International Journal of Latest Trends in Engineering and Technology
Vol.(23)Issue(2), pp.001-005
DOI: http://dx.doi.org/10.21172/1.232.01
e-ISSN:2278-621X

PERSONALITY PREDICTION USING NATURAL LANGUAGE PROCESSING (NLP)

Himanshi Joshi[1], Himanshu Agrawal[1], Amrit Kaur[1]

Abstract - **This study explores the use of Natural Language Processing (NLP) for predicting personality traits based on text analysis. By leveraging machine learning and deep learning techniques, we aim to infer the Big Five personality traits from social media posts. Our findings indicate significant correlations between linguistic patterns and personality traits, suggesting the potential of NLP in psychological assessments. In this paper, we identified the personality from the reviews entered by the user in Hindi language about any person by identifying the personality keywords.**

Keywords –Personality, Sentiment analysis, Review, NLP, Machine learning, Hindi.

I. INTRODUCTION

1.1 Personality

Personality [1] is one of the interesting characteristics that can be considered for adaptation purposes. In the field of research, the personality of a person can be described as a set of specifications that compels a tendency on the behavior of the person; this tendency is unchanged through time and positions. Having information about one's personality period gives hints about how he would react when encountering different situations. Detecting a user's personality can facilitate knowing his potential needs on different occasions. Therefore, adaptive applications may take advantage of models of users' personality to adapt their behavior accordingly. There are a wide marketing, healthcare or recommender systems, among others. On the other hand, we believe that the personality can be traced by investigating users' interaction in Online Social Networks (OSNs). In this context, an essential consideration would be whether virtual relationship and communication reflect user personality in the real world or offline life.

Personality prediction is crucial for various applications, including targeted marketing and mental health assessment. Social media is an area where users symbolize themselves to the world. The accounts of social media are private and personal then it can describe the personal life of the users. There are a lot of bounces such as posting, commenting, and updating status can uncover personal information. Text commented or posted by users can be analyzed to procure information.

Personality is substantive as a rightly fixed feature of an individual which mentions individual's priorities. Personality has been shown to be pertinent to many types of interactions such as social relationships, predicting songs preferences, and linking between the job tasks and the personality.

[1]*Department of Computer Science Engineering, Punjabi University, Patiala, Punjab, India*

The data gathered from social media users are used to represent the user behavior based on the different situations of everyday life. Machine learning techniques can be used to understand the mood.
Sentiment Analysis also known as Opinion Mining [4] is the process of defining the emotional mode behind a series of words (blogs, user comments, etc.) using natural language processing to determine whether it is negative, positive, or neutral and to examine the problem of studying texts [2].

1.2 NLP and Sentiment Analysis in Hindi

NLP in Hindi presents unique challenges due to the language's complex morphology and syntax. Despite these challenges, recent advancements in NLP tools and techniques for Hindi have opened new avenues for research. Sentiment analysis in Hindi involves determining the emotional tone of text, which is crucial for understanding personality dimensions [3].

II. LITERATURE REVIEW

Sumner et al. [5] extended the research of personality prediction beyond the Big Five Personality traits to the anti-social traits of narcissism, Machiavellianism and psychopathy, collectively known as the Dark Triads of personality. Language use and profile attributes were analyzed of 2927 Twitter users to predict dark triads of their personalities. The authors stated that they were the first to study the relationship between Twitter use and dark triads of personality. The study was conducted using custom made Twitter application which collected self-reported ratings on the Short Dark Triad. A maximum of 3200 tweets were collected and were analyzed using LIWC which resulted into a selection of 337 features for machine prediction usage. A comparative study of total six models was conducted by the authors namely, SVM using SMO and a polynomial kernel, Random Forest, J48 algorithm, Naïve Bayes Classifier and two Kaggle models, standard benchmark model and a competition winner model which was held by the authors respectively in context to predict psychopathy and other seven traits together in two different competitions. As stated by the authors, predictive models may not work well for predicting an individual's personality but may work well for predicting the trend of anti-social traits over a subset of population. However, the study resulted into new findings of strong relationships between anti-social traits and language use. The study also showed certain limitations such as selection bias of the subjects and the ever existing issues related to linguistic usage in social media. On the other hand, this study provides endless ports of opportunities for researchers to refine personality prediction from tweets and profile attributes. Rigorous work is required in linguistics used in social media and refinements on individual level predictions is open for study to build robust models of individual's personality prediction. The study also puts forward a greater need of better evaluation metrics for prediction models.

Reitan et al. [6] discussed that proposed framework was assessed on existing corpora from different areas and on a corpus of English Twitter information commented on for nullification. The framework comprises of two sections: a negation cue detector and a negation scope classifier. The cue detector uses a lexicon lookup that yields high recall, but modest precision. The negation scope classifier creates preferable outcomes than in other area. The negation cue variation in the Twitter data was quite low. Due to part-of-speech ambiguity it was unclear for some tokens whether they functioned as a negation signal or not.

Dadvar et al. [7] studied the impact of negation detection in SA in movie reviews [7]. The issue of deciding the polarity of opinions in movie review when negative words, for example, not happen in the sentences is researched. Different negation scopes (window sizes) that affect the classification accuracy are examined to investigate how it would affect the polarity identification of the sentences. The results show that there is no significant difference in classification accuracy when different window sizes have been applied.

Yi et al. [8] has stated that sentiment Analyzer (SA) that extracts sentiment about a subject from online text documents. Instead of classifying the sentiment of an entire document about a subject, SA detects all references to the given subject, and determines sentiment in each of the references using natural language processing (NLP) techniques. Their sentiment analysis consists of 1) A topic specific feature term extraction, 2) Sentiment extraction, and 3) Subject, sentiment association by relationship analysis. SA utilizes two linguistic resources for the analysis: the

sentiment lexicon and the sentiment pattern database. The performance of the algorithms was verified on online product review articles (“digital camera” and “music” reviews), and more general documents including general web pages and news article.

Liu [9] discussed that the academics have undertaken a diverse range of related research, due to its practicality in opinion monitoring and business competitive intelligence. Sentiment analysis on online reviews has become increasingly popular. A multidisciplinary research field in nature, sentiment analysis includes multiple fields such as natural language processing (NLP), computational linguistics, information retrieval, machine learning and artificial intelligence etc. No research on the commercial value of online product reviews.

III. METHODOLOGY

Data Collection

We collected a dataset comprising social media posts, blogs, and other user-generated content. Each text sample was labeled with the author's personality traits using the Big Five Inventory (BFI).

Preprocessing

The text data underwent preprocessing steps including tokenization, stop-word removal, stemming, and lemmatization. Sentiment scores were calculated using a sentiment analysis tool to gauge the emotional tone of each text.

Feature Extraction

We extracted linguistic feature such sentiment scores. Additionally, syntactic and semantic features were considered to enhance the model's predictive power.

Model Training

Several machine learning algorithms were employed for the training of data.

IV. RESULTS

This paper explores the development of a system for predicting personality traits from user-entered feedback in the Hindi language. Leveraging Natural Language Processing (NLP) and sentiment analysis, the proposed approach aims to perform personality classification based on the Big Five personality traits framework. The system is implemented using C#.net for the frontend and SQL Server for the backend. Comparative analysis of various techniques highlights the most effective methods for personality prediction, providing valuable insights for decision-making processes.

In ‘Personality Analyse’ form, we analyze the personality of any person according to his/her DOB and reviews.

Firsly, we enter the DOB of the person whose personality we want to analyze. After clicking on the submit button, it will show the person’s friendly number, unfriendly number, planet and personality as shown in figure 1.

Secondly, we enter the reviews of the person whose personality we want to analyze. After clicking on the submit button, the results are shown (as shown in figure 2). In which red colored word shows the negative personality keywords and green coloured words shows the positive personality keywords.

From these personality keywords, we detect the personality of the person. It also shows the top two personalities of the person according to their respective score and it also shows the total score of the personality. In this example, our proposed algorithm detects two personalities i.e. सुंदर and अभिमान. In this form we can also reset our analyse.

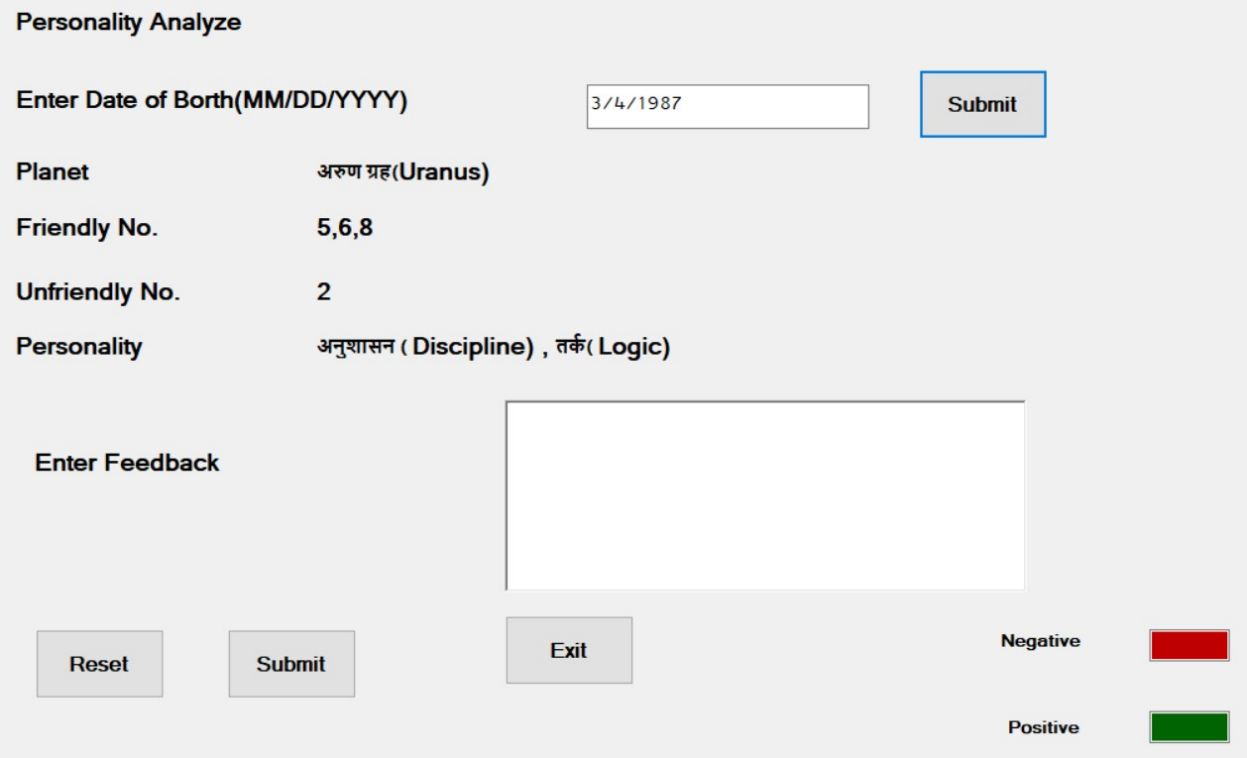

Fig. 1. Analyze Personality according to the DOB

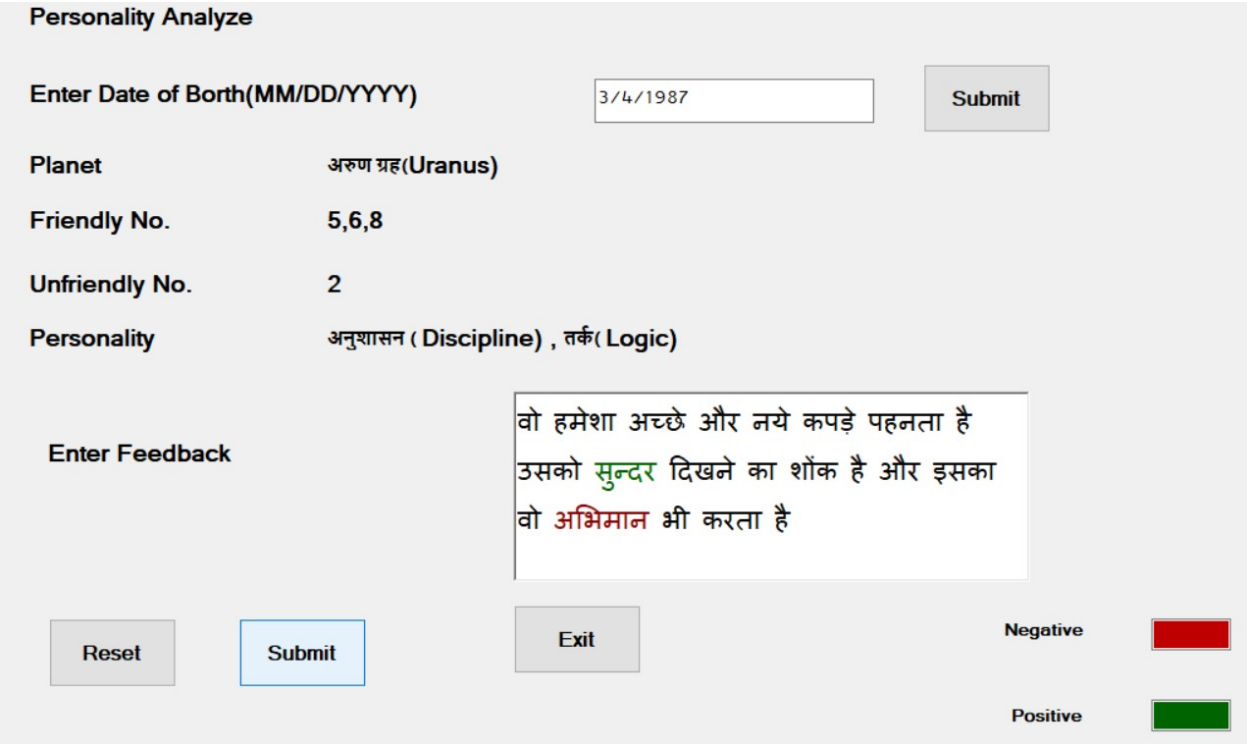

Fig. 2. Personality Analyze according to the Reviews

V.CONCLUSION

This paper demonstrates the potential of using NLP and sentiment analysis for personality prediction in the Hindi language. Future research should focus on expanding the dataset, improving NLP tools for Hindi, and exploring deep learning techniques to enhance predictive accuracy.

REFERENCES

[1] Souri, A., Hosseinpour, S. & Rahmani, A. M. (2018). Personality classification based on profiles of social networks' users and the five-factor model of personality. Human-centric Computing and Information Sciences, Vol. 8, No. 24, https://doi.org/10.1186/s13673-018-0147-4

[2] Ngatirin, N. Rahayu, Z. Zainol & T. Lee Chee Yoong (2016). A comparative study of different classifiers for automatic personality prediction. In 6th IEEE international conference on control system, computing and engineering (ICCSCE). Batu Ferringhi, Malaysia.

[3] Singh, A., & Patil, P. (2019). Natural Language Processing for Hindi: A Review. International Journal of Computer Applications, 182(22), 1-6.

[4] Garg, P., Singh, R. & Goel, M. B. (2020). A Novel Approach to Implement Natural Language Interface for Database in Punjabi Language. Journal of Xidian University, Vol. 14, No. 4, pp. 1230-1238.

[5] Sumner, C., Byers, A., Boochever, R. & Park, G. J. (2012). 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.

[6] Reitan, J., Faret J., Gamback, B. & Bungum L. (2015). Negation Scope Detection for Twitter Sentiment Analysis. Proceedings of the 6th Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis (WASSA 2015), pp. 99–108, Lisboa, Portugal, Association for Computational Linguistics.

[7] Dadvar, M., Hauff, C. & Jong, F. D. (2014). Scope of Negation Detection in Sentiment Analysis.

[8] Yi, J. & Nasukawa, T. (2003). Sentiment Analyzer: Extracting Sentiments about a Given Topic using Natural Language Processing Techniques". Proceedings of the Third IEEE International Conference on Data Mining. IEEE.

[9] Liu, M. H. (2004). Mining and Summarizing Customer Reviews. Washington. ACM.