ABSTRACT

Mismatch between students' personalities and their chosen college majors is a critical issue that impacts learning motivation and increases the risk of dropping out. This study proposes a machine learning-based solution in the digital era, particularly aimed at high school students. A lack of self-awareness, external pressure from parents or career trends, and excessive gadget use have made it difficult for many students to recognize their potential and choose the right major. Research indicates that such mismatches can lead to academic dissatisfaction, decreased motivation, and even college dropout. Therefore, a technology-based solution is needed to assist students in making more informed decisions.

This study proposes a college major recommendation system based on machine learning, utilizing personality analysis through the Myers-Briggs Type Indicator (MBTI). The methods employed include text preprocessing (tokenization, stopword removal, lemmatization), feature extraction (word frequency, sentiment, topics), and modeling using boosting algorithms (XGBoost, AdaBoost, Gradient Boosting, Support Vector Machine (SVM), Naive Bayes, and Logistic Regression). The system is integrated into a Flask-based website to provide user-friendly access to major recommendations aligned with personality profiles. This approach combines analysis of communication patterns and user preferences on social media to deliver more personalized recommendations.

The experimental results show that the accuracy of the models in predicting MBTI personality types ranged between 60–80%, with major challenges including noisy data and limited labeled datasets. Among the six algorithms tested, only Support Vector Machine (84%) and Logistic Regression (83%) surpassed the minimum accuracy target of 80%, while others such as XGBoost, Gradient Boosting, AdaBoost, and Complement Naive Bayes achieved results within the 60–72% range. Based on evaluation metrics, the SVM model achieved an average F1-score of 0.83 with precision of 0.84 and recall of 0.84, while Logistic Regression reached an F1-score of 0.82 with precision of 0.83 and recall of 0.83. These findings suggest that the proposed machine learning and MBTI-based system has the potential to reduce the risk of mismatched major selection and support the transformation of technology-driven education in Indonesia.

***Keywords**: Personality Analysis, MBTI, College Major Recommendation, Algorithmic Models, Flask.*