

Predicting Personality from Tweets using Machine Learning

Nathan Holmes

NHOLMES@EMAIL.UNC.EDU

University of North Carolina at Chapel Hill

Namita Krishna

EMAIL HERE

University of North Carolina at Chapel Hill

Andrew Mu

EMAIL HERE

University of North Carolina at Chapel Hill

Nabeel Rahman

EMAIL HERE

University of North Carolina at Chapel Hill

Abstract

This paper explores the use of machine learning in analyzing social media data, specifically Tweets, to accurately identify Myers-Briggs Type Indicator (MBTI) personality types. Leveraging diverse data encompassing both qualitative and quantitative dimensions, our study employs K-Nearest Neighbors, Logistic Regression, Multinomial Naive Bayes, and Random Forest models to discern nuanced patterns. Through rigorous evaluation, we contribute insights into the effectiveness of these models for personality type identification. This research advances the understanding of the interplay between language use on social media and personality traits, with practical implications for social sciences and human behavior analysis.

Keywords: MBTI, KNN, Logistic Regression, Multinomial Naive Bayes, Random Forest

1. Introduction

Classification is a fundamental domain in realm of machine learning research, permeating our daily lives through advanced models that process data with refined sophistication. From discerning spam emails to categorizing patient diagnoses, we are constantly immersed in the practical applications of these advanced technologies. This project delves into the realm of social media, leveraging a diverse dataset of Tweets that encapsulate both qualitative and quantitative dimensions. Within this paper, we undertake the construction and evaluation of accuracy in identifying MBTI types based on users' tweet content. Our approach involves employing a K-Next Neighbor classifier model, a Logistic Regression model, a Multinomial Naive Bayes model, and a Random Forest classifier model to discern nuanced patterns and trends.

2. Data Collection and Model Development

Our dataset, obtained from Kaggle, consists of a collection of Tweets paired with their corresponding user’s MBTI types. This provided us with diverse and invaluable data for identifying patterns between tweet content and a user’s personality. After downloading the dataset, we employed preprocessing techniques, including tokenization and vectorization using tools like CountVectorizer. Subsequently, we split the dataset into training and testing sets, facilitating the development of various classification models. The K-Nearest Neighbors classifier, Logistic Regression, Multinomial Naive Bayes, and Random Forest classifier were trained and evaluated to capture nuanced patterns in the data, and each underwent fine-tuning of hyperparameters. Accuracy metrics and confusion matrices were then generated, providing a comprehensive assessment of each model’s performance in predicting personality types from social media content.

3. Evaluating Model Results

To assess the performance of our classification models, we employ key metrics including accuracy, precision, recall, and F1-score.

The accuracy of a model is determined by the ratio of correctly predicted instances to the total number of instances:

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Additionally, we generate confusion matrices, visual representations of a model’s performance, to provide insights into the number of true positive, true negative, false positive, and false negative predictions for each class. This allows us to see which MBTI types we are identifying correctly, and which ones we are confusing. We also take this a step further, and with each model calculate the number of incorrectly predicted MBTI types were only 1 trait, or letter, away from being correct.

3.1 Multinomial Naive Bayes

The Multinomial Naive Bayes (MNB) model, well-suited for text classification tasks, emerged as a fitting choice for our Twitter dataset. Rooted in Bayes’ theorem, the probability estimation equation for MNB forms the core of our implementation:

$$P(\mathbf{x}|y, \theta) = \prod_{i=1}^n \frac{N_{yi} + \alpha}{N_y + \alpha \cdot N_{total}}$$

Here, the probability of a given tweet belonging to a personality class given model parameters, $P(\mathbf{x}|y, \theta)$, is equal to the product ($\prod_{i=1}^n$) of all possible probability estimates of a word being in tweets by a specific MBTI type ($N_{yi} + \alpha$) out of the total count of all words in tweets of that personality type ($N_y + \alpha \cdot N_{total}$).

We utilized CountVectorizer to convert text data into feature vectors, focusing on term frequencies in each document. To optimize the model’s performance, hyperparameter fine-tuning was implemented, specifically exploring different values for the Laplace smoothing parameter α . Laplace smoothing, crucial for handling the zero probability problem that

occurs when the model encounters words it was not trained on, was determined to be most effective with α set at 0.1.

In our evaluation, the MNB model demonstrated an accuracy of 0.48 in predicting Myers-Briggs personality types based on tweet content. Complementing accuracy, a confusion matrix visually represented the model's performance across various personality types, offering a nuanced understanding of the model's strengths and potential areas for improvement. For example, this matrix revealed that 50% of all mistakes were confusions of MBTI types that were only a single trait apart.