Personality Prediction from Social Media Posts

PROJECT SYNOPSIS

Department of CSE (DS)

Submitted by

Shreya Mishra

Sushant Mishra

Guided by

Ms. Dhanshri Parihar

(Assistant Professor, AKGEC)

Ajay Kumar Garg Engineering College

Academic Year: 2024-25

Contents

Chapter 1

Introduction

The project involves developing a system to analyze social media posts to predict the personality traits of users. This is based on the Big Five Personality Traits model, using natural language processing (NLP) techniques. The system leverages Python programming, with libraries like NLTK and Scikit-learn, for text analysis and machine learning. The project can assist in areas such as targeted marketing, behavioral analysis, and HR candidate profiling.

In essence, this project taps into the growing field of psychoinformatics and has numerous practical applications in industries ranging from HR to entertainment, mental health, and marketing.

Chapter 2

Literature Survey

2.1 Personality Prediction Models

Authors: Kosinski M., Stillwell D., & Graepel T. (2015)
Title: Computer-based Personality Judgments Are More Accurate Than Those Made by Humans
Summary: This study demonstrated how digital footprints on social media, such as Facebook Likes, can be used to predict Big Five personality traits. The results showed that computational models based on user activity were highly accurate, surpassing human judgments in personality prediction.
Relevance: This work emphasizes the potential of computational models for predicting personality traits from digital behavior, validating the feasibility of the proposed project.
DOI: 10.1073/pnas.1218772110

2.2 Natural Language Processing for Personality Prediction

Authors: Vora Hetal & Bhamare Mamta & Dr. Kumar (2020)
Title: Personality Prediction from Social Media Text: An Overview
Summary: This research focused on analyzing Twitter data to predict Big Five traits using text mining and machine learning. By combining content and stylistic features, the study achieved significant accuracy in personality classification, particularly for Openness and Extraversion.
Relevance: This study reinforces the use of NLP techniques such as TF-IDF, along with stylistic features, for personality prediction from social media text.
DOI: 10.17577/IJERTV9IS050203

2.3 Big Five Personality Traits in Text Analysis

Authors: Settanni M., & Marengo D. (2015)
Title: Sharing Feelings Online: Studying Emotional Well-being via Automatically Ana-

lyzed Facebook Status Updates
Summary: This study analyzed Facebook status updates to assess personality traits and emotional well-being. It utilized linguistic and emotional features for accurate prediction, focusing on traits like Neuroticism and Agreeableness.
Relevance: This research validates the importance of analyzing emotional and linguistic features in text for personality prediction, directly supporting the project's approach.
DOI: 10.3389/fpsyg.2015.01045

2.4 Machine Learning Approaches for Text-Based Personality Prediction

Authors: Tandera Tommy & Hendro & Suhartono Derwin & Wongso Rini & Prasetio Yen. (2017).
Title: Personality Prediction System from Facebook Users
Summary: This study explored personality prediction from Facebook data using machine learning algorithms, including Random Forest and Support Vector Machines. Features such as n-grams, sentiment scores, and part-of-speech tags were used to classify personality traits effectively.
Relevance: This work highlights the effectiveness of Random Forest and Support Vector Machines in personality prediction, suggesting robust algorithms for your project.
DOI: 10.1016/j.procs.2017.10.016

2.5 Feature Extraction and NLP Methods

Authors: A. A. Marouf, M. K. Hasan and H. Mahmud
Title: Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media
Summary: This study employed various NLP techniques, including TF-IDF and word embeddings, to predict personality traits from Twitter text. It demonstrated that combining lexical features with embeddings improved accuracy.
Relevance: This research validates the use of TF-IDF as a baseline and suggests exploring embeddings to enhance prediction models.
DOI: 10.1109/TCSS.2020.2966910

2.6 Summary

The reviewed literature highlights the effectiveness of using social media text for personality prediction based on the Big Five model. Studies since 2015 emphasize the combination of NLP methods like TF-IDF and embeddings with machine learning models such as Random Forest and Support Vector Machines. These techniques form a solid foundation for building robust personality prediction systems.

Chapter 3

System Requirements

Hardware Requirements	Software Requirements
Processor: Intel i3 or higher	Programming Language: Python
RAM: 4 GB or higher	Libraries: NLTK, Scikit-learn
Storage: 500 GB or more	Tools: Jupyter Notebook, Colab

Table 3.1: Hardware and Software Requirements

Chapter 4

Methodology

1. **Data Collection:** Social media posts gathered from publicly available datasets.
2. **Preprocessing:** Text cleaning, tokenization, and lemmatization.
3. **Feature Extraction:** Techniques include TF-IDF, sentiment analysis, and part-of-speech tagging.
4. **Model Training:** Supervised learning with Random Forest, SVM.
5. **Evaluation:** Metrics like accuracy, precision, recall, F1-score.

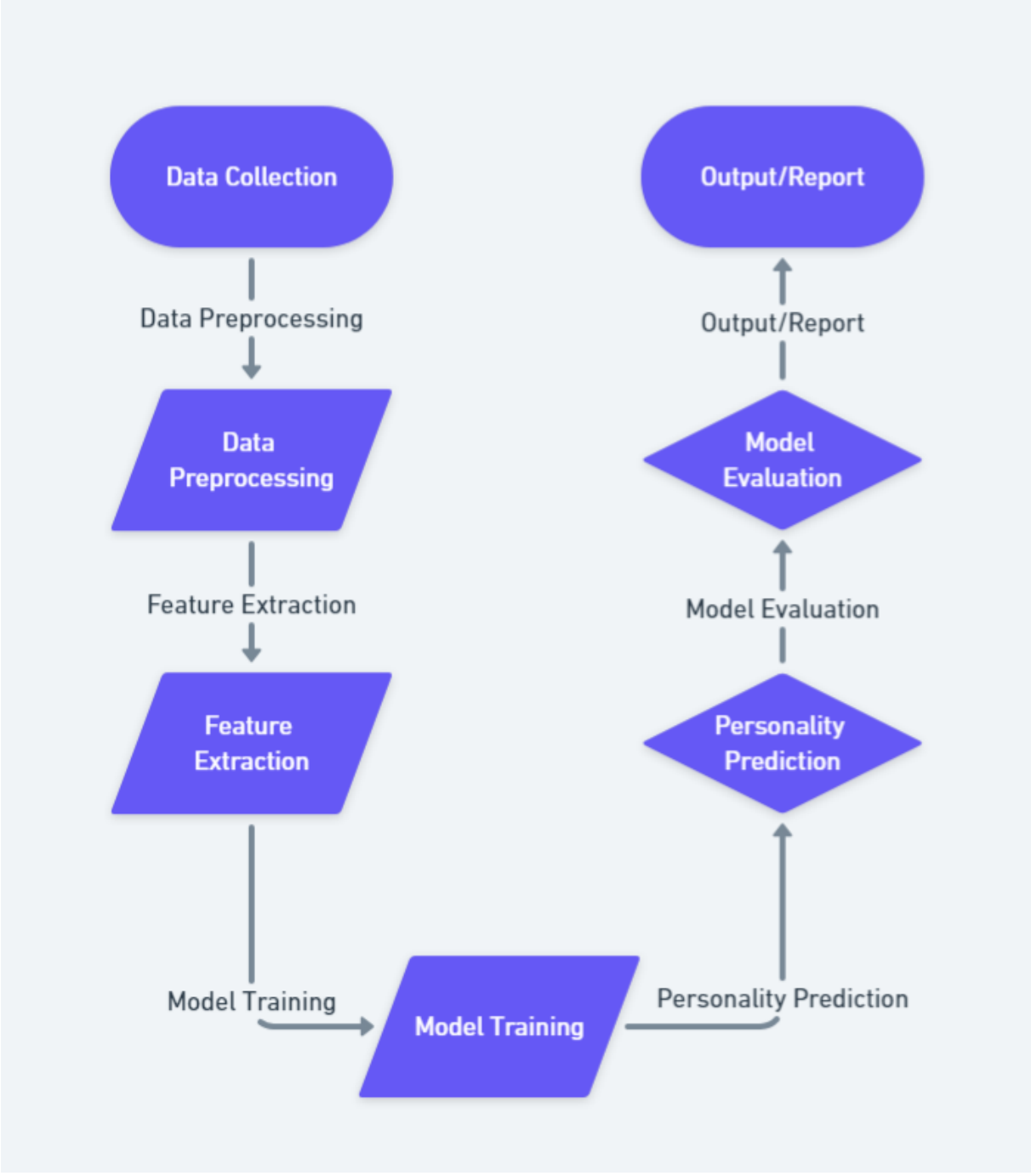

Figure 4.1: Flowchart of the proposed methodology.

Chapter 5

References

- nltk.org
- skikit-learn.org
- kaggle.com
- github.com
- google scholar

Bibliography

[1] Kosinski M, Stillwell D, Graepel T. Private traits and attributes are predictable from digital records of human behavior. Proc Natl Acad Sci U S A. 2013;110(15):5802-5805. doi:10.1073/pnas.1218772110

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

[3] Settanni Michele & Marengo Davide (2015). Sharing feelings online: studying emotional well-being via automated text analysis of Facebook posts. Frontiers in Psychology. 6. 1045. 10.3389/fpsyg.2015.01045.

[4] Tandera Tommy & Hendro & Suhartono Derwin & Wongso Rini & Prasetio Yen (2017). Personality Prediction System from Facebook Users. Procedia Computer Science. 116. 604-611. 10.1016/j.procs.2017.10.016.

[5] A. A. Marouf, M. K. Hasan and H. Mahmud, "Comparative Analysis of Feature Selection Algorithms for Computational Personality Prediction From Social Media," in IEEE Transactions on Computational Social Systems, vol. 7, no. 3, pp. 587-599, June 2020, doi: 10.1109/TCSS.2020.2966910