Literature Survey

Group no.2
Aswathi E
Gelsa Greenson
Neethu Shankar

Vora, Hetal & Bhamare, Mamta & Kumar, Dr. (2020). Personality Prediction from Social Media Text: An Overview. International Journal of Engineering Research and. V9. 10.17577/IJERTV9IS050203.

Advantages

1. Leverages social media data for analysis.
2. Explores machine learning and NLP methods.
3. Connects personality traits with online behavior.

Disadvantages

1. Challenges with noisy and inconsistent social media data.
2. Limited discussion on ethical concerns like privacy and consent.
3. Potential bias in data sources and demographic representation.

[10]A.S. Khan, H. Ahmad, M. Z. Asghar, F. K. Saddozai, A. Arif, and H. A. Khalid, ‘‘Personality classification from online text using machine learning approach,’’ Int. J. Adv. Comput. Sci. Appl., vol. 11, no. 3, pp. 460–476, 2020. Available: https://www.researchgate.net/publication/340399695

Advantages

1. High accuracy
2. Employs advanced text preprocessing techniques like TF-IDF to enhance results.
3. Uses a publicly available dataset, making the study reproducible.
4. Provides insights for recruitment, marketing, and customer profiling.

Disadvantages

1. Focuses only on social media data, missing broader personality aspects.
2. Results are not validated on diverse datasets or languages.
3. Does not extensively compare with modern deep learning models.
4. High accuracy may indicate overfitting, especially with resampled data.

M. N. Sahono, F. U. Sidiastahta, G. F. Shidik, A. Z. Fanani, Muljono, S. Nuraisha, and E. Lutfina, ‘‘Extrovert and introvert classification based on Myers-Briggs type Indicator(MBTI) using support vector machine (SVM),’’ in Proc. Int. Seminar Appl. Technol. Inf. Commun. (iSemantic), Sep. 2020, pp. 572–577, doi: 10.1109/iSemantic50169.2020.9234288

Advantages

1. Uses Support Vector Machine (SVM), a robust algorithm for classification.
2. Leverages the widely recognized Myers-Briggs Type Indicator (MBTI).
3. Focuses on extroversion and introversion, key personality dimensions.

Disadvantages

1. Limited to classifying only extroversion vs. introversion.
2. SVM may struggle with noisy or unbalanced datasets.
3. Binary classification oversimplifies personality traits.

[1] Alkhelil, A. H. (2016). The Relationship between Personality Traits and Career Choice: A Case Study of Secondary School Students, International Journal of Academic Research in Progressive Education and Development. 5(2). ISSN: 2226-6348.DOI: 10.6007/IJAR PED/v5- i2/2132, 2

Advantages

1.Empirical Case Study:-Uses a case study approach, offering practical, real-world data to support the analysis of personality-career relationships.

2.Wide Applicability:-The findings can be applied to career guidance programs for secondary school students, helping them make more informed decisions.

3.In-depth Analysis:-Examines the correlation between various personality traits and career preferences, providing a comprehensive view of factors influencing career decisions.

Disadvantages

1.Limited to Secondary School Students:

2.Potential Cultural Bias

3.Scope of Study:The study is limited to a single case study.

K.Yang, R. Y. K. Lau, and A. Abbasi, ‘‘Getting personal: A deep learning artifact for text-based measurement of personality,’’ Inf. Syst. Res., vol. 34, no. 1, pp. 194–222, Mar. 2023, doi: 10.1287/isre.2022.1111. Available: https://www.researchgate.net/publication/358396484

Advantages

1. Achieves 10-30% better personality detection accuracy than existing methods like BERT and IBM Insights.
2. Combines psychology-based concepts with advanced deep learning for better personality insights.
3. Useful for tasks like predicting company performance and health trends.
4. Uses transfer learning to overcome the lack of labeled personality data.

Disadvantages

1. Requires advanced knowledge and significant resources to implement.
2. Relies on high-quality data, which may not be available everywhere.
3. Less Transparent
4. Performance in other languages or contexts is not well-tested.

N. D. Almalis, G. A. Tsihrintzis, N. Karagiannis and A. D. Strati, "FoDRA — A new content-based job recommendation algorithm for job seeking and recruiting," *2015 6th International Conference on Information, Intelligence, Systems and Applications (IISA)*, Corfu, Greece, 2015, pp. 1-7, doi: 10.1109/IISA.2015.7388018.

Advantages

1. Content-based recommendation for personalized job matching.
2. Customizable for different job seekers and recruiters.
3. Improved accuracy by matching job seeker profiles with job descriptions.

Disadvantages

1. Scalability challenges with large datasets.
2. Risk of overfitting to certain keywords or attributes.
3. Potential bias if input data is biased.

M. A. Akber, T. Ferdousi, R. Ahmed, R. Asfara, and R. Rab, ‘‘Per
sonality prediction based on contextual feature embedding SBERT,’’ in
Proc. IEEE Region Symp. (TENSYMP), 2023, pp. 1–5, doi: 10.1109/TEN
SYMP55890.2023.10223609

Advantages

1.The SBERT-based paper uses Sentence-BERT, which captures contextual and semantic sentence-level features, enabling nuanced personality prediction.

2.SBERT embeddings simplify the input to downstream machine learning models, reducing the need for complex architectures like LSTM or transformers for feature extraction

3.By focusing on embeddings, this paper may have achieved faster training and inference compared to deeper architectures

Disadvantages

1.The SBERT paper focuses primarily on embeddings and does not experiment with a wide range of machine learning and deep learning models.

2.The analysis may be constrained to specific datasets, limiting generalizability.

3.While SBERT is efficient, the absence of advanced architectures like Bi-LSTM or BERT could limit its performance in capturing long-term dependencies.

I. A. Mirza, S. Mulla, R. Parekh, S. Sawant and K. M. Singh, "Generating personalized job role recommendations for the IT sector through predictive analytics and personality traits," *2015 International Conference on Technologies for Sustainable Development (ICTSD)*, Mumbai, India, 2015, pp. 1-4, doi: 10.1109/ICTSD.2015.7095894.

Advantages

1. Generates personalized job role recommendations using predictive analytics.
2. Incorporates personality traits for more tailored career suggestions.
3. Focuses specifically on the IT sector, enhancing relevance.
4. Combines data-driven insights with psychological profiling for more accurate recommendations.
5. Offers a practical solution for recruitment and career guidance in the IT industry.

Disadvantages

1. Focused only on the IT sector, limiting applicability to other industries.
2. Depends on the quality of personality trait data for accurate predictions.
3. Potential privacy concerns with collecting and analyzing personal data.
4. May not account for external factors such as industry trends or personal preferences.
5. Requires comprehensive data collection, which could be resource-intensive.

B. Fieri, J. La'la, and D. Suhartono, "Introversion-extraversion prediction using machine learning," Int. J. Inform. Vis., vol. 7, no. 4, pp. 2154–2160, 2023.
Available: https://www.researchgate.net/publication/377013314

Advantages

1. Compared multiple machine learning models to identify the best-performing ones.
2. Random Forest model achieved high accuracy (95.5%) for introversion-extraversion prediction.
3. Feature Optimization

Disadvantages

1. Focuses only on introversion and extraversion traits without exploring other personality dimensions.
2. The original dataset's imbalanced nature affected certain models' predictive power.
3. Results may not apply to other datasets or personality prediction scenarios without further validation.

Selected paper: AI Knows You: Deep Learning Model for Prediction of Extroversion Personality Trait ANAM NAZ1 , HIKMAT ULLAH KHAN , SAMI ALESAWI , OMAR IBRAHIM ABOUOLA , ALI DAUD , AND MUHAMMAD RAMZAN

This paper Utilize deep learning and natural language processing (NLP) techniques to analyze social media textual data for predicting extroversion personality traits.

Advantages:

1. High accuracy in predicting extroversion traits (up to 92.52%).
2. Comprehensive use of traditional ML and advanced DL models.
3. Applicable in various domains like career counseling, marketing, and mental health.

Disadvantages:

1. Relies heavily on textual data, which may introduce biases.
2. Requires large datasets and computational resources.
3. May not generalize well to non-English languages or different cultural contexts.

THANK YOU