ISSN 2319-5991 www.ijerst.com

Peer Reviewed, Referred & Indexed Journal

Research Paper

AI-POWERED COGNITIVE PROFILING AND HIDDEN TRAIT DISCOVERY SYSTEM FOR PERSONALITY MAPPING.

Devarayala Manisha[1], G. Narasimham[2]

[1]*Post Graduate Student, M. Tech, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India*

[2]*Associate Professor, Department of Computer Science and Engineering, Jawaharlal Nehru Technological University, Hyderabad, India.*

[1]manishadevarayala@gmail.com

[2]gnssitjntu@gmail.com

ABSTRACT

Personality assessment has historically relied on psychology-based questionnaires (MBTI) such as this one which is widely used in education, recruitment and counseling. Psychometric testing today is rapidly being adopted in India, especially in HR, EdTech, and mental wellness sectors such as so many large organizations use personality tests and these are in use as the basis for the development of the present study. In the research, we are designing an automated system for predicting MBTI personality traits from user-generated text to be based on artificial intelligence to provide scalable, real-time and unbiased personality profiles without lengthy questionnaires or manual evaluation. In existing systems XGBoost is used with text embeddings and linguistic features to classify personality traits. This method is effective for tracking the interaction of features and nonlinear relationships but is not sufficient for sequential dependencies and contextual nuances in natural language. To overcome this problem, we have proposed a Bidirectional Long Short-Term Memory (BiLSTM) model that learns contextually and sequentially from text. The system uses Sentence Transformers to generate high-dimensional semantic embeddings that capture context, meaning, and writing style. Then it feeds these embedded data into the BiLSTM network and they process the text in both forward and backward directions to understand linguistic dependencies. This deep learning method helps the model detect subtle personality signals in unstructured text. The BiLSTM-based architecture addresses the challenges of questionnaire and machine learning approaches and reduces self-report bias, enhances context understanding, and provides real-time personality prediction at scale. It is a robust and efficient tool for automated personality profiling and is suitable for real-world work with recruitment, psychological counseling, and personalized learning systems.

Keywords: MBTI, XGBoost, BiLSTM, Sentence transformer, Text Embedding, Deep learning, Manual Evaluation, Personality Prediction, Long term Dependencies.

1.INTRODUCTION

In the last few years, personalized learning has become an increasingly important aspect of education that provides learning that is tailored to individual needs, preferences and cognitive characteristics. Traditional adaptive learning systems are concerned with performance data, which often can miss personality-related preferences that affect learning behavior. So, the need for intelligent systems that can better understand learners and offer personalized learning experiences is growing. Myers-Briggs Type Indicator (MBTI) is a standard personality assessment tool that categorizes people into sixteen personality types based on four dimensions of Extraversion/Introversion, Sensing/Intuition, Thinking/Feeling and Judging/Perceiving. MBTI can help to understand how people process information and how they make decisions and interact with their environment. Integrating personality traits into learning systems allows for better personalization and better learner engagement and outcomes. With the rapid growth of digital communication, a lot of user-generated content from social media, forums, and online platforms has become an excellent source for personality analysis. Recently, AI, Natural Language Processing (NLP) and Deep Learning have made it possible to automatically detect personality traits from text. Normal machine learning, such as XGBoost, has good performance, but has limitations in capturing the context and the sequential relationships of text. To overcome these limitations, this project aims to create a cognitive profile and hidden trait discovery system using a Bidirectional Long Short-Term Memory (BiLSTM) model. The system uses Sentence Transformers to generate semantic embeddings and BiLSTM to capture contextual information from text in both forward and backward directions. This approach improves personality prediction accuracy and allows for the identification of hidden personality traits through personality mapping. The proposed system can be applied in personalized learning, recruitment, career guidance, counseling, and self-development and is an intelligent and scalable approach for automated personality assessment.

II. LITERATURE SURVEY

[1] Bousalem, S., Benchikha, F., and Chelghoum, M., "Ontology-Based Learner Profiling Using Machine Learning Techniques for Personalized Learning Systems", International Journal of Intelligent Information Technologies, 2024. The purpose of this research is to address the shortcomings of traditional

learner modeling methods that cannot be used to capture the dynamic learner preferences, behaviors, and characteristics of learners. Based on the knowledge of the learner profiles, we develop a semantic ontology framework for learning models and we train machine learning models to train them for learner classification and personalization. These ontologies are semantically and machine learning models can be used to model learner skills, learning styles, and preferences, while machine learning algorithms can be used to identify patterns and predict learner requirements. The results of this work show that the integration of ontologies with machine learning to learner profiling is more flexible and accurate compared to individual methods alone; more personalized. The study makes clear that semantic representation with predictive analytics is the best solution for adaptive and intelligent learning environments.

[2] A. Muñoz, J. Lasheras, A. Capel, M. Cantabella, and A. Caballero, "OntoSakai: On the optimization of a learning management system using semantics and user profiling," Exp. Syst. Appl., vol. 42, nos. 15–16, pp. 5995–6007, Sep. 2015. Muñoz and colleagues aim to enhance Learning Management Systems (LMS) using semantic technologies and ontology-based user profiling. They find that the current LMS platforms do not offer personalization and do not adapt to individual learner needs and preferences. To address this problem, they propose the semantic-based model, called "OntoSakai," for the Sakai LMS platform. The ontologies are used to describe learner behavior, interactions, and preferences, and semantic reasoning techniques are used to make recommendations and adaptive learning resources. On an experimental basis, the proposed system improves learner engagement, content delivery, personalization accuracy, and learning outcomes. From this research, we conclude that the ontology-based semantic technology can transform traditional LMS platforms into intelligent and adaptive learning systems that offer personalized educational experiences. [3] K. Stancin, P. Poscic, and D. Jaksic, "Ontologies in education— State of the art," Educ. Inf. Technol., vol. 25, no. 6, pp. 5301–5320, Nov. 2020

The research of K. Stancin, P. Poscic and D. Jaksic shows how ontology can be used in education to improve structured knowledge representation and intelligent learning systems.The study identifies that the traditional education systems do not have common ways of representing educational data and learner profiles as well as to make them interoperable and adaptive but rather, have no standardized techniques. The research reviews various ontology-based educational systems in literature reviews. They find that ontologies enhance semantic interoperability, knowledge representation and personalized learning and conclude that ontology-based approaches can greatly enhance intelligent educational environments and adaptive learning systems. [4] P. T. Palomino, A. M. Toda, L. Rodrigues, W. Oliveira, L. Nacke, and S. Isotani, "Anontology for modelling user' profiles and activities in gamified education," Res. Pract. Technol. Enhanced Learn., vol. 18, p. 18, Nov. 2022. P. T. Palomino and co-authors aim to improve user profiling and activity modeling in gamified educational systems using ontology-based approaches. They show that traditional profiling approaches are unable to capture dynamic user interactions and behaviors in gamified learning environments. The goal is to develop an ontology for user profiles, achievements, activities, and engagement metrics in a gamified educational system.

III. METHODOLOGY

A. System Overview

The system architecture of our proposed personality prediction system is designed as a modular and scalable pipeline that processes user-generated text and predicts personality traits based on Myers-Briggs Type Indicator. The architecture begins with the data input layer, where user text or dataset posts are taken from files or real-time input. This input is fed into the preprocessing layer, which performs text cleaning of the text and removes noise, stopwords, URLs and special characters, followed by normalization. The resulting text is converted into numerical representations through feature extraction techniques. The Sentence Transformers have been proposed in the previous work, where the text is embedded in high-dimensional semantic structures and a few more linguistic features like text length, average word length, and sentiment polarity. In the proposed system, we also include a tokenization and sequence padding layer to prepare text data for deep learning models. We implement two modeling layers: our existing XGBoost model and our proposed Bidirectional Long Short-Term Memory (BiLSTM) model. The XGBoost model handles structured feature vectors and learns nonlinear relationships while the BiLSTM model captures sequential and contextual dependencies in text by processing input from both forward and backward directions. Each MBTI trait pair (Introversion/Extroversion, Sensing/Intuition, Thinking/Feeling, Judging/Perceiving) is modeled separately with our model storage layer and tokenizers and encoders to perform multi-label classification effectively. The prediction layer takes user input and applies the same preprocessing and feature transformation steps and feeds it into the trained models to generate predictions for each personality trait. They are combined to form the final MBTI personality type. The output layer displays the predicted personality in a user-friendly format, which can be integrated into applications such as recruitment systems, counseling tools or personalized learning platforms. The architecture is flexible and scalable to handle batch processing as well as real-time prediction, and is efficient, accurate and easy to integrate with external systems.

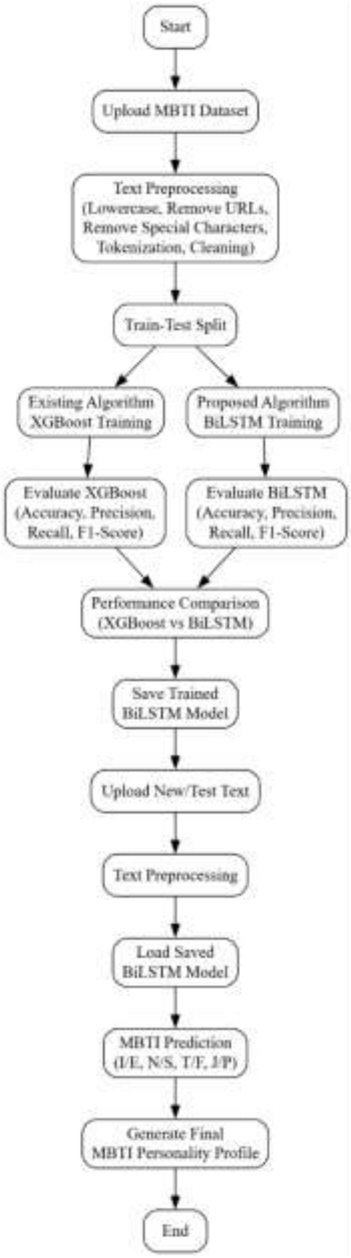

Figure 1: System Architecture for AI based Personality Prediction from Text.

B. Data Acquisition and Preprocessing

The data acquisition stage involves collecting the textual dataset required for training and evaluating the personality prediction model. The dataset contains user-generated text along with corresponding Myers–Briggs Type Indicator (MBTI) personality labels. The textual data may include social media posts, comments, or other forms of written communication that reflect an individual's language usage and behavioral patterns. The dataset is stored in a structured CSV format, where each record consists of the user's text and its associated MBTI personality type. The data preprocessing stage transforms raw textual data into a clean and standardized format suitable for deep learning. Initially, all text is converted to lowercase to maintain consistency across the dataset. Unnecessary elements such as URLs, punctuation marks, special characters, extra spaces, and irrelevant symbols are removed to reduce noise. Common stopwords that carry little semantic value are also eliminated, allowing the model to focus on more informative words. The cleaned text is then tokenized into individual words or subword units, and each token is mapped to an integer index. These token sequences are padded or truncated to a fixed length so that all input samples have a uniform size. Next, the **Embedding Layer** transforms these token indices into dense numerical vectors that capture semantic and contextual relationships between words.

C Proposed System

The working of the proposed system begins when a user provides input text. The input text is automatically processed by the system without any manual questionnaire. Input text is first cleaned and standardized so that the data is identical. The Sentence Transformer then encodes this text into a numerical vector representation that captures deep semantic and contextual information. These embeddings are meaningful features that represent the user's language patterns and cognitive expressions. The BiLSTM model then takes these embeddings as input and analyzes them in a sequence by learning the relationships between words and phrases in the whole text. Because BiLSTM processes data from both sides, it has a complete picture of the context so it can detect subtle personality signals such as emotional tone, logical reasoning, or abstract thinking. The model outputs predictions for each personality trait dimension and the predictions are then decoded and combined to generate the final MBTI personality type. The entire process is automated and can produce results in real time, which makes it efficient and scalable for large-scale applications.

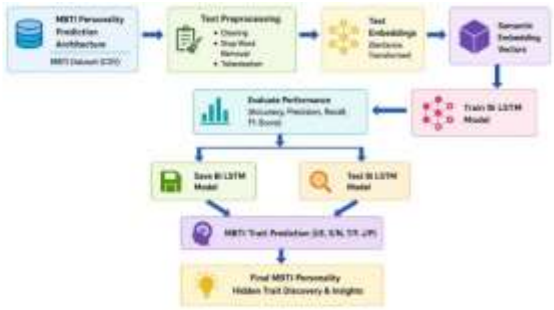

Fig 2: Architecture of BiLSTM Model

D Model Development

The model development stage focuses on building and training the deep learning model for MBTI personality prediction. After preprocessing, the tokenized text is converted into dense vector representations using an Embedding Layer, which captures the semantic meaning of words. These embeddings are then passed to a Bidirectional Long Short-Term Memory (BiLSTM) network, enabling the model to learn contextual information from both forward and backward directions.

E Performance Evaluation

The proposed system is evaluated using standard classification metrics such as Accuracy, Precision, Recall, and F1-Score. The dataset is divided into training and testing sets to assess model performance on unseen data.

V. RESULTS AND DISCUSSION

The results of the proposed personality prediction system show that AI-based approaches can effectively identify personality traits from user-written text based on the Myers–Briggs Type Indicator. The existing model using XGBoost provides good performance by utilizing semantic embeddings and linguistic features, but it has limitations in understanding the sequence and context of words. In contrast, the proposed Bidirectional Long Short-Term Memory (BiLSTM) model achieves better accuracy because it can capture contextual relationships and word dependencies more effectively. The experimental results indicate that the BiLSTM model outperforms the existing approach in most trait predictions, providing more reliable and consistent outputs. This improvement demonstrates that deep learning techniques are more suitable for analyzing natural language in personality prediction tasks. The proposed personality prediction model is evaluated using multiple performance metrics to ensure its effectiveness, reliability, and generalization capability. Accuracy measures the overall percentage of correctly predicted personality traits, while Precision evaluates the proportion of correctly predicted personality classes among all predicted instances. Recall measures the model's ability to correctly identify the actual personality traits present in the dataset, and the F1-Score provides a balanced evaluation by combining both precision and recall into a single metric. In addition, a Confusion Matrix is used to visualize the distribution of correct and incorrect predictions across different MBTI personality dimensions, enabling a detailed analysis of the model's classification performance and identifying areas for improvement.

MBTI Personality Dimension	Accuracy (%)	Precision	Recall	F1-Score
I/E	96.33	0.96	0.96	0.96
S/N	96.33	0.98	0.98	0.98
T/F	96.00	0.96	0.96	0.96
J/P	95.00	0.95	0.95	0.95
Average	96.15	0.96	0.96	0.96

Table 1: Performance Summary of the Proposed BiLSTM Model.

I/E-Introvert/Extrovert

S/N- Sensing/Intuition

T/F-Thinking/Feeling

J/P- Judging/Perceiving

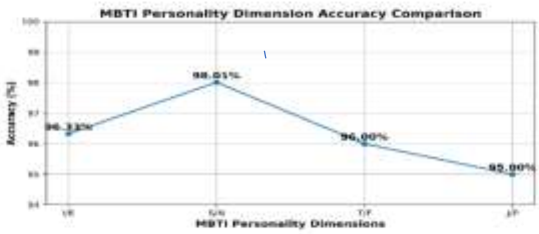

Fig 3: Accuracy of the proposed BiLSTM model for each MBTI personality dimension.

Accuracy Values:

I/E = 96.33%

S/N = 98.01%

T/F = 96.00%

J/P = 95.00%

The accuracy graph presents the performance of the proposed model across the four MBTI personality dimensions. Among these, the S/N (Sensing–Intuition) category recorded the highest classification accuracy of 98.01%, whereas the J/P (Judging–Perceiving) category achieved the lowest accuracy of 95.00%. The remaining dimensions, I/E and T/F, also demonstrated strong predictive performance. With an overall average accuracy of 96.15%, the results confirm the effectiveness and reliability of the proposed personality prediction model.

Figure 4: User Input Interface of the AI MBTI Personality Predictor System.

Figure 4 illustrates the user input interface of the AI MBTI Personality Predictor System, which provides a simple and user-friendly platform for personality analysis. The interface contains a large text input area where users can enter their thoughts, opinions, or any written content for analysis. Once the text is entered, the user clicks the **Predict** button to initiate the prediction process. The system then preprocesses the input text, converts it into numerical representations using the embedding layer, and passes it to the trained **BiLSTM** model for personality classification.

Figure 5: Output of the Proposed AI-Powered Cognitive Profiling and Hidden Trait Discovery System

Figure 5 illustrates the prediction output generated by the AI MBTI Personality Predictor System after processing the user's input text. Once the user submits the text by clicking the **Predict** button, the system preprocesses the input, converts it into semantic representations, and analyzes it using the trained **BiLSTM** model.

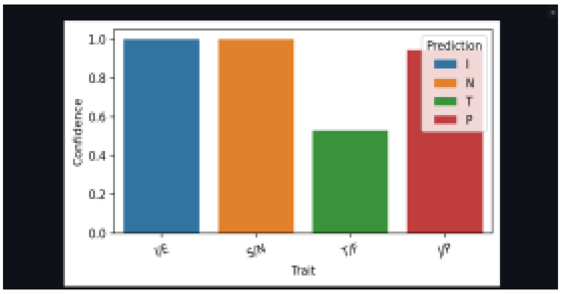

Figure 6: Trait Confidence Score Visualization

Figure 6 presents the confidence score visualization for each of the four MBTI personality dimensions predicted by the proposed AI-based personality prediction system. The bar chart displays the confidence values for the Introversion/Extroversion (I/E), Sensing/Intuition (S/N), Thinking/Feeling (T/F), and Judging/Perceiving (J/P) traits, providing a detailed view of the model's prediction certainty for each dimension. Higher confidence values indicate stronger confidence in the predicted trait, while comparatively lower values suggest greater uncertainty. In the illustrated example, the model exhibits very high confidence for the I/E and S/N dimensions, moderate confidence for the J/P dimension, and comparatively lower confidence for the T/F dimension.

Figure 7 : Personalized Insights Generated through Cognitive Profiling and Personality Mapping.

Figure 7 illustrates the personalized insights generated by the proposed AI-powered cognitive profiling and personality mapping system after predicting the user's MBTI personality type. Based on the personality classification, the system provides a comprehensive interpretation that includes the personality title, key strengths, suggested career paths, areas for improvement, lifestyle recommendations, and the meaning of each MBTI personality dimension. In the illustrated example, the user is identified as **INTP**, and the system highlights strengths such as analytical thinking, curiosity, and innovative problem-solving abilities.

V. CONCLUSION

AI-powered Cognitive Profiling and Hidden Trait Discovery System demonstrates the use of Artificial Intelligence (AI) and Natural Language Processing and Deep Learning methods to predict human personality with user-generated text. By employing Sentence Transformers in combination with BiLSTM based Sentence Transformer and BiLSTM model, the system can detect semantic meaning, context, and language patterns, which is more accurate than traditional machine learning methods. The proposed system is scalable, efficient, and reliable for personality mapping and hidden trait discovery which is useful for recruitment, personalized learning, career advice, counseling and psychological analysis. Future work can be applied to enhance our system with more advanced transformer-based models like BERT, RoBERTa, and GPT to improve prediction performance. Other improvements may include multilingual personality prediction, multimodal analysis with text, voice and facial expressions, and explainable AI techniques for more transparent and interpretable results. The system can also be integrated with chatbots, adaptive learning platforms, and mental health support applications to offer personalized recommendations and real-time personality insight, which makes it more intelligent and complete profiling solution in practice.

References

[1] S. Bousalem, F. Benchikha, and M. Chelghoum, ‘‘Modeling learner profiles using ontologies and machine learning,’’ in Proc. 2nd Int. Conf. New Technol. Inf. Commun. (NTIC), Dec. 2022, pp. 1–6.

[2] A. Muñoz, J. Lasheras, A. Capel, M.

Cantabella, and A. Caballero, ‘‘OntoSakai: On the optimization of a learning management system using semantics and user profiling,’’ Exp. Syst. Appl., vol. 42, nos. 15–16, pp. 5995–6007, Sep. 2015.

[3] K. Stancin, P. Poscic, and D. Jaksic, ‘‘Ontologies in education— State of the art,’’ Educ. Inf. Technol., vol. 25, no. 6, pp. 5301–5320, Nov. 2020.

[4] P. T. Palomino, A. M. Toda, L. Rodrigues, W. Oliveira, L. Nacke, and S. Isotani, ‘‘An ontology for modelling user’ profiles and activities in gamified education,’’ Res. Pract. Technol. Enhanced Learn., vol. 18, p. 18, Nov. 2022.

[5]M. Kordahi, “Ontology for the user-learner profile personalizes the search analysis of online learning resources: The case of thematic digital universities,” Inf. Technol. Libraries, vol. 41, no. 2, pp. 1–21, 2022. [Online]. Available: https://doi.org/10.6017/ital.v41i2.13601