

Abstract

Modern interactions do not require face-to-face meetings, as social media has become a platform for indirect communication. X, as one of the social media platforms with 27.5 million users, serves as a space for text-based social interactions, providing a vast amount of data that can be collected and utilized in this study to compare the performance of Logistic Regression and Naive Bayes in predicting personality traits based on social media data from X.

In this study, experiments were conducted under three scenarios: without interaction data, with interaction data, and with data augmentation. Logistic Regression performed better in capturing patterns in text data, achieving the highest accuracy of 52.63%, while Naive Bayes was more stable when dealing with imbalanced data, with a highest accuracy of 51.58%. The addition of interaction data did not provide significant improvements due to a lack of variation in distribution. However, data augmentation slightly increased accuracy, particularly for Naive Bayes. The key factors affecting model performance were feature extraction methods, data quality and variation, and data imbalance. Therefore, based on this study, selecting the right features and optimizing data processing strategies play a crucial role in improving personality prediction accuracy.

Keywords: X, Interaction, Logistic Regression, Naive Bayes.