Predicting the next word using the Markov Chain Model according to profiling personality

Bahaa Eddine ELBAGHAZAOUI (elbaghazaoui.bahaa@gmail.com)
Université Ibn-Tofail
Mohamed AMNAI
Université Ibn-Tofail
Youssef FAKHRI
Université Ibn-Tofail

Research Article

Keywords: Personality, Big five Model, Data profiling, Prediction, Markov Chains

Posted Date: July 29th, 2022

DOI: https://doi.org/10.21203/rs.3.rs-1879234/v1

Predicting the next word using the Markov Chain Model according to profiling personality

Bahaa Eddine Elbaghazaoui[1*], Mohamed Amnai[1†] and Youssef Fakhri[1†]

[1*]Laboratory of Computer Sciences, Faculty of Sciences, IbnTofail University, Kenitra, Morocco.

*Corresponding author(s). E-mail(s): elbaghazaoui.bahaa@gmail.com;
Contributing authors: mohamed.amnai@uit.ac.ma; fakhri@uit.ac.ma;
†These authors contributed equally to this work.

Abstract

Understanding human data has been the focus of philosophers and scientists. Social media platforms encourage people to be creative and share their personal information. By analyzing data, we will be able to identify people's personalities and information that is also important to specific profils. The aim of this paper is to propose an approach that predicts the next word during writing a sentence based on the user's personality. To achieve this goal, our approach is illustrated by two points: (1) An approximate extraction of the Big Five model for a specific user from his tweets. (2) Predicting the next word while a user is writing a new tweet depends on his personality using the Markov Chain Model. On the basis of these two notions, our approach makes writing posts easier by predicting and suggesting next words based on the user's personality. Experience represents the ease of predicting the next word during the writing of a new post related to individual potential.

Keywords: Personality, Big five Model, Data profiling, Prediction, Markov Chains

1 Introduction

Personality recognition has been an interesting topic in the domain of psychology [1] as it has profound implications for studying personal interactions. The personality influences life choices, happiness, and many other behaviors. It's critical to anticipate the preferences and actions of the people we interact with in order to establish effective cooperation. Personality recognition is applied in many fields and has several methods to extract it. Most of those studies have emerged from texts in which they focus on the analysis and examination of textual samples. Several studies have found a strong correlation between personality traits and linguistic characteristics. There are many models used to profile user personalities [9].

Numerous strategies have been put out to determine user personality automatically from the content produced [23]. The accuracy of the data gathered is a key factor in how well these strategies perform. In this paper, we chose the Big Five Personality as the base model to extract user personality because it is the most widely used and appropriate in predicting [22]. Dreamtalent writes a post on his own blog with the headline "Stop Using MBTI & DISC, They're Not That Good" to recommend using the big five personalities in recruiting.

Alternatively, people text each other frequently, and every time users try to compose a text, a suggestion pops up, trying to predict the next word they want to enter. One of the applications that NLP deals with is the prediction process [38]. However, When a user is entering text on a mobile device, it can be helpful to suggest the next word so that typing time is reduced and errors are avoided. The problem we found is that the suggested prediction words sometimes are not compatible with user needs. However, smart phones suggest words based on historical user data or selected language dictionary [30] and for search engines like "Google, Yahoo, Bing..." suggest next words based on geographic data or user profile [29]. For our approach, we aim to inject user personality as a parameter to predict the next word using markov chains.

Knowing one's personality can help a person gain a better understanding of himself and the people around him. It can assist him in identifying his strengths and shortcomings, comprehending his emotions and actions, and regulating his conduct in various settings [39]. The availability of a significant amount of high-dimensional data has cleared the path for marketing initiatives to be more effective by targeting specific consumers. Personality-based communications are extremely effective at increasing product and service acceptance and attractiveness.

In our approach, we focus on profiling personality characteristics from users' tweets. Then, our program predicts and suggests words based on his personality when writing a new tweet. This facilitates the writing of tweets for each user based on their personality, and mainly to give suggestions and orientations for not posting aggressive texts.

By summarizing, the document structure is as follows: In the first section, we provide an overview of relevant publications as well as the main problem that our technique addresses.The second section introduces the main idea of

the solution. In the third section, we introduce our implementation and discussion. Finally, we provide a general conclusion and suggestions for additional research.

2 Related Works

Much research has been conducted to predict personality on social media [24]. For example, Agarwal used text from the myPersonality corpus to detect personality in his research [7]. Pednekar and Duny conducted a data mining method using social media to identify the essence of personality [2]. In their research, Kosinski et al, have identified the personality patterns of Facebook users [15]. Various methods using different classifiers and feature spaces have been proposed for clustering human's personality. Until recently, the majority of models relied on shallow learning techniques like Support Vector Machines (SVM) [8], Naive Bayes classifier [12] , K-Nearest Neighbors (kNN) [11], and Logistic Regression (LR) [16].

Currently, there are several ways used to profile the user's personality, such as the Big Five Personality or OCEAN model [13], Myers Briggs Type Indicator (MBTI) [19] and Dominance Influence, Steadiness, Conscientiousness (DISC) [10].

Generally, most of previous Next Word Suggester/Predictor concentrate on two models, N-grams model or Long Short Term Memory (LSTM). However, The statistical language model has been presented for a long time, with S. Katz introducing a nonlinear recursive algorithm to solve ngram language prediction in 1987 [31], and Stolcke proposing a language model using a method called Bayesian Learning [32]. Yoshua et al. went on to develop a distributed representation for words that improves the n-gram model [33]. Great feed-forward networks were built using LM in [34]. Recently, M. Sundermeyer et al. [35] developed an upgraded LSTM, while Mikolov et al. [36] designed an RNN model.

Traditional models are easy to use and perform better than deep learning models in some situations [37]. We chose the n-gram model based on Markov chain for our project for the following reasons: Markov chain is very insightful. It can identify the areas of any process where we are deficient, allowing us to make changes in order to improve. The memoryless quality of a stochastic process is referred to as the Markov property. Also, any size of system can readily determine it's very low or modest computation requirements.

In general, our approach would focus on profiling the big five personalities from Twitter posts. Using the Markov Chain Model, we will be able to predict the next word when writing a new tweet targeted at a user's personality.

3 Problem formulation

To completely understand their users' activity, several research initiatives are working on gathering metadata from their products and platforms [25]. Understanding user behavior, on the other hand, pushes companies to improve the

quality of their various products and services [3], and then present their products to fit the user's desire [6].When searching in search engines or writing on a specific platform, however, we notice that most solutions suggest words or sentences that follow what we have already written as the proposition of what we want to write. Its recommendations are based on past research done by other users and even by their regions [26]. Sometimes the user does not object if he finds suggestions that contradict his principles, culture, or general personality. this situation, companies will lose their users. Therefore, our objective is to analyze the user's personality through their historical data and suggest to him targeted data appropriate to his personality and behavior. In this paper, we will focus on extracting the big five personality traits from previous tweets of a user and then suggest the next word for his sentences when he is writing a tweet using the Markov Chain Model.

4 Research Method

By the 1990s, it had been commonly understood that both situational and personality factors influenced short-term behavior [14]. The research and improvement of the OCEAN, or big five model, was still in progress, proving its influence until today [27]. The big five personality traits, or "ocean traits," are presented as follows:

- **Openness:** This indicates a willingness to try new things and think outside the box, and is sometimes referred to as intelligence or imagination. Insight, inventiveness, and curiosity are all qualities to look for.
- **Conscientiousness:** Need to control the desire for immediate gratification by being careful, vigilant, and self-disciplined. Ambition, discipline, consistency, and dependability are all characteristics.
- **Extroversion:** Rather than a person, a state in which an individual pulls energy from others and wants social relationships or contact (introverted). Outgoing, active, and self-assured are characteristics.
- **Agreeableness:** The way a person interacts with others, as measured by their level of compassion and collaboration. Wit, sociability, and loyalty are among the traits.
- **Neuroticism:** Negative personality traits, emotional instability, and self-destructive thoughts are all risk factors. Pessimism, anxiety, insecurity, and fear are all characteristics of the human personality.

Our approach has two main points, as shown in the following figure 1. First point (1: Extract Big Five Personalities): We're going to focus on a real database from one of the social networks, then extract all the posts filtered by their users. Through data profiling [20] [21], we need to extract an approximate personality score of a user from an existing library and store the results in our repository. Second point (2: calcul distance & predict next word), predict the next word for the new post based on the user's personality, and we can also adapt our algorithm to suggest words from another personality.

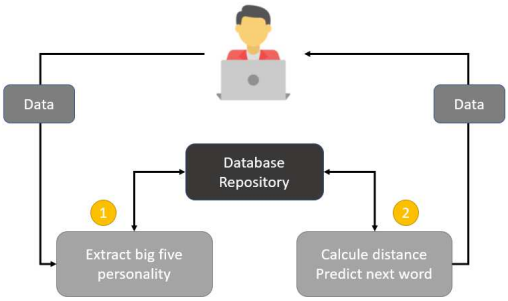

Fig. 1 Workflow approach

In the first point of our approach, we were inspired by Navonil Majumder et al. paper [17]. Its main idea is to use deep learning algorithms to detect personality from text based on document modeling [18]. Moreover, the approach focuses on processing the input data in a hierarchical manner by analyzing and evaluating every single word, then combine words to make n-grams, n-grams to make sentences, and sentences to make a complete document.

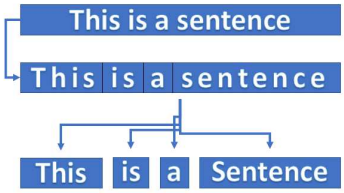

Fig. 2 N-Grams concept

After editing the program to extract the five characteristics necessary to detect the personality, We did a separate analysis for each user to detect their personality and classify them among other users. In our case, we take all the tweets that are made by a single user and calculate the average value as a representation of his permanent personality. This analysis could be done for an interval of time. However, when a user wants to write a new post, the system already knows his personality, so it will suggest the next words for his tweet depending on his personality and his behavior.

Jack Dorsey, the founder of Twitter, posted the first tweet on March 21, 2006. The billionth tweet was not reached until the end of May 2009, after a three-year wait. One billion tweets are sent in less than two days nowadays. So, we are speaking of data, which is extremely hard to calculate. For that, we propose a method to classify the results of the personalities that we have already profiled. However, the main framework of personality psychology is the Big Five model. Characteristics of personality are conceptualized as five independent continuous dimensions in this model. If we divide each dimension at the median to produce personality types, we get 32 different types, in which

individuals are above or below the median in **neuroticism, extroversion, openness, agreeableness, and seriousness**. If these five dimensions are completely independent of each other, we will see that individuals are equally assigned to one of 32 types.

In the second point of our approach, we have to predict the next words when the user writes a new tweet. In fact, the system already knows the average personality of each user as well as their classification. Therefore, the system will suggest the next word in the cluster to which the user belongs. In this part, we used the Markov chain algorithm.

Markov chain is a type of stochastic process that is distinct from others, in that it must be "memory-less" [28]. Future acts, on the other hand, are not reliant on the steps that lead to the current situation. The Markov property is the name for this. Markov chains theory is important, because so many discrete processes $(X_n)_{n\geq 0}$ satisfy the following Markov property:

$$P(X_n = i_n | X_{n-1} = i_{n-1}) = P(X_n = i_n | X_0 = i_0, X_1 = i_1, \ldots, X_{n-1} = i_{n-1}) \tag{1}$$

This indicates that all of the knowledge needed to forecast the future is included in the current state of the process and is not dependent on previous states.

The most obvious example in probability theory is a time-homogeneous Markov chain [5], in which the likelihood of any state transition is unaffected by time A labeled directed graph can be used to visualize such a process, for which *the labels of any vertex's outgoing edges add up to 1.*

For example, the Markov chain (homogeneous in time) based on the foundations of states A and B as shown in the figure 3 below. To move from A to B after 2-steps, the process must stay on A in the first step and then move to B in the second step, or move to B first and then to B again. According to the figure, the probability is $0.3 * 0.7 + 0.7 * 0.2 = 0.35$. Alternatively, the likelihood of the process being on A after two movements is $0.2 * 0.8 + 0.8 * 0.3 = 0.65$.Because the chain has just two states, if the process is not on A, it must be on B. As a result, the likelihood that the process will be on B after two steps is $1 - 0.65 = 0.35$.

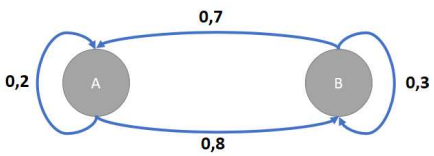

Fig. 3 Markov Chain Graph

Note again that the last formula 1 expresses the fact that, for a given historical record, only the current state, not the previous state, the probability distribution of the following state is influenced.

Now, we will focus on predicting words. Assume we wish to create a system that, when given an incomplete sentence, will make an attempt to guess the next word in the phrase. To deal with word prediction cases like this, we model it as a Markov model problem. Each word must be treated as a state (i_t) and the next word must be predicted based on the previous state ($i_t - 1$).This situation is very suitable for the Markov Chain Model. To emphasize this point, all the unique words from our database could form different states. Moreover, the probability distribution consists of determining the probability of a transition from one word to another.

To fully understand the application of markov chains in our approach. Considering the following example, let's take three sentences as follows:

- I like Engineering.
- I like Science.
- I love Mathematics.

All of the words in the preceding sentences are unique, "I", "Like", "Love", "Engineering", "Science" and "Mathematics" can form different states. In other words, probability distributions are all related to figuring out the chances of a transition from one state to another, in this example, the transition from one word to another state. In this scenario, it is clear from the above example that the first word always starts with the word "I". Therefore, the first word of the sentence has a 100% chance of being "I". In the second state, we must choose between the two words "like" and "love". The probability distribution now represents the likelihood that the next word is "like" or "love", assuming the previous word is "I". The word "like" appears in two of the three phrases after "I" in our example, although the word "love" appears only once. Therefore, there is approximately a 67% or 2/3 probability of successfully obtaining "like" after "I", and a 33% or 1/3 probability of "love" occurring. Similarly, the probability of success for "Science" and "Engineering" is fifty fifty. In our case, "mathematics" is always after "love". The following figure 4 represent graphically the probability of a transition between words.

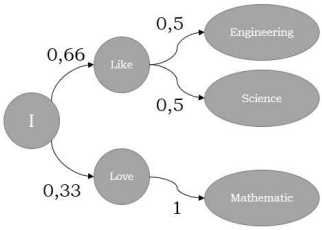

Fig. 4 State transition diagram for our example

To generalize, we propose the following algorithm 1 as the basic procedure for our approach.

Algorithm 1 Markov Model Chain to predict next word

Require: *tweets*
Ensure: $first\_possible\_words, transitions$

```
first_possible_words = {}
transitions = {}
for tweet in tweets do
    length = tweet.split().length()
    for word in tweet.split() do
        if word.isFirst() = True then
            first_possible_words[word] = first_possible_words.get(word)
        else
            prev_word = word.previous()
            if word.isLast() = True then
                Expand(transitions, (prev_word, word), "END")
            else
                Expand(transitions, (prev_word, word), word)
            end if
        end if
    end for
end for
procedure EXPAND(dict, key, value)
    if key not in dict then
        dict[key] = []
        dict[key].add(value)
    end if
end procedure
```

The concept of our algorithm is to start by defining two variables:

- **first_possible_words:** contains the first words of each sentence of the tweets with its transistions properties.
- **transitions:** all the possible transitions between states (the words which are related to each others).

Our algorithm starts by decomposing every tweet from the previous suggestions into words. Through analysis of the words, we save all the first unique words in the first variable with their possible probabilities of starting. Then, if the word is located at the end of some sentence, the program will store the word and propose the next word as "END". Otherwise, we will store all words that have a relation between them and their transition probability.

5 Implementation and Discussion

We started the realization by implementing the first point of our approach. However, we have uploaded a database from the Kaggle platform under the name [4], this database was extracted through the Twitter api. It contains 1,600,000 tweets from 659,775 different users. We have been inspired by a library called "bigfive" by the official package manager for NodeJs programs (NPM). This library was created by Peter Hughes, which used the big five lexica data from the World Well-Being Project and obtained a Creative Commons Attribution-Non-commercial-Same Sharing 3.0 Sharing License (CC BY-NC -SA 3.0). Through this library, we could analyze and extract approximately five-factor model features from each tweet.

To optimize our analysis, we performed the five personality traits on 100, 1000, and 10,000 tweets from the dataset. We get similar results from one input to another. The error between each conversion is almost non-existent (the standard error does not exceed 2%). The figure 5 below shows that we could say, the mean average is close between one stage and another. Also, throught analyzing dataset, we profiling some metadata concerning the tweets of the population, as shown in the table 1 below. The result concluded that we could generalize the mean average for the big five personality traits for the whole dataset.

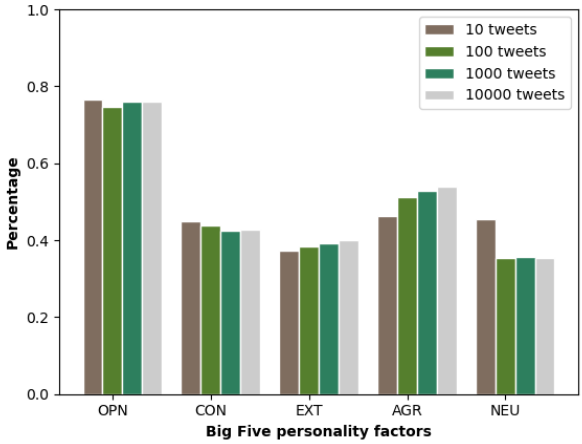

Fig. 5 Extract the mean five personality traits for 10, 100 ,1000 and 10000 tweets

Now, we will extract the approximate personality from a random user. First, and using our data source as a search engine,The user who has multiple tweets and is also active in Twitter is (lost_dog). Between May 1 2009 and June 25 2009, he posted 549 tweets. After profiling his personality, we found the results as shown in the following figure 6.

Table 1 Metadata about dataset's five personality

	OPN	CON	EXT	AGR	NEU
Mean (Average)	0.75786925	0.43517075	0.38722125	0.5108205	0.354028
Sample Standard Deviation	0.00790346	0.01118990	0.01118083	0.0345270	0.001233
Standard Error (SE x)	0.00395173	0.00559495	0.00559041	0.0172635	0.000712

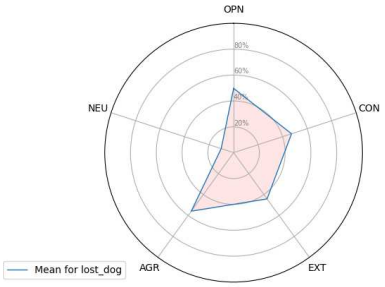

Fig. 6 Lost_dog's mean big five personality

Through all the metadata that we extracted in the previous process. We could suggest decomposing the database into 32 clusters. However, each personality trait could break the database down into two groups, one of which has a higher than average component value and the other lower. Therefore, five components give $2^5 = 32$ clusters. So, applying this classification to our most frequent user in our dataset (lost_dog) with 9917 tweets, we found the results illustrated in the following table 2. However, the table can be described as follows. Indeed, the vertical columns are the information for each type of big five personality, for the vertical columns, we have "top" which describes if the user frequents on a positive or negative value for each type. the row "freq" describes how many times to have this type negative or positive and "avg" describes the average value for each type of personality.

Table 2 Comparing lost_dog status with the dataset

	EXT	NEU	AGR	CON	OPN
top	Negative	Negative	Positive	Negative	Positive
freq	5707	6200	5268	5361	7370
avg	0.441674	0.101543	0.558951	0.472545	0.497207

We will now establish a relationship between the user's personality and the personality of others. and all of the terms in the database. We all know that this dataset is only a small part of the actual database. Therefore, we suggest calculating the distance between the user's average personality and the tweets

that are included in his cluster or even in the clusters closest to his personality. For example, through the Euclidean distance, we will predict the words only from the 1000 first tweets at the closest distance.

In the second part of our approach, when the user composes a tweet, we must recognize the next word prediction based on his personality. In fact, through the Markov chain algorithm. The system is always on, so when the user writes a word, the system analyzes the data (words and sentences) that depends on his personality, and suggests the most relevant words to continue the sentence, thus accepting his personality. These suggestions are listed in order, from most appropriate to least. In our system, we configure the program to suggest only the first four words, as shown in the figure 7 below.

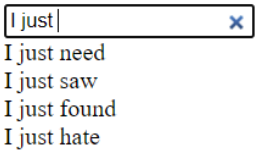

Fig. 7 Application of the approach

Through our approach, we could suggest words when writing a sentence according to a personality, either from the user himself or to give another personality. In fact, we have applied our program to two datasets, the first without any modification. We get the tweets randomly from the database, and then we apply the second part directly. And in the second experience, we respect the procedure of our approach and we apply both parts. The results obtained after training our algorithm are illustrated in the following figure 7. The program is based on the fact that each time the user enters a word, it reviews the suggested words depending on their probabilities.

Application 1 : *Applying the second part without profiling personality.*

- **I** [('just', 0.0666), ('hate', 0.0631), ('need', 0.0368), ('had', 0.0192), ('really', 0.0333)]
- **just** [('re-pierced', 0.0070), ('hate', 0.0070), ('thought', 0.0070), ('found', 0.0354), ('saw', 0.0354)]
- **hate** [('saying', 0.25), ('you', 0.336)]
- **saying** [('bye', 1.0)]
- **bye** [('END'), 1.0]

Application 2 : *Applying approach with profiling user personality.*

- **I** [('really', 0.0333), ('have', 0.0666), ('wish', 0.0596), ('wanna', 0.0122), ('miss', 0.0456)]
- **really** [('dont', 0.0307), ('miss', 0.0461), ('feel', 0.0153), ('really', 0.0461), ('just', 0.0307)]
- **miss** [('my', 0.2), ('real', 0.2), ('u', 0.2), ('you', 0.2), ('the', 0.2)]

- **you** [('too', 0.0937), ('terribly!', 0.0312), ('on', 0.0312), ('END', 0.1562), ('too!', 0.0625)]

Based on our observations,Applications (1) and (2) have the same vision, which is to predict the next word when writing a new post. The difference is that the first application (1) does not filter when using the database, but relies on historical tweets. The second application (2) uses our approach (predicting the next word based on the user's personality). In fact, we found a huge difference between the outputs of the experience. We focused on writing both sentences with the meaning "*I just hate saying goodbye*" and "*I really miss you*". Indeed, depending on the big five of the "lost_dog" personality, we calculate the distance between the two sentences and his personality. We notice that the first result (without profiling) is far from 0.54, but the second (with profiling) is only far from 0.24. This means that the second sentence is more relevant than the first. Therefore, our approach has successfully outperformed the results of the previous forecasting methods.

6 Conclusion & Future works

In this paper, we profile user personalities based on the Big Five Model to predict the next word when writing a new tweet using the Markov Chain Model. Through our approach, we found that there is a big difference between predicting the next word with and without profiling the user's personality. Moreover, from the comparison, it might be said that the suggested technique was successful in identifying and detecting the accuracy of the personality prediction and suggesting the target words for each user.

In the future, we plan to apply this approach to speech recognition and to develop our real model to precisely profile the user's personality. Therefore, we improve our system power using other prediction methods such as the hidden Markov model.

7 Declarations

- **Funding.** This work supported by the authors.
- **Conflict of interest.** All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or non-financial interest in the subject matter or materials discussed in this manuscript.
- **Availability of data and materials.** Data sharing is not applicable to this article as no datasets were generated or analyzed during the current study.
- **Ethical Approval.** Not applicable.

References
References

[1] J. V. Haxby, E. A. Hoffman, and M. I. Gobbini, "The distributed human neural system for face perception," Trends Cogn. Sci., vol. 4, no. 6, pp. 223–233, 2000.

[2] J. Pednekar and S. Dubey, "Identifying Personality Trait using Social Media : A Data Mining Approach," Int. J. Curr. Trends Eng. Res., vol. 2, no. 4, pp. 489–496, 2016.

[3] M. Giannakis, R. Dubey, S. Yan, K. Spanaki and T. Papadopoulos, "Social media and sensemaking patterns in new product development: demystifying the customer sentiment", Annals of Operations Research, https://doi.org/10.1007/s10479-020-03775-6, (2020)

[4] A Go, R Bhayani, L Huang, "Twitter Sentiment Classification using Distant Supervision", CS224N project report, Stanford, 2009.

[5] A. Semmouri, M. Jourhmane and Z. Belhallaj, "Discounted Markov decision processes with fuzzy costs", Annals of Operations Research volume 295, pages 769–786, https://doi.org/10.1007/s10479-020-03783-6, (2020).

[6] A. TalwarRadu, J. JurcaBoi and F. Faltings, "Understanding user behavior in online feedback reporting", 10.1145/1250910.1250931, 2007

[7] B. Agarwal, "Personality Detection from Text : A Review," Int. J. Comput. Syst., vol. 1, no. 1, pp. 1–4, 2014.

[8] B. Verhoeven, W. Daelemans, and T. De Smedt, "Ensemble methods for personality recognition," in Proceedings of the workshop on computational personality recognition, 2013, pp. 3538.

[9] E. El Sayed, "Exploiting Social Annotations for Personalizing Retrieval," vol. 10, no. 6, pp. 192–202, 2017.

[10] E. Utamia, A. Hartantob, S. Adib, I. Oyongb and S. Raharjo, "Profiling analysis of DISC personality traits based on Twitter posts in Bahasa Indonesia", https://doi.org/10.1016/j.jksuci.2019.10.008, 2019.

[11] F. Alam, E. A. Stepanov, and G. Riccardi, "Personality traits recognition on social network-facebook," WCPR (ICWSM-13), Cambridge, MA, USA, 2013.

[12] G. Farnadi, S. Zoghbi, M.-F. Moens, and M. De Cock, "Recognising personality traits using facebook status updates," in Proceedings of the workshop on computational personality recognition (WCPR13) at the 7th

international AAAI conference on weblogs and social media (ICWSM13). AAAI, 2013.

[13] I. Sorič, Z. Penezič and I Burič. "The Big Five personality traits, goal orientations, and academic achievement" - https://doi.org/10.1016/j.lindif.2017.01.024, 2017 - Elsevier

[14] John, O. P., Donahue, E. M. and Kentle, R. L. (1991). Big Five Inventory (BFI) dDatabase record. APA PsycTests. https://doi.org/10.1037/t07550-000.

[15] M. Kosinski and D. Stillwell, "Personality and Patterns of Facebook Usage," 2012.

[16] M. T. Tomlinson, D. Hinote, and D. B. Bracewell, "Predicting conscientiousness through semantic analysis of facebook posts," Proceedings of WCPR, 2013

[17] N. Majumder, S. Poria, A. Gelbukh and E. Cambria, "Deep Learning-Based Document Modeling for Personality Detection from Text", IEEE Intelligent Systems (Volume: 32, Issue: 2, Mar.-Apr. 2017), Page: 74 - 79, doi: 10.1109/MIS.2017.23

[18] R. Talib, M. K. Hanif, S. Ayesha, and F. Fatima, "Text Mining : Techniques , Applications and Issues," Int. J. Adv. Comput. Sci. Appl., vol. 7, no. 11, pp. 414–418, 2016.

[19] S. Keh and I. Cheng, "Myers-Briggs Personality Classification and Personality-Specific Language Generation Using Pre-trained Language Models", 2019.

[20] Ziawasch Abedjan, "Data profiling", Encyclopedia of Big Data Technologies, Springer, Cham, DOI: https://doi.org/10.1007/978-3-319-77525-8, 2019

[21] Bahaa Eddine Elbaghazaoui, Amnai Mohamed & Abdellatif Semmouri. "Data Profiling over Big Data Area: A Survey of Big Data Profiling: State-of-the-Art, Use Cases and Challenges". In book: Intelligent Systems in Big Data, Semantic Web and Machine Learning. Springer. 2021.

[22] Feher, Anita, and Philip A. Vernon. "Looking beyond the Big Five: A selective review of alternatives to the Big Five model of personality." Personality and Individual Differences 169 (2021): 110002.

[23] Garg, Shruti, and Ashwani Garg. "Comparison of machine learning algorithms for content based personality resolution of tweets." Social Sciences & Humanities Open 4.1 (2021): 100178.

[24] Mitra, Arion, et al. "A Machine Learning Approach to Identify Personality Traits from Social Media." Machine Learning and Deep Learning in Efficacy Improvement of Healthcare Systems. CRC Press 31-59.

[25] Su, Yu-Sheng, and Sheng-Yi Wu. "Applying data mining techniques to explore user behaviors and watching video patterns in converged IT environments." Journal of Ambient Intelligence and Humanized Computing (2021): 1-8.

[26] Shakhovska, Khrystyna, et al. "An Approach for a Next-Word Prediction for Ukrainian Language." Wireless Communications and Mobile Computing 2021 (2021).

[27] Dhelim, Sahraoui, et al. "Big-Five, MPTI, Eysenck or HEXACO: The Ideal Personality Model for Personality-aware Recommendation Systems." arXiv preprint arXiv:2106.03060 (2021).

[28] Shang, E. R. I. C. "Introduction to markov chains and markov chain mixing." (2018).

[29] Savelev, Aleksei O., et al. "The high-level overview of social media content search engine." IOP Conference Series: Materials Science and Engineering. Vol. 1019. No. 1. IOP Publishing, 2021.

[30] Zhao, Sha, et al. "AppUsage2Vec: Modeling smartphone app usage for prediction." 2019 IEEE 35th International Conference on Data Engineering (ICDE). IEEE, 2019.

[31] Katz, Slava. "Estimation of probabilities from sparse data for the language model component of a speech recognizer." IEEE transactions on acoustics, speech, and signal processing 35.3 (1987): 400-401.

[32] Stolcke, Andreas. "Bayesian learning of probabilistic language models". Diss. University of California, Berkeley, 1994.

[33] Bengio, Yoshua, et al. "A neural probabilistic language model". journal of machine learning research, 3 (Feb): 1137-1155, 2003.

[34] Le, Hai-Son, et al. "Structured output layer neural network language model." 2011 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP). IEEE, 2011.

[35] Sundermeyer, Martin, Ralf Schlüter, and Hermann Ney. "LSTM neural networks for language modeling." Thirteenth annual conference of the international speech communication association. 2012.

[36] Mikolov, Tomas, et al. "Recurrent neural network based language model." Interspeech. Vol. 2. No. 3. 2010.

[37] Panzner, Maximilian, and Philipp Cimiano. "Comparing hidden markov models and long short term memory neural networks for learning action representations." International workshop on machine learning, optimization, and big data. Springer, Cham, 2016.

[38] Ganai, Aejaz Farooq, and Farida Khursheed. "Predicting next word using RNN and LSTM cells: Stastical language modeling." 2019 Fifth International Conference on Image Information Processing (ICIIP). IEEE, 2019.

[39] Costantini, Giulio, et al. "State of the aRt personality research: A tutorial on network analysis of personality data in R." Journal of Research in Personality 54 (2015): 13-29.