IJNRD.ORG ISSN : 2456-4184
INTERNATIONAL JOURNAL OF NOVEL RESEARCH AND DEVELOPMENT (IJNRD) | IJNRD.ORG
An International Open Access, Peer-reviewed, Refereed Journal

Personality Analysis Using XGBoost Classifier

[1]K Indu, [2]A Akshaya Madhuri, [3]A Amrutha Varshini,[4]A Bindu Harini,[5]B Jasmini

[1]Guide, [2]student, [3]student,[4]student,[5]student
[1]Computer science and system engineering,
[1]Andhra University College Of Engineering For Women, Visakhapatnam, India

Abstract **:** The personality of a person plays a major role in his/her life. A person's personality tells what kind of person he is and his behavioral aspects. The personality of a particular person should be described and it is necessary to tell them about their personality different organization helps to perform personality analysis. Personality analysis can be done using machine learning (ML) algorithms where we can analyze the personalities of various persons.
.

IndexTerms **-: XGBoost, LogisticRegression, SVM, MBTI, Ensembling Learning, Python.**

INTRODUCTION

Personality analysis helps a person to know one's behavioral aspects it plays a vital role in our day-to-day life here, it contains data from various social media platforms like Facebook, Youtube, and Instagram, this data helps to analyze people's personalities where different machine learning algorithms are used one can make predictions of the particular person.XGBoost and Logistic Regression are two major algorithms that are used to analyze personality accuracies and the percentage of people from a particular personality.MBTI personality dataset is used to analyze the personality which categorizes individuals into one of 16 personality types based on their preferences in four different dichotomies: **extraversion (E)** vs. **introversion (I)**, **sensing (S)** vs. **intuition (N), thinking (T)** vs. **feeling (F)**, and **judging (J)** vs. **perceiving (P)**.

REVIEW OF LITERATURE.

Data sources: social media posts, written responses to personality questionnaires, and speech transcripts have been used as data sources in this MBTI datasets social media posts have shown to be a promising data source due to their large volumes and easily accessible
Machine Learning algorithms: Various Machine Learning algorithms have been used to analyze personalities such as decision trees, random forests, and neural networks. Ensemble learning methods such as stacking, bagging, and xgboost have been used to improve model performance.
Model evolution: Various metrics such as precision, recall, and F1 score have been used to evaluate the model's performance. cross-validation has been used to ensure the model's generalizability.
Feature engineering: Feature engineering is a critical step in personality Analysis projects. Text-based features such as word frequency, word embeddings, and sentiment analysis have been used as input features. Non-text-based features such as age, gender, and education level have also been used

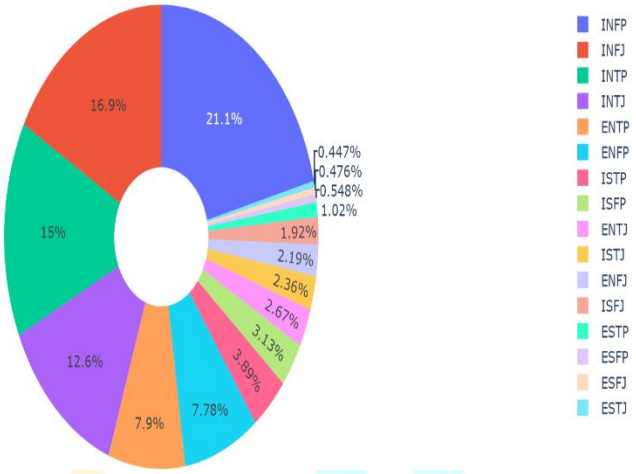
RESEARCH METHODOLOGY

Data collection: the data is collected from various social media platforms and is generated by the MBTI dataset it is a famous dataset that is used to analyze personality.
Data preprocessing: Clean and preprocess the data by removing irrelevant information, converting text to lowercase, or removing accents Convert the data into a format that can be used by machine learning algorithms, such as a numerical representation and other irrelevant symbols and other syntax errors using regular expressions.
Feature engineering: extracting meaningful features from the data, such as word frequency, sentiment analysis, or syntactic features.
Model selection: select the appropriate machine learning algorithm based on the characteristics of the data. common algorithms include decision trees, logistic regression, support vector machines, neural networks, XGBoost
Model training: train the model using a training dataset.use cross-validation techniques, such as k-fold cross-validation, to ensure the model's generalizability and make it accurate.
Model evaluation:
Calculating the model's performance using various metrics such as accuracy, precision, recall, and F1 score
relationship between interest rate and stock returns. Nguyen (2010) studies Thailand market and found thatInterest rate has aninverse relationship with stock prices.

IV. RESULTS AND DISCUSSION

The data collected gives you the percentage of different personalities present in the given MBTI dataset

Figure -1: shows the percentage of different personalities present in the given dataset.



Calculating the accuracies of personalities using two different algorithms Logistic Regression and XGBoost.

Logistic Regression Results:

	precision	recall	f1-score	support
ENFJ	0.42	0.12	0.19	41
ENFP	0.72	0.60	0.66	125
ENTJ	0.73	0.36	0.48	44
ENTP	0.71	0.58	0.64	135
ESFJ	0.00	0.00	0.00	7
ESFP	0.00	0.00	0.00	8
ESTJ	0.00	0.00	0.00	7
ESTP	1.00	0.13	0.24	15
INFJ	0.65	0.69	0.67	288
INFP	0.61	0.86	0.71	370
INTJ	0.62	0.72	0.67	193
INTP	0.69	0.82	0.75	293
ISFJ	1.00	0.27	0.42	45
ISFP	0.75	0.23	0.35	53
ISTJ	0.69	0.25	0.37	44
ISTP	0.75	0.45	0.56	67
accuracy			0.65	1735
macro avg	0.58	0.38	0.42	1735
weighted avg	0.66	0.65	0.63	1735

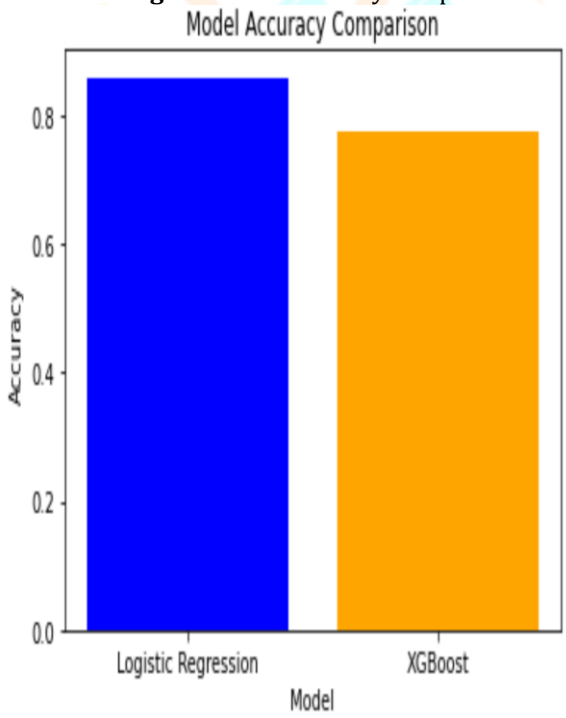
XGBoost Results:

	precision	recall	f1-score	support
ENFJ	0.60	0.22	0.32	41
ENFP	0.65	0.60	0.62	125
ENTJ	0.58	0.43	0.49	44
ENTP	0.62	0.62	0.62	135
ESFJ	1.00	0.29	0.44	7
ESFP	0.00	0.00	0.00	8
ESTJ	1.00	0.29	0.44	7
ESTP	0.89	0.53	0.67	15
INFJ	0.67	0.66	0.66	288
INFP	0.65	0.81	0.72	370
INTJ	0.65	0.73	0.69	193
INTP	0.72	0.77	0.75	293
ISFJ	0.79	0.49	0.60	45
ISFP	0.76	0.55	0.64	53
ISTJ	0.77	0.55	0.64	44
ISTP	0.71	0.63	0.67	67
accuracy			0.67	1735
macro avg	0.69	0.51	0.56	1735
weighted avg	0.67	0.67	0.67	1735

Table 1. Comparison of the algorithms for calculating accuracies

The graphical representation of the accuracies of the algorithm's logistic regression and boost.

Fig -1: model accuracy comparison



I. CONCLUSION

the personality Analysis project aims to predict a person's Myers-Briggs Type Indicator (MBTI) personality type based on their social media activity. While the project has the potential to be interesting and informative, there are some potential concerns to consider. For example, the accuracy of personality prediction based on social media activity is not well-established, and there is a risk of perpetuating stereotypes and biases if the predictions are not based on sound scientific evidence. Additionally, there are ethical considerations around the use of social media data, including issues of privacy and informed consent. It's important to ensure that any data used in the project is obtained.

REFERENCES

[1] Celli, F., Pianesi, F., Stillwell, D., & Kosinski, M. (2013). Workshop on computational personality recognition (shared task). In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Vol. 2, pp. 13-18).

[2] Golbeck, J., Robles, C., & Turner, K. (2011). Predicting personality from Twitter. In Proceedings of the 2011 Annual Conference on Human Factors in Computing Systems (pp. 2531-2534).

[3] Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., & Levy, O. (2019). Roberta: A robustly optimized BERT pretraining approach. arXiv preprint arXiv:1907.11692.
[4] Park, G., Schwartz, H. A., Eichstaedt, J. C., Kern, M. L., Kosinski, M., Stillwell, D. J., ... & Seligman, M. E. (2015). Automatic personality assessment through social media language. Journal of personality and social psychology, 108(6), 934-952.
[5] Youyou, W., Kosinski, M., & Stillwell, D. (2015). Computer-based personality judgments are more accurate than those made by humans. Proceedings of the National Academy of Sciences, 112(4), 1036-1040.