative Sentence	Assent Terms	Sexual Termsprocedure	Social Procedures Terms	Biological
	0.17	0.09	0.19	-0.14	0.11
Neuroticism	Prepositions	Social Procedures Terms	Affective Procdure Terms	Anger	Favorable Emotion Terms
	-0.14	-0.11	0.12	0.15	-0.12

Pearson interconnection is a metric based on continuous relationships within two variables that forecast the link between personality assessment and extricated qualities. The linear correlation coefficient r for a pair of variables (x, y) is given by the Equation 1:

$$r = \frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2}\sqrt{\sum_{i=1}^{n}(y_i - \bar{y})^2}} \qquad Eq(1)$$

$\bar{x}$ and $\bar{y}$ indicates sample average

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \qquad Eq(2)$$

In equation 2, n stands for unique size, while xi and yi are unique values labeled with I, with r ranging from -1 to 1. If x and y are entirely firmly related, then r has a positive relationship of one and a negative relationship of one. The value of 'r' is '0' if x and y are individualistic.

5. Experimental Results

First, we assess each attribute's importance and calculate the Pearson correlation coefficients for three feature sets and personality scores. The outcomes are shown in Tables 4, 5, and 6. Due to text space constraints, The characteristics having an important relationship (0.05) within the feature value system and personality assessment on the r-value are investigated. They are highlighted. Many intriguing discoveries were made due to the relationship between LIWC characteristics and personality attributes. Extraverted people are more prone to utilize second (0.15) and third (0.16) person singular pronouns, as well as verbs in the past tense (0.14). The tweets were updated with terms of dictionary (0.17), terms of social interaction (0.15), and commonly used adjectives (0.18). Extraverts like to compose brief messages since they have an inverse relationship with the length of words (-0.08) and length of sentences (-0.09), indicating the one with top levels of extraversion tweet with a small number of words and brief phrases. They employ more expressing favorable emotion terms (0.6), like "love," "nice," and "sweet"; treaty (0.14), like "OK," "yes," and "agree"; and terms of social interaction (0.12), like"friend," "family," and interpersonal relations, meaning that those who are more extraverted are more inclined to utilize these words (0.19). According to the findings of [60], extroverts updated their Twitter posts with more expressive phrases than neurotic users.

Neurotic users update their tweets utilizing phrases that convey unpleasant sentiments, like anger (0.15), anxiety (0.08), or emotional activity (0.13). They use fewer words for social interaction terms (-0.13) or favorable emotions terms (-0.14). They tend to utilize "we", and "our" (0.05) when talking to others. Extroverts prefer using the pronouns "he" and "she". Neuroticism was positively correlated with word and sentence length (0.06), indicating that users with top levels of neuroticism would write more negative mood, anxiety, and irritability-related phrases. The conclusion is confirmed in Table 6. Pearson connection valuation within SNA and character attributes was shown to be favorable between neurotic and extroverted users. Regarding biological processes, you can share information relevant to your body and health on social media networks. Users with a higher level of openness use more frequent feature words than meaningful words. Tweets carry publications, prepositions (0.07), and pronouns like "them" (0.13), "he," "she," and "his" (0.16). For more significant expressions in the proposal, the useful connection between the average word (0.046), the word of the proposal, and the word exceeds 6 characters (0.087).

Table 5.Values Between SNA and Personality Trait by Pearson Correlation

Traits	Density	Betweenness	Network size	Brokerage	Transitivity
Openness	0.06	0.05	0.03	0.05	-0.07
Conscientiousness	**-0.15**	**0.12**	**0.15**	**0.12**	-0.04
Extraversion	**-0.25**	**0.26**	**0.32**	**0.26**	-0.4
Agreeableness	-0.09	0.06	0.08	0.06	**-0.16**
Neuroticism	**0.12**	-0.04	**-0.19**	**-0.14**	**0.15**

They often update their tweets as a preliminary action word (0.15), the word (0.13), emotional process (0.14), and the process of signaling (0.18), and confident spare words "unsure", "uncertain", (-0.06) and "never" (-0.02) appear as low frequency. Results are Sumner et al. Assuming that persons with high degrees of openness are more receptive to speaking about potentially sensitive subjects. They frequently have a good association with the term that expresses unpleasant feelings (0.16), religion (0.18), and death (0.071).

Words reflecting negative emotions, like "hurt," "ugly," and "disgusting," (-0.16) were inversely connected with conscientious users (-0.14). In contrast, they used to avoid discussing depressing themes. Tweets that describe emotional procedures, like "see," "hear," and "feel" (0.12), as well as words connected to interactional procedures (0.12). This indicates that sincere users share everything they see or hear with others. In addition, conscientiousness users were favorably linked with terms of the dictionary (0.12), proposing that they would be more likely to use good spelling terms than common ones like "btw," "lol," and "thx." (-0.13) or usual computerized phrases. They also employ fewer verbs (-0.13) and pronouns in the first person singular (-0.15). More questioning sentences and question marks are used by agreeable users (0.17). They employ the pronoun "I" as well as terms for biotic procedures like physical (0.11) and sexual (0.12). They are unlikely to discuss processes with others (-0.14).

Table 6. Pearson Association Values Between Splice Characteristics and Personality Attribute

Openness	Pleasantness	SWG Negativity	Activation	Imagery	Present Tense
	0.15	0.13	0.12	0.13	0.05
Conscientiousness	Verbal words	SWG Positivity	Submissiveness Ratio	Total Submissiveness	Agreement Ratio
	-0.07	0.01	-0.14	-0.12	-0.1
Extraversion	Pleasantness	Complexity Composite	Activation	Average Sentence Length	I Can Do It
	0.12	-0.16	0.1	-0.09	0.12
Agreeableness	Question Count	Num Interjections	Num Agreement	Pausility	Complexity Composite
	0.04	0.06	0.09	-0.12	-0.12
Neuroticism	Pleasantness	Expressivity	Activation	ProSelfImage	Complexity Composite
	-0.14	0.11	-0.12	-0.05	0.12

Results between SPLICE and enormous five-character qualities show that extroverts positively correlate with the word PosSelfImage (0.08) compared to neuroticism and describe themselves positively in text. Users with a higher degree of openness showed a strong correlation with images (0.13), so they are generally considered imaginative individuals with personal curiosity, an open mind, and a willingness to explore new ideas. Those who get high scores for openness and agreeableness used to update tweets with the activation word (0.12, 0.1).

Words used for pleasantness (0.15, 0.12) and present the tense verbs (0.05, 0.013). Open users have a positive connection with complex mixed words (0.12). On the other hand, favorable users and external users are unlikely to make use of them (-0.12 and -0.16). Conscientious users have a negative connection with AgreementRatio (-0.1), TotalSubmissiveness (-0.12), and SubmissivenessRatio (-0.14), and others do not easily handle the previously mentioned types of things and are planned initially, which I want to maintain. As stated by the results, few corresponding resemblances exist between the LIWC outcomes. SPLICE appeared in Tables 4 and 5, respectively. Extraverts exhibit a strong affinity for phrases like ICanDoIt (SPLICE 0.13) and terms of achievement like "win," "success,", and "better" (LIWC 0.19). Extraverts are unlikely to utilize Length of averageSentence (SPLICE -0.09), Length of averageWord (SPLICE -0.06), and word count (LIWC 0.08) in tweets, stating how extraverts prefer to communicate information in brief sentences with few words. The use of verbal words was negatively linked with the help of conscientious users (SPLICE-0.037) (LIWC-0.13). Both dictionaries agree on this. Using different dictionaries can improve prediction results, as shown by our results.

We discovered that extraversion is the most highly correlated trait based on the correlation results for the social network features. Extraversion is a personality trait that is strongly linked to the usage of Twitter. These findings back up the claims of [35] that extroverts and social networks have a positive relationship. They have a proclivity to go online searching for new and exciting experiences [36] in online and social media networks. Our findings back up this conclusion. Table 6 shows that extraverts have a positive correlation with network size (0.31), reflecting their proclivity to have many friends. Unlike neurotic users, they belong to distinct social circles, and their companions are frequently unfamiliar with one another. This observation is supported by the inadequate results for transitivity (-0.4) as a link quality in social media sites. Interconnections within friends are limited, as denseness has an inverse connection (-0.25). According to the findings, extroverted people are responsive to reward cues. They are engaging in a vast range of interactional activities. They have larger friendship networks as a result of their socializing tendencies. However, there are a lot of missing connections among their contacts.

Other transitivity features with negative correlations included openness (-0.07), consciousness (-0.04), and agreeableness (-0.16). Communicated people are more likely to be at the center of large, loosely connected networks, which explains why all four traits have a negative effect on transitivity. Neuroticism was associated with a larger degree of common neighbors who shared the same experience and had a positive relationship with transitivity (0.14). For all of the SNA features, neuroticism showed the opposite connection results. When compared to extroverted users, neurotic users have a negative relationship with network size (-0.18). They are recognized as unpopular interaction partners in online discussion networks and form tiny friendships. Their friends used to know each other due to the positive connection between transitivity (0.14) and density (0.11). Because of their high levels of anxiety, worry, and insecurity, neurotic persons use the internet less frequently and engage in information-based activities less frequently. Neurotic users avoid social interactions, particularly under stressful occasions, and express unhappiness with the help they receive from their social media friends. Paraphrase formalized: People high in extraversion are emotionally stable and likely to maintain persistent communications with their friends. They are less likely to seek out new experiences and are more likely to struggle with self-efficacy and self-esteem. When confronted with new challenges, such as learning a new technology, these individuals will likely encounter difficulties or try to avoid the new situation altogether. Because their friend's network is scattered and wide, and the social circle is dispersed, diligent users have a positive relationship with network size (0.14) but a negative relationship with density (-0.14). Users with high levels of agreeableness usually prefer friends who share their degrees of agreeableness, extraversion, and openness [69]. The degree to which an individual man is connected to others or groups of people who are not connected is referred to as brokerage and betweenness.

According to our findings, extroverted people had a higher connection between brokerage and betweenness among the enormous five-character personality traits (0.25). Individuals with high degrees of conscientiousness are regarded to be orderly and diligent. They're also good at brokerage (0.11) and betweenness (0.11), particularly in instrumental networks and as colleagues. When colleagues from various organizational areas request support in addressing work-related challenges, they are picked in particular. Agreeableness users can benefit from positive brokerage and betweenness outcomes (0.05). They are empathetic, cooperative, and eager to build positive relationships with others. They are frequently chosen as friends because of these characteristics. They are invited into the friendship networks of the team [69]. They are more expected to respond favorably to various ties leading to the node and to assist in integrating conflicting partners' opinions and needs. Users having a top level of openness are also favorably associated with a brokerage (0.05) and betweenness (0.05), indicating a wide range of interests. They are more expected to be chosen for friendships because of their curiosity as conversational partners. They may act as network brokers in their search for connections with people from various unrelated social circles. Neurotics who are worried, insecure, and aggressive have unfavorable associations with brokerage (-0.13) and betweenness (-0.03), which are not shared by other personality types. They also have a negative link with any group friendship and advice networks. They are likely to be avoided as expensive interconnected partners since they frequently display unpleasant feelings. As revealed in Tables 4 and 5, The majority of neurotic users add fury phrases to their tweets (0.20), happy emotions terms (-0.14), and social interaction terms (-0.13).

After analyzing the significant connections set in the earlier section, an experimental evaluation form on two standard methods was conducted. The test was designed to examine how effectively the models predicted outcomes using the myPersonality dataset, which comprised the 96-dimension attribute space for both single and combination traits. Before putting the models' prediction abilities on the test's merged data, evaluate each procedure with unique feature sets, like SNA, LIWC, and SPLICE attributes. It produce the personality prediction findings reported in the following section, and two algorithms were employed for classification. We used an open Python project to train our data using denary turns cross-training with denary repetitions. Each time, an individual turn was used for testing, and nine turns were employed throughout the training. We compared it to the main classification, and we employed Logistic Regression and Random Forest machine learners as a reference.

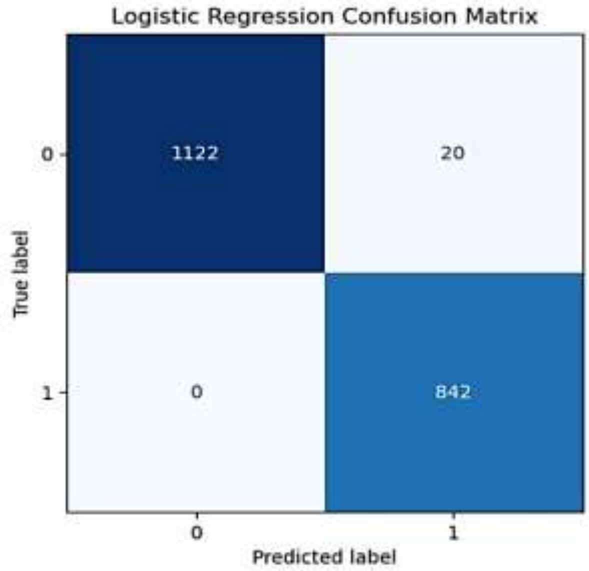

Figure 4. Confusion Matrix of Logistic Regression

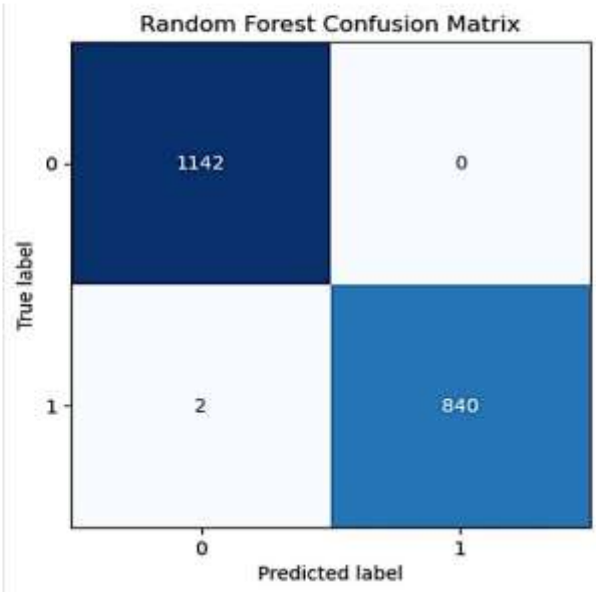

Figure 5. Confusion Matrix of Random Forest

Fig 4. shows the confusion matrix of the Logistic Regression Model. It demonstrates that 1122 instances are predicted TN (True Negative), 842 TP (True Positive), 20 FP (False Positive), and 0 FN (False Negative) out of a total of 1982 instances. Fig 5. illustrates the confusion matrix of the Random Forest Model. It shows that 1142 instances are predicted TN (True Negative), 840 TP (True Positive), 0 FP (False Positive), and only 2 FN (False Negative).

$$\text{Accuracy} = \frac{\text{True Positive} + \text{True Negative}}{\text{True Positive} + \text{True Negative} + \text{False Positive} + \text{False Negative}} \qquad \text{Eq(3)}$$

$$\text{Precision} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Positive}} \qquad \text{Eq (4)}$$

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}} \quad \text{Eq (5)}$$

$$\text{F1} - \text{Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{precision} + \text{Recall}} \quad \text{Eq (6)}$$

Eq (3)-Eq(6) are used to find Accuracy, Precision, Recall, and F1-Score, respectively. After finding the values of these metrics a performance comparison of both models is shown in Table 7. To restrain over-fitting [40], Random Forest employs a well-ordered pattern for interpretation that outperforms Logistic Regression. Personality traits were predicted with an average accuracy of 99.8 percent using all of the features extracted.

Table 7. Comparison of Performance Metrics

Model	Accuracy	Precision	Recall	F1-Score
Logistic Regression	0.989919	0.976798	1.0	0.988263
Random Forest	0.998992	1.0	1.0	0.998811

According to the previous study [39], features from different categories are merged to improve the model's capability. The outcome shows that both models exhibit high reliability, But Random Forest performs slightly better than Logistic Regression. Table 7. explains the outcome of both models as Logistic Regression gained recall of 100%, precision of 97.68%, F1-Score of 98.83%, and accuracy of 98.99%.

On the other hand, Random Forest shows outstanding performance with recall and precision of 100%, F1-Score 99.88%, and an accuracy of 99.90%. These stats show that both models are efficient, but Random Forest is a robust option for our model. Fig 6. illustrates a comparison of performance metrics (Recall, Precision, F1-Score, and Accuracy) of Random Forest and Logistic Regression. Due to outstanding performance, Random Forest is selected as the final model. If we compare our results with those of previous studies, the prediction of robustness and accuracy significantly improved. With an accuracy of 98.9%, Logistic Regression outperformed previous techniques like I.F. Iatan's ANN (NRMSE: 0.079), Golbeck et al.'s ZeroR (MAE: 0.118) and Tandera et al.'s SVM approach (Accuracy: 70.4%, Precision: 0.71).

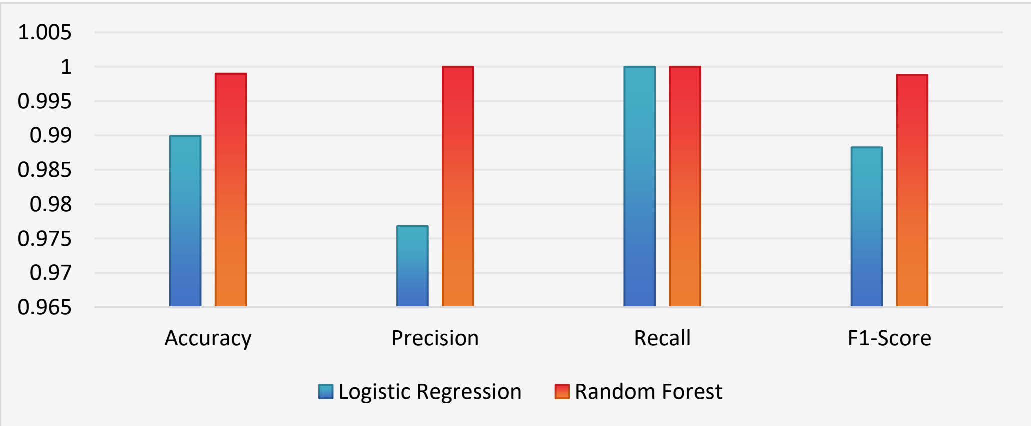

Figure 6. Illustration of Performance Metrics

Furthermore, with an ideal accuracy of 99.9%, Random Forest surpassed Ong et al.'s Xgboost, which achieved a remarkable 97.9% accuracy. In addition, Random Forest gained 100% recall and precision compared to Schwartz's topic extraction (R: 0.42) and Yarkoni's LIWC and n-gram analysis (P: 0.32). In performance metrics, our results, particularly for Random Forest, highlight remarkable improvements compared to previous studies. Assessing the Pearson correlation values in the dataset and every character dimensional confirmed that specific personalities fit specific functions. Our study uses two machine learning algorithms to foresee the character rankings from extricated functions [42]. Consequences confirmed that using social community functions for character prediction can reap a better overall performance than using linguistic functions.

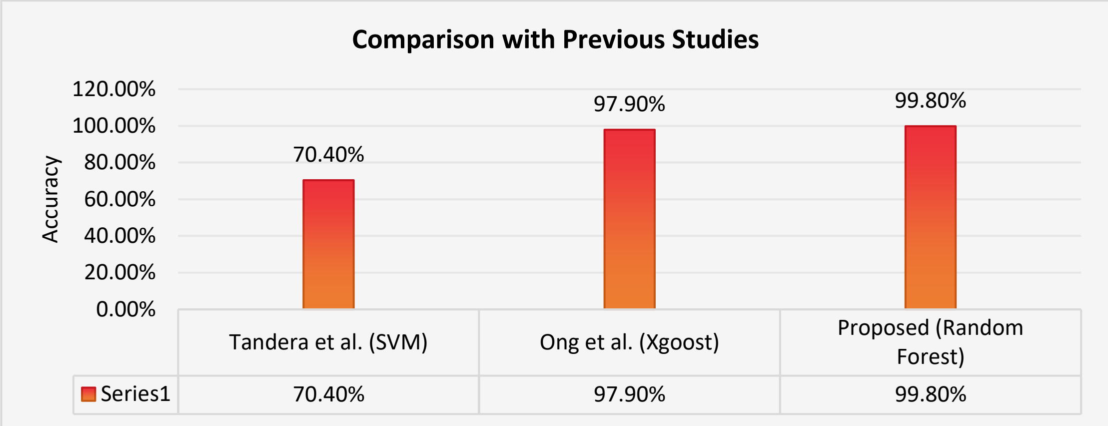

Figure 7. Performance Comparison with Previous Studies

The maximum character prediction is completed with the Random Forest device, which gains knowledge of the method [41] through the usage of person SNA function sets. Specifically, the very best accuracy is achieved for extraversion, which is the trait most customarily shown through Tweets. Individual feature set performance accuracy for openness, neuroticism, and agreeableness was likewise higher than combined feature set performance accuracy. Overall, Personality traits were predicted with an average accuracy of 99.8 percent using the combination of all the variables we retrieved.

6. Conclusion

The study investigates the literature on using social media frameworks as behavioral feature studies by looking at the link between users' personalities and their actions on social networks. We proposed a framework for social network and personality psychology research. To forecast users' personalities, we manage comparison research of optimum behavior signals for Twitter utilization of identical attributes to get how users interact, convey, and engage with one another. MyPersonality dataset is utilized to establish a large set of attributes appropriate for determining our study's distinct personality traits. Our findings suggest that examining social and linguistic indices of personality can provide a wealth of information. We may use a bigger training dataset to allow the system to integrate a more excellent range of characteristics, improving the system's efficiency. We discovered that employing multiple language dictionaries can enhance the correlation findings. Due to their great quantity, we discovered that linguistic variables provide a broad variety of correlation variants; when compared to social media network attributes, it is used to produce unconscious sense. These findings show that individual social network traits have more potential for personality prediction. Inferring user personality qualities on Twitter can be useful in recognizing users' online activity and providing suggestions for future individualized service upgrades. There are various crucial areas where we can expand our research scope. The precision of the results was limited because our experiment relied on a few data from the myPersonality dataset (260 users, 9920 Tweets). To improve the system's efficiency, we need to employ a large number of training datasets that allow the system to interact with a wide range of attribute sets. To solve further practicable difficulties with this innovation, such as plans for the public benefit and well-expressed content to consumers form on the mutual understanding of nodes in their social media network groupings. By researching personalities from Tweets, recommended systems may be able to enhance their predictive performance by suggesting items like Series and movies, music, or athletic occasions that are suited to the user's character. Based on mutual relationship ratings, things could be suggested to a specific user. Furthermore, we may identify folks with similar likes and recommend things to them utilizing a collective clarification technique. It could be used to group people together who have comparable personality qualities. Establishing and assessing the worth of these methodologies is a newfangled field of research that can be pursued in the future.

References

1. Twitter Investor Relations, "Q216 Selected Company Metrics and Financials," 2016.

2. Verhoeven, B., Daelemans, W., & De Smedt, T. (2013). Ensemble the methods for personality recognition. The Proceedings of WCPR13, in conjunction with ICWSM-13.
3. Farnadi, G., Zoghbi, S., Moens, M. F., & De Cock, M. (2013, January). Recognizing personality traits using Facebook status updates. InProceedings of the workshop on computational personality recognition (WCPR13) at the 7th international AAAI conference on weblogs and social media (ICWSM13) AAAI.
4. Alam, F., Stepanov, E. A., & Riccardi, G. (2013, June). Personality traits recognition on social network-facebook. In Proceedings of the Workshop on Computational Personality Recognition (pp. 6-9).
5. Khadim, U., Iqbal, M. M., & Azam, M. A. (2021). An intelligent three-level digital watermarking method for document protection. Mehran University Research Journal Of Engineering & Technology, 40(2), 323-334.
6. M. M. Tadesse, H. Lin, B. Xu, and L. Yang, "Personality Predictions Based on User Behavior on the Facebook Social Media Platform," in IEEE Access, vol. 6, pp. 61959-61969, 2018, DOI: 10.1109/ACCESS.2018.2876502.
7. A. Laleh and R. Shahram, "Analyzing Facebook Activities for Personality Recognition," 2017 16th IEEE International Conference on Machine Learning and Applications (ICMLA), 2017, pp. 960-964, DOI: 10.1109/ICMLA.2017.00-29.
8. J. Golbeck, C. Robles, and K. Turner, "Predicting personality with social media," inProc.ExtendedAbstr.Hum.FactorsComput.Syst.(CHI),2011, pp. 253–262
9. C. Ross, E. S. Orr, M. Sisic, J. M. Arseneault, M. G. Simmering, and R. R. Orr, "Personality and motivations associated with Facebook use," Comput. Hum. Behav., vol. 25, no. 2, pp. 578–586, 2009.
10. Javeed, M., Aslam, S., Farhan, M., Aslam, M., & Khan, M. (2023). An Enhanced Predictive Model for Heart Disease Diagnoses Using Machine Learning Algorithms. Technical Journal, 28(04), 64-73. Retrieved from https://tj.uettaxila.edu.pk/index.php/technical-journal/article/view/1828.
11. C. Sumner, A. Byers, and M. Shearing, "Determining personality traits & privacy concerns from Facebook activity," Black Hat Briefings, vol. 11, pp. 197–221, Dec. 2011.
12. Y. Kalish and G. Robins, "Psychological predispositions and network structure: The relationship between individual predispositions, structural holes, and network closure," Social Netw., vol. 28, no. 1, pp. 56–84, 2006.
13. J. W. Pennebaker, R. L. Boyd, K. Jordan, and K. Blackburn, "The development and psychometric properties of LIWC2015," Tech. Rep., 2015.
14. Nazir, T., Iqbal, M. M., Jabbar, S., Hussain, A., & Albathan, M. (2023). EfficientPNet—an optimized and efficient deep learning approach for classifying disease of potato plant leaves. Agriculture, 13(4), 841.
15. S. Argamon, S. Dhawle, M. Koppel, and J. Pennebaker, "Lexical predictors of personality type," Tech. Rep., 2005.
16. R. R. McCrae and P. T. Costa, "Personality, coping, and coping effectiveness in an adult sample," J. Pers., vol. 54, no. 2, pp. 385–404, 1986.
17. Hassan, M., Al-Awady, A. A., Ali, A., Iqbal, M. M., Akram, M., Khan, J., & AbuOdeh, A. A. (2023). An efficient dynamic decision-based task optimization and scheduling approach for microservice-based cost management in mobile cloud computing applications. Pervasive and Mobile Computing, 92, 101785.
18. B. R. Karney and T. N. Bradbury, "The longitudinal course of marital quality and stability: A review of theory, method, and research," Psychol. Bull., vol. 118, no. 1, pp. 3–34, 1995.
19. S. D. Gosling, A. A. Augustine, S. Vazire, N. Holtzman, and S. Gaddis, "Manifestations of personality in online social networks: Self-reported Facebook-related behaviors and observable profile information," Cyberpsychol., Behav., Social Netw., vol. 14, no. 9, pp. 483–488, 2011.
20. H. Davis et al., in Proc. 3rd ACM Web Sci. Conf. (WebSci), Paris, France, 2013.
21. C. Ross, E. S. Orr, M. Sisic, J. M. Arseneault, M. G. Simmering, and R. R. Orr, "Personality and motivations associated with Facebook use," Comput. Hum. Behav., vol. 25, no. 2, pp. 578–586, 2009.
22. C. Sumner, A. Byers, and M. Shearing, "Determining personality traits & privacy concerns from Facebook activity," Black Hat Briefings, vol. 11, pp. 197–221, Dec. 2011.

23. ul Hassan, M., Al-Awady, A. A., Ali, A., Sifatullah, Akram, M., Iqbal, M. M., ... & Abdelrahman Ali, Y. A. (2024). ANN-Based Intelligent Secure Routing Protocol in Vehicular Ad Hoc Networks (VANETs) Using Enhanced AODV. Sensors, 24(3), 818.
24. Tauqeer, H., Iqbal, M. M., Ali, A., Zaman, S., & Chaudhry, M. U. (2022). Cyberattacks detection in iomt using machine learning techniques. Journal of Computing & Biomedical Informatics, 4(01), 13-20.
25. Cristani, M., Vinciarelli, A., Segalin, C., & Perina, A. (2013). Unveiling the multimedia unconscious:
26. Implicit cognitive processes and multimedia content analysis. Proceedings of the 21st ACM int'l
27. conference on Multimedia (pp. 213-222).
28. Nguyen, T., Phung, D. Q., Adams, B., & Venkatesh, S. (2011). Towards the discovery of influence and personality traits through social link prediction.
29. Aran, O., & Gatica-Perez, D. (2013). Cross-domain personality prediction: from video blogs to small group meetings. Proceedings of the 15th ACM on International Conference on Multimodal
30. Interaction (pp. 127-130).
31. Tim, P., Katrin, S., Sebastian, M., Florian, M., Mohammadi, G., & Vinciarelli, A. (2012). On
32. speaker-independent personality perception and prediction from speech.
33. J. Golbeck, C. Robles, M. Edmondson, and K. Turner, "Predicting personality from Twitter," in Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on, 2011, pp. 149–156.
34. A. Wijaya, I. Prasetia, N. Febrianto, and D. Suhartono, "Sistem Prediksi Kepribadian 'The Big Five Traits' Dari Data Twitter," Bina Nusantara University, 2016.
35. C. Sumner, A. Byers, R. Boochever, and G. J. Park, "Predicting dark triad personality traits from Twitter usage and a linguistic analysis of tweets," in Machine Learning and Applications (ICMLA), 2012 11th International Conference on, 2012, vol.2, pp. 386–393.
36. G. Farnadi, S. Zoghbi, M. Moens, and M. De Cock, "Recognising Personality Traits Using Facebook Status Updates," Work. Comput. Personal. Recognit. Int. AAAI Conf. weblogs Soc. media,
37. pp. 14–18, 2013.
38. H. A. Schwartz et al., "Personality, gender, and age in the language of social media: The open-vocabulary approach," PLoS ONE, vol. 8, no. 9, p. e73791, 2013
39. Y. Bachrach, M. Kosinski, T. Graepel, P. Kohli, and D. Stillwell, "Personality and patterns of Facebook usage," in Proc. 4th Annu. ACM Web Sci. Conf., 2012, pp. 24–32.
40. G. Farnadi, S. Zoghbi, M.-F. Moens, and M. De Cock, "Recognizing personality traits using Facebook status updates," in Proc. Workshop Comput.Pers. Recognit. (WCPR13), 7th Int. AAAI Conf. Weblogs Social Media
41. (ICWSM), 2013.
42. I. Cantador, I. Fernández-Tobías, and A. Bellogín, "Relating personality types with user preferences in multiple entertainment domains," in Proc.CEUR Workshop Shlomo Berkovsky, 2013.
43. L. R. Goldberg, "The structure of phenotypic personality traits," Amer. Psychol., vol. 48, no. 1, pp. 26–34, 1993.
44. T. Morton, J. Kottmann, J. Baldridge, and G. Bierner, "Opennlp: A java-based NLP toolkit," in Proc. EACL, 2005
45. K. Moffitt, J. Giboney, E. Ehrhardt, J. Burgoon, and J. Nunamaker, "Structured programming for linguistic cue extraction (SPLICE)," in Proc. HICSS Rapid Screening Technol., Deception Detection Credibility Assessment Symp., 2012, pp. 103–108.
46. J. H. Fowler and N. A. Christakis, "Dynamic spread of happiness in a large social network: Longitudinal analysis over 20 years in the Framingham heart study," Brit. Med. J., vol. 337, p. a2338, Sep. 2008.
47. Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality," Comput. Hum. Behav., vol. 26, no. 6, pp. 1289–129

48. Y. A. Hamburger and E. Ben-Artzi, "The relationship between extraversion and neuroticism and the different uses of the Internet," Comput. Human Behavior, vol. 16, no. 4, pp. 441–449, 2000.