Journal of Innovative Computing and Emerging Technologies

Real Time Personality Analysis by Tweet Mining

Hirra Mustafa[1], Sobia Yaqoob[2], Tuba Mansoor[3]

Received: 06 December 2022; Accepted: 24 January 2023; Published: 08 February 2023

Abstract: Social Media addiction is becoming a Commonplace. A user gets deeply involve with an application or website that he share it thoughts , ideas and viewpoints on these platforms more comfortably then communicating with an people face to face . Through text mining these viewpoint and thoughts from social Media(psychological profiling) can be used to improve user experience & personality assessment .The aim is to exhibits the construct validation framework for Personality Analysis using machine learning approaches like Term Frequency-Inverse Document Frequency (TF-IDF) & SVM. We proposed a methodology by which user's personality can be analyzed using publicly available information on user's Twitter account using the Myers_Briggs Type Indicator (MBTI). This study will contribute & help in many ways such as customization of Content displayed and product list , Recruitment and Information Retrieval.

Index Terms-MBTI, Social Media, Tweets , Personality Analysis ,user experience.

I. INTRODUCTION

Real-time personality analysis is the process of using natural language processing and machine learning techniques to analyze text data and determine the personality traits of the person who wrote it. This can be used for a variety of applications, such as customer service chatbots, social media monitoring, and recruitment.Personality analysis using Twitter data can be done by analyzing the language and behavior of an individual on the platform. This can include analyzing the content of tweets, the frequency of certain words or phrases, and engagement with other users. However, it's important to note that such assessments are not always accurate.

A. Internet platform

The Internet platform is a place where people introduce themselves to other people, telling private details and awareness about their lives. We are starting to recognize how this information can be used to help user experience[1]. Web-based information is a series of interactions between people all around the world that is stored on a website and can be accessed by anyone with an Internet connection. This can be in large volume and continue to grow in size[2]. These series of interactions can contain text,Audio, Figure , videos, or any other type of digital content. Nowadays people use social media to share information, to reveal themselves, their points of view and experience content selection mechanism and positioning product & services.

B. Social media(twitter)

According to current research around 4.5 billion people are internet users,among which 3.9 billion are social network users . It has been used as a tool to share news and information and to connect people around the world with sustained connectivity. 60% of world population is already online, and the research indicate that more than 50% of

[1] Sharif College of Engineering and technology, Lahore Email: hirramustafa@gmail.com
[2] University of Okara, Okara Email: sobia.yaqoob@uo.edu.pk
[3] Riphah International University Email: tubamansoor786@gmail.com

mentioned population will be marked social media users by the 2023[3]. Social media use is common among young adults. In a 2015 Pew Research Center survey, 90% of 18- to 29-year-olds reported using at minimum of one social media site, and 80% used two or more[4]. Twitter is an internet platform that allows users to send and receive short messages called tweets. These tweets can include text, images, videos, and links and have a maximum length of 280 characters.Purposed methodology is analysis of how social media such as twitter can be utilized for personality assessment to to boost the user experience . A personality of an individual is their qualities and parts of others observations. The power to seek out associations between actions got from the knowledge gathered from web-based life. By the assistance of this data users are divided into different groups which depend upon their personality.

C. Personality Assessment

Personality refers to singular qualities example of feeling, behaving, thinking which makes them unique concerning another person. The likes, dislikes, thoughts, tests, and feelings expressed on Twitter as it is a micro-blogging website Therefore,this information is composed to see the personality of the user.Up to the present, to exactly estimate users' personalities, they' go through a personality Test, which made it unserviceable to use personality evaluation in various of social media domains. The proposed system will predict and analyze the user’s personality through the publicly available information on his Twitter account.

D. Real-time application

The real-time application will help to speed up the process by allowing developers to work on an application while it is being used by its user without any interruption. This can help to improve the standard of the application and make it more user- friendly[5].Real-time personality analysis can be used in a variety of real-time applications, such as:

Chatbots: Companies can use personality analysis to create more personalized customer service experiences by tailoring responses based on the customer's personality traits.

Social Media Monitoring: By analyzing the text data of social media posts, businesses can gain insights into the personality traits of their audience, which can be used to create more effective marketing campaigns.

Recruitment: Employers can use personality analysis to screen job candidates and determine which ones are the best fit for the company culture and specific roles.

The Twitter API will be used by the proposed system to fetch public Tweets from the platform in real-time so that a personality assessment algorithm can be implemented on them[6].Analyzing Real-time tweet data is different from Analyzing available corpus.Developers have to sign-up for the Twitter developers account to get access of twitter APIs, which then allows the developer to register the application for which they are accessing data from twitter database[25]. For application registration a form is provided to the Developer which incorporates the information and undertaking of the application , intention ,aim, URL[26] .The proposed model will predict personalities that will be appropriate for all organizations or businesses. The fields of these companies can be quite diverse. it can be often used ostensibly for marketing and predilection.

```
CONSUMER_KEY        = 'XXX'
CONSUMER_SECRET     = 'XXX'
ACCESS_TOKEN        = 'XXX'
ACCESS_TOKEN_SECRET = 'XXX'

AUTH = tweepy.OAuthHandler(CONSUMER_KEY, CONSUMER_SECRET)
AUTH.set_access_token(ACCESS_TOKEN, ACCESS_TOKEN_SECRET)

api = tweepy.API(AUTH)
```

Figure 1 Twitter Api for Text Mining (Tweets)

E. Personality Prediction models

There are various personality Prediction models available such as Myers Briggs Type Indicator (MBTI) [7], Strength Finder [9], and Big 5 Personality Traits [10] DiSC ((D)ominance,(i)nfluence,(S)teadiness,(C)onscientiousness) Model [8].The MBTI is a extensively recognized personality Analysis test for over 50 years. It is used to help people understand themselves and others better. MBTI is formulated on the typology assessment of C.Jung that is based on four dichotomies: extraverted (E), introverted (I), sensing (S), and thinking (T)[7]. respectively all types have their own positive and negative aspects. The MBTI can be used to help people identify their own personality type and understand how they interact with others. As it is descriptive and analytical,which makes it useful in a aggregation of many context including. Offices,School,Health-care.

DISC(D)ominance,(i)nfluence,(S)teadiness,(C)onscientiousness assessment is a tool introduced in 1928, used to help individuals understand themselves, their behaviors, and how to effectively interact with others. DISC is four different categories of behaviors as the benchmark: The DISC assessment is a great tool to help you learn more about the behavior of yourself and your team members, and how to communicate with them more effectively[8]. The DISC assessment is not a test, but rather a tool to help you understand your personality type and how you can best interact with others. Strength Finder.In 2001 a new personality categorization method was introduced by Gallup [13]. The strength discovery technique divides the people in thirty four different classes based on their strength. This technique can be used in several different ways, but it is most commonly used in the context of career development. The Strength Finder technique is based on Positive psychology that everyone has a unique set of strengths and that these strengths can be used to help individuals achieve success in their careers[14]. This test is more regulative on a dynamic strategy, in comparison to MBTI [12]. Data Clustering is a type of data mining that groups similar data points together . Clustering is an unsupervised learning technique, which means that it does not require a training set of data with known labels.Text Clustering is a sub-type of data clustering which embeds Information Retrieval,machine learning and NLP [11].

In the proposed system, the user's personality is projected by analyzing the Twitter tweets using MBTI assessment . The Myers-Briggs Type Indicator (MBTI) is a widely-used personality assessment that helps in understanding the personality traits and how one interacts with others.

The proposed System will have 5 steps in its implementation Starting with training the algorithm on selected training data set , next step is to do Necessary prepossessing to remove irrelevant,noisy and ambiguity information which includes symbols and slang. To establish connection with twitter database using twitter API is the third step and most important step. By establishing the connection recent tweets can be extracted for analysis. in fourth step machine learning algorithms will be used for the classification of tweets into different categorizes and in the last step results will be displayed once user personality is mapped on MBTI. The data mining technique such as clustering and personality analysis will be performed on more than 3000 tweets. In section 2,Literature review and background

regarding data mining and classification of users based on different attributes will be delineated. Section 3, Discusses the methodology furthermore Section 4 elaborates the Social media platforms and analytics. Section 5 elaborates the results in detail. Section 6 concludes the research.

II. BACKGROUND AND LITERATURE REVIEW

Little work has been done on personality analysis domain in Pakistan as compare to other countries. Using big five personality Model is a lengthy work . Efforts are being made to acquire concise quantitative relation in psychology. As Gosling introduced Ten Item Personality Inventory (TIPI) [15] which discuss the 10 questions to specify the Big-Five personality traits. Various studies of Chittaranjan [16] implements TIPI to measure self-perceived personality. According to him its more appropriate to inquire a person how extroverted he is than to interrogate whether he delighted to being the company , attends Gatherings often , is expansive, forthcoming, social, and enthusiastic. Multiple questions reflecting one trait is minimized to one question to avoid verbosity, demoralization and accelerate the procedure , so that more people can participate in survey. According to S. Srivastava[17], People frequently updating and modify their social media profile with time, This action can be taken under consideration while to predicting the personality of the user . For some cases researchers have observed the behavior of users and concluded that personality alter over time [18].

There is still ample scope for social behavior-based analysis.Golbeck[19] stated in his paper that Performance needs to be improved for several personality predict scores from 11% to 18% of their actual values display very different characteristics.

Study of mentioned aspect is a topic of future research. J. Golbeck et al.[20]stated in their research that they were pioneers to light on the correlation between personality traits and social profile statistics. They developed a Twitter application through which they collected data-set oft 2000 tweets of 50 users. The users were presented with the 45-question version of the Big Five Personality Model. The Prepossessing of 2000 tweet corpus was done using 2 techniques, one LIWC (Linguistic Inquiry and Word Count) fro features extraction and other , MRC Database which which yielded 14 language features. They also performed a word-by- word sentiment analysis with the help of General Inquirer data-set.Paper shows that the authors performed various methods on corpus such as Correlation analysis ,regression analysis ,Gaussian Process and ZeroR with K-fold which resulted in prediction of scores from 10% to 18% of their actual values.

D. Quercia et al. [21] presented a research based upon on a group of 335 users having profile on 2 social-media platforms i.e Twitter and Facebook with construct that on us with multiple profiles on 2 social-media platforms. The Correlation of personality Twitters users was analyzed and the personality Assessment outcome was predicted on the basis of input of three features, Number of people following , Number of people followers and number of times the user is listed..Paper shows that the calculation done with k-fold cross validation taking k =10 using M5' rules. Th outcome of mean square error between the predicted and ascertained values is up to 0.88.For the sake of data integrity author further suggests to conduct a Test with validated users. C. Sumner et al. [22] took the research of personality prediction to the anti-social traits of the Big Five known as the Dark Triads of personality i.e narcissism, Machiavellianism and psychopathy. Author grouped profile attributes of 2927 Twitter users conducting a comparative study of total six models (SVM integrated with SMO and a polynomial kernel, Random Forest, J48 algorithm, Naïve Bayes Classifier with two Kaggle models, standard benchmark model and a competition winner model).After implementation of 6 models author concluded that models might not perform well for

predicting personality of a user but might work more efficiently predicting the trend of anti-social traits over a subset of population. In a research paper [23]Ana C.E.S Lim et al. Has predicted personality traits in group of tweets using machine learning techniques like muti-label classification This is a new conceptualization as a users may have multiple personality traits.The Naïve Bayes Algorithm is used to analyze tweets.The methodology has been divided into three Parts (prepossessing, transformation and classification) First few features will be extracted from twitter Data in prepossessing , then next step is to map the multi label sets into five dimensions of Five Factor Model training set. Finally, training sets and meta-attributes as input for the semi supervised classification. This paper [24] represents the concept of hash-tagging in twitter data and how supervised approach can b helpful for building data-set using this hash-tags. It further explains the issues regarding the compatibility of Training Data with Testing Data .Training data plays a vital part in performance of the complete system as training the system is entirely counts on it and classification of testing data is finished on the basis of Training data result only. Author stated 2 main problems that might occur . First similar format of Training data should be gathered for efficient and accurate results.Second main issue is Emoticons. Each Emoticons has its on meaning that is why thy are very important symbols present tweets, they emphasize the flings in tweets.

III. METHODOLOGY AND AND ANALYSIS REQUIREMENTS

The proposed model will predict the personalities that will be appropriate for all organization or businesses for better or more positive response. The domain of these companies can be quite diverse so recruitment team is analyze better what kind of personality is required for that certain job , it can be used externally for targeted marketing and preference.

A. *Myers Briggs Model*

The Myers–Briggs Type Indicator (MBTI) is a theory-based introvert self-report exam indicating different mental preferences in how people judge the world and make choices. It is based on Carl Jung's theory of personality types which is developed by Katherine Briggs and her daughter Isabel Myers. In this model people are identified as having one of 16 personality types. Basically, the model is classified in 4 different classes and in each class, there lies 4 personality types.

IS_ITE_J	IS_IFE_J	IN_IFE_J	IN_ITE_J
Inspector	Protector	Counselor	Mastermind
IS_ETIP	IS_EFIP	IN_EFIP	IN_ETIP
Crafter	Composer	Healer	Architect
ES_ETIP	ES_EFIP	EN_EFIP	EN_ETIP
Promoter	Performer	Champion	Inventor
ES_ITE_J	ES_IFE_J	EN_IFE_J	EN_ITE_J
Supervisor	Provider	Teacher	Fieldmarshal

Table 1 Temperament Mapping on Myers–Briggs groupings SP, SJ, NF, and NT

The main web application is developed using Atom IDE and MongoDB Web framework, with a connected document similarity module implemented in Flask. MongoDB is a embedded with the proposed system as it's a Python-based publicly available web framework, which follows the model-template-view architectural pattern.

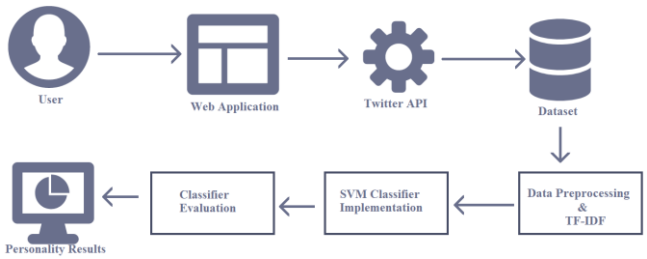

Figure 2 Proposed System

B. Collection of Data

Data is divided in two categories -Training and Testing. For training purpose Publicly Available data on Kaggle was used "Myers–Briggs personality type data-set", which contains 6000 data entries, each entry comprise of 2 columns. (MBTI personality type , fifty tweets from each user).

Average no.of tweets/user	15
Average no.of words /user	401
Average no of Characters/ user	2020
Average word_length of tweets/user	4

Table 2 Data Sets Statistics

For Testing we will gather data-set from the individual's social media. The process begins by collecting the twitter data. Twitter API provides us likes, tweets and number of followers. The tweets are expanded by classifying them into four Main categories i.e extraverted (E), introverted (I), sensing (S), and thinking (T). Myers–BriggsModel has evidenced predictive validity and compatibility with earlier psychological Sciences & Marketing with understandable dimensions.

Text mining and personality analysis are implemented on each user public profile based on their recent tweets Twitter API allows developers to use of twitter features without accessing website interface. It also facilitates automated tweet posting or message sending .By using a a set of URLs provided by API several Tweets were collected to analyze the user's personality.

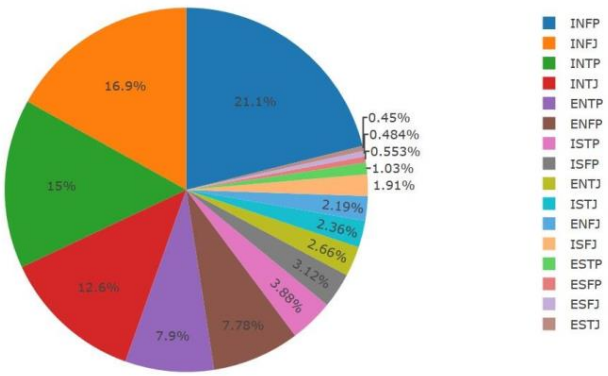

Figure 3 Personality occurrences Distribution

C. *Types indicators in the data-set*

To classify and understand the distribution of the type indicators in data-set, 4 new Rows was added in data-set. Each row will represent a different category of type indicator in distribution. The first category was for is_Extrovert(E), the second
category is_Sensing (S), the third is_Thinking](T)]and the fourth category is_Judging (J). the output on computation will be that one Alphabet from the assigned will return for each category, on Assignment of 4 alphabets one Personality out of 16 MBTI personality types will be mapped for example if the first loop is returning E, the second is returning S, the third is returning F and the last category is returning J, the resultant MBTI personality type will be ESFJ.

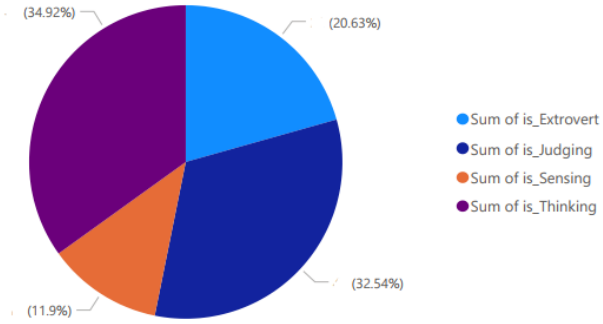

Figure 4 : Classification Type indicators

D. *Data Preprocessing*

After data is collected , data cleaning process is the next step. Texts which are written by humans usually contain noise, hence ewe have to remove the noise first , to extract the features out of the Tweets .To make computer understand the human language commands need to be translated using Natural Language Processing which translates high level language to language that machines understand[32].
Inspection of the data-set,uncover any issues like noisy and redundant data including spelling mistakes in the raw data. To address this issue,the whole data was pass inspected by spellchecker . For Tokenization and punctuation Removal, Open-CV was used.Removal of URLs, symbols, names, spaces, and lower cases was include in getting rid of Stop-words.Sometimes, in tweets the user writes a lot about his feelings. So, in these type of tweets data is complex as compared to a normal tweet. To deal with this, text summarization is done before passing it to the data soothing process. Missing Values were checked using

```
df.isnull().any()
```

Figure 5 To Check Missing and Null Value

Confirmation of data-set shape is done to check if there is any ambiguity.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5000 entries, 0 to 5000
Data columns (total 2 columns):
 #   Column  Non-Null Count  Dtype
---  ------  --------------  -----
 0   personality 5000 non-null   object
 1   posts       5000 non-null   object
dtypes: object(2)
```

Figure 6 Data-set Shape Code

E. Data Visualization

Visualization tools are one of the important components when dealing with data-set as help in communicating the information more clearly as compared to table and text .Tools provide us with the visualization of data in some graphical descriptor[31] . Below is the overall process of the data analytics.The process starts with analytic Dashboard,which can be designed using Microsoft Power BI and for just Plotting we used Python which includes data analysis with Data visualization.

Data → Data Analysis→ Data Visualization

After Preprocessing it is clearly seen that there is visible unbalance in Introvert/Extrovert and Intuition/Sensing pairs. Whereas other two are quite balanced. For confirmation we plotted a graph.

```
# visually showing the class imbalance
personality_data["personality"].value_counts().plot(kind="bar", color="#61BED6")
plt.show()
```

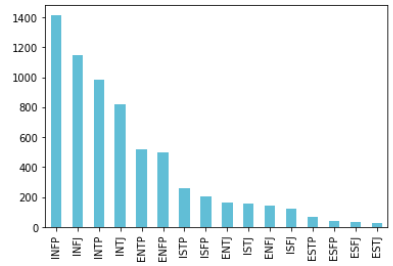

Figure 7 Data Imbalanced (Visual Representation)

Personality Type	Frequency in Population
ISFJ	13.8%
ESFJ	12.3%
ISTJ	11.6%
ISFP	8.8%
ESTJ	8.7%
ESFP	8.5%
ENFP	8.1%
ISTP	5.4%

Table 3 MBTI Personality distribution in population

F. Feature Extraction

WordNet is a lexical database for the English language that was developed by Princeton University. It organizes English words into sets of synonyms called synsets, providing a hierarchical structure of meaning through relationships such as hypernyms (more general terms) and hyponyms (more specific terms). WordNet is used for enhancing the quality of the feature aggregation in Proposed mode[31]. WordNet supports the mapping of English words onto their synonym (synset). By performing this ,the words having same content corresponds to same personality Class.One column is added for each MBTI Personality Indicator pair,because for training ,independent classifier model will be used for each pair Python Library that will be used is sklearn.feature_extraction.text, CountVectorizer, Tfidf Transformer. The reason is the imbalance Nature of database:

```
personality = df.personality.values
```

```
yIE = df.ie.values
yNS = df.ns.values
yFT = df.ft.values
yPJ = df.pj.values
y = df.type
```

- TF-IDF:Term Frequency-Inverse Document Frequency is a numerical statistic used to weight the importance of words in a document.It was observe pattern's words like , he, or; what, will she, and, it repeats commonly. As they have no contribution in personality analysis,they are removed. TF-IDF transformer is a tool that is used to convert a collection of raw text documents into a matrix of TF-IDF features[35]. The transformer takes in recent tweets and calculates the TF-IDF score for each word.After that TF-IDF is creating document frequency matrix:

```
cv = CountVectorizer().fit(posts)
X = cv.transform(posts)
tf = TfidfTransformer()
```

Figure 8 TFIDF code

- Classifier: Support vector machines (SVM) classification algorithm is used in proposed system . After text passes through feature extraction process using NLP techniques. Then, the features along with label are passed to supervised ML models. After the validation and training process system classify the given unknown inputs. the sklearn library is used for feature extraction process.The aim is to acquire the best hyper-plane using SVM that will separates the classes in linear and non-linear SVM .

```
from sklearn.svm import SVC
svm = SVC(random_state = 1)
svm.fit(X_train, y_train)

Y_pred = svm.predict(X_test)

predictions = [round(value) for value in Y_pred]
# evaluate predictions
accuracy = accuracy_score(y_test, predictions)
accuracies['SVM'] = accuracy* 100.0
```

Additionally, data-set will be divided into testing and training in multiple ratios to find out which gives the best results.

IV. TESTING

When it comes to text classification Support vector machine (SVM) are best suited as the SVM algorithm finds the best boundary that separates the data into different classes. This is the reason that SVM was formulated for purposed system , to improve the speed along with performance as the data-set is not linearly separable.The algorithm uses a technique called the kernel trick to transform the data into a higher-dimensional space where it may be linearly separable.SVM finds the hyperplane that maximizes the margin, which is the distance between the hyperplane and the closest data points from each class. These closest points are called support vectors.

FINDINGS

In Table 4 it can be seen that when the classifier Predicted the System model ,it performed poorly ,but the distinct model performed well in prediction of Type indicators.

Type indicators	SVM results

IE (distinct model)	0.7863241
NS(distinct model)	0.8526425
FT(distinct model)	0.8412365
PJ(distinct model)	0.7284636
Model	0.3847622

Table 4 SVM Score

Naive Bayes classifier has been used a benchmark in many studies[34] due to its out-performance when compared with outer classifies.The data was always tested with Naive Bayes to Compare the Results of purposed model.

Type indicators	SVM	Naive Bayes
IE	78%	77%
NS	85%	83%
FT	84%	77.9%
PJ	72%	62.3%
Model	39%	44%

Table 5: Comparison of SVM & Naive Bayes

V. CONCLUSION

The aim was to find a correlative link between user's post on Twitter and their personality type. Social media gives its users a platform to put forward themselves openly and therefore those tweets can be used as an indicator of their personality type.Based on the scores, users are assigned a four-letter personality type, such as INTJ or ESFP. The assessment is often used in self-improvement, career development, and team building, but it's important to note that the MBTI assessment is not scientifically validated .The purposed system is an automatic Process of Predicating one's personality and the results can be used in various domain such as HR , Psychological Science , Business.For the development process TF-IDF,Word Vector SVM were used along with Python libraries like Pandas, Numpy and Sklearn . The Results of SVM model are documented and the performance was compared with Naive Bayes implemented on same data-set. The Implemented SVM gave 39% overall accuracy.

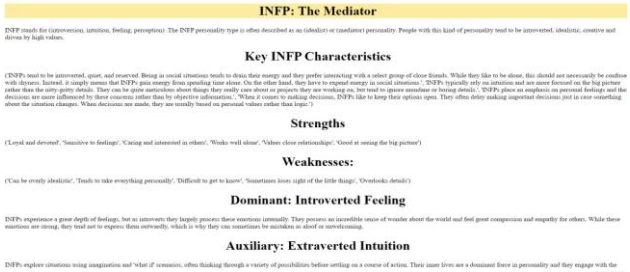

Figure 9 :Output

Further work can be done in improving the results by implementing multiple classifiers with different filtering techniques such as collaborative approach.Now twitter has many Roman Urdu user's also .NLP can be merged also in this model.

REFERENCES

[1] Bakos, Y., & Katsamakas, E. (2008). Design and ownership of two-sided networks: Implications for Internet platforms. Journal of Management Information Systems, 25(2), 171-202.*)*

[2] Rumpf, H. J., Effertz, T., & Montag, C. (2022). The cost burden of problematic internet usage. Current Opinion in Behavioral Sciences, 44, 101107. 3. Chaffey, D. (2016). Global social media research summary 2016. Smart Insights: Social Media Marketing.

[3] Chaffey, D. (2016). Global social media research summary 2016. Smart Insights: Social Media Marketing

[4] Perrin, A. (2015). Social media usage. Pew research center, 125, 52-68..

[5] Buttazzo, G. (2006). Research trends in real-time computing for embedded systems. ACM SIGBED Review, 3(3), 1-10.

[6] Carvalho, J. P., Rosa, H., Brogueira, G., & Batista, F. (2017). MISNIS: An intelligent platform for twitter topic mining. Expert Systems with Applications, 89, 374-388.

[7] Myers, I. B., McCaulley, M. H., & Most, R. (1985). Manual: A guide to the development and use of the Myers-Briggs Type Indicator(Vol. 1985). Palo Alto, CA: Consulting Psychologists Press

[8] Shaffer, D., Schwab-Stone, M., Fisher, P., Cohen, P., Placentini, J., Davies, M., ... & Regier, D. (1993). The diagnostic interview schedule for children-revised version (DISC-R): I. Preparation, field testing, interrater reliability, and acceptability. Journal of the American Academy of Child & Adolescent Psychiatry, 32(3), 643-650

[9] Rath, T. (2007). StrengthsFinder 2.0. Simon and Schuster.

[10] Judge, T. A., Higgins, C. A., Thoresen, C. J. and Barrick, M. R. (1999), THE BIG FIVE PERSONALITY TRAITS, GENERAL MENTAL ABILITY, AND CAREER SUCCESS ACROSS THE LIFE SPAN. Personnel Psychology, 52: 621–652.

[11] Scurlock-Evans, L., & Upton, D. (2015). The role and nature of evidence: A systematic review of social workers' evidence-based practice orientation, attitudes, and implementation. Journal of evidence-informed social work, 12(4), 369-399

[12] .Reynierse, J. H., Ackerman, D., Fink, A. A., & Harker, J. B. (2000). The effects of personality and management role on perceived values in business settings. International Journal of Value-Based Management, 13(1), 1-13

[13] .Buckingham, M., & Clifton, D. O. (2001). Now, discover your strengths. Simon and Schuster

[14] Rios, G. R., & Craven, H. J. (2019). Organizational Dynamics with StrengthsFinder® Facilitation.

[15] S.D. Gosling, P.J. Rentfrow, W.B. Swann Jr, "A very brief measure of the big-five personality domains", Journal of Research in Personality, Volume 37 Issue 6, pp. 504- 528, 2003.

[16] G. Chittaranjan, J. Blom, D. Gatica-Perez, "Mining largescale smartphone data for personality studies", Journal Personal and Ubiquitous Computing, Volume 17 Issue 3, 2013.

[17] Y. Amichai-Hamburger and G. Vinitzky, "Social network use and personality", Computers in Human Behavior, Volume 26 Issue6, pp. 1289–1295, 2010.Buttazzo, G. (2006). Research trends in real-time computing for embedded systems. ACM SIGBED Review, 3(3), 1-10.

[18] .S. Srivastava, O. John, S. Gosling, and J. Potter, "Development of personality in early and middle adulthood: "Set like plaster or persistent change?"", Journal of Personality and Social Psychology, Volume 84 Issue 5, pp. 1041–1053, 2003

[19] S. Adali and J. Golbeck, "Predicting Personality with Social Behavior", Proc. the 2012 International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2012), pp. 302-309, 2012

[20] Golbeck, Jennifer, et al. "Predicting personality from twitter." Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011.

[21] .Quercia, Daniele, et al. "Our Twitter profiles, our selves: Predicting personality with Twitter." Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third Inernational Conference on Social Computing (SocialCom), 2011 IEEE Third International Conference on. IEEE, 2011.

[22] Sumner, Chris, et al. "Predicting dark triad personality traits from Twitter usage and a linguistic analysis of tweets." Machine Learning and Applications (ICMLA), 2012 11th International Conference on. Vol. 2. IEEE, 2012.

[23] Lima, Ana CES, and Leandro N. De Castro. "Multi-label Semi-supervised Classification Applied to Personality Prediction in Tweets." Computational Intelligence and 11th Brazilian Congress on Computational Intelligence (BRICS-CCI & CBIC), 2013 BRICS Congress on. IEEE, 2013

[24] Kouloumpis, E., Wilson, T., & Moore, J. (2011). Twitter sentiment analysis: The good the bad and the omg!. In Proceedings of the international AAAI conference on web and social media (Vol. 5, No. 1, pp. 538-541).

[25] D'Andrea, E., Ducange, P., Lazzerini, B., & Marcelloni, F. (2015). Real-time detection of traffic from twitter stream analysis. IEEE transactions on intelligent transportation systems, 16(4), 2269-2283.

[26] Lin, L., Lin, C., & Lai, Y. (2018, October). Realtime event summarization from tweets with inconsistency detection. In International Conference on Conceptual Modeling (pp. 555-570). Springer, Cham.

[27] Washha, M., Qaroush, A., Mezghani, M., & Sedes, F. (2019). Unsupervised collective-based framework for dynamic retraining of supervised real-time spam tweets detection model. Expert Systems with Applications, 135, 129-152.

[28] Obschonka, M., Fisch, C., & Boyd, R. (2017). Using digital footprints in entrepreneurship research: A Twitter-based personality analysis of superstar entrepreneurs and managers. Journal of Business Venturing Insights, 8, 13-23.

[29] Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012, December). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In 2012 11th international conference on machine learning and applications (Vol. 2, pp. 386-393). IEEE.

[30] Guntuku, S. C., Lin, W., Carpenter, J., Ng, W. K., Ungar, L. H., & Preoţiuc-Pietro, D. (2017, June). Studying personality through the content of posted and liked images on Twitter. In Proceedings of the 2017 ACM on web science conference (pp. 223- 227)

[31] Nazir, M., Ishtiaq, M., Batool, A., Jaffar, M. A., & Mirza, A. M. (2010, June). Feature selection for efficient gender classification. In Proceedings of the 11th WSEAS international conference (pp. 70-75).

[32] Masood, S., Soto, I., Hussain, A., & Jaffar, M. A. (2014, November). Statistical features based noise type identification. In Mexican International Conference on Artificial Intelligence (pp. 231-241). Springer, Cham.

[33] Tahir, M. U., Naqvi, M. R., Shahzad, S. K., & Iqbal, M. W. (2020, February). Resolving data de-duplication issues on cloud. In 2020 International Conference on Engineering and Emerging Technologies (ICEET) (pp. 1-5). IEEE.

[34]] L. C. Lukito, A. Erwin, J. Purnama and W. Danoekoesoemo, "Social media user personality classification using computational linguistic," 2016 8th International Conference on Information Technology and Electrical Engineering (ICITEE), Yogyakarta, 2016, pp. 1-6.

[35] Koloski, B., Pollak, S., Škrlj, B., & Martinc, M. (2021). Extending neural keyword extraction with TF-IDF tagset matching. arXiv preprint arXiv:2102.00472.

.