Analysis of the relationship between social media and mental health

Alice Sturaro, *DTM student, University of Bologna*

Abstract—This project investigates the potential of using social media usage patterns and demographic data to predict the risk of depression in individuals. Following the CRISP-DM methodology, a survey-based dataset (*smmh.csv*) was analyzed, containing self-reported information on age, gender, relationship status, social media habits, and mental health indicators.

Data preprocessing included handling missing values, encoding categorical features, and selecting the most relevant variables using a Random Forest-based feature selection technique. Three models were trained and evaluated: Random Forest, Decision Tree, and Logistic Regression (as a baseline). The models were compared using multiple metrics, including accuracy, precision, recall, F1-score, and ROC AUC.

The results showed that the Random Forest and Decision Tree models achieved comparable performance, with the Decision Tree offering greater interpretability. The study highlights the importance of feature selection and model evaluation in the context of mental health prediction. Future improvements could include handling class imbalance, exploring more advanced models, and validating results on external datasets.

Index Terms—Social media, mental health, Machine learning, DTM

1 Introduction

The rise of social media has significantly impacted modern society, influencing communication patterns, behavior, and even mental health. In this project, we aim to explore the potential of using social media data and demographic information to predict the risk of depression in individuals. The dataset used for this analysis is "smmh.csv", which contains self-reported responses on social media usage habits, mental health indicators, and demographic characteristics such as age, gender, and occupation.

This project was conducted as an individual assignment, ranging from data exploration and cleaning to model development and evaluation. The work follows the CRISP-DM methodology (Cross-Industry Standard Process for Data Mining), a widely accepted framework in data science projects. The main objectives of this study are:

- To investigate whether social media usage patterns can predict the risk of depression.
- To identify the most relevant features contributing to the prediction.
- To compare the performance of different machine learning models on the task.

The work involved several key phases:

1) Data Understanding: exploring the dataset, identifying potential issues, and preparing it for analysis.
2) Data Preparation: cleaning, encoding, feature selection, and splitting the data into training and test sets.
3) Modeling: training and tuning various classifiers (Random Forest, Decision Tree, Logistic Regression).
4) Evaluation: assessing the models' performance using appropriate metrics (accuracy, precision, recall, F1-score, ROC AUC).

The project demonstrates the ability to manage a real-world dataset, handle preprocessing challenges, and evaluate different models for a classification problem in the mental health domain.

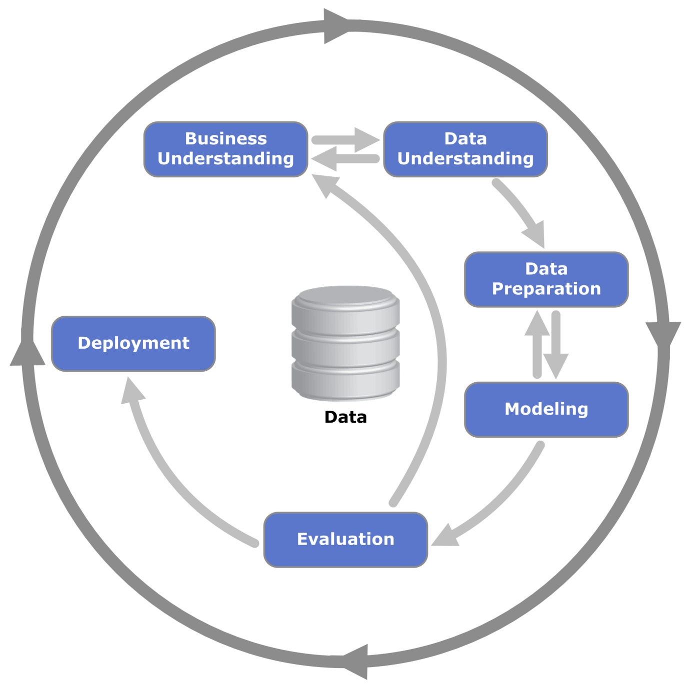

Fig. 1. Crisp-DM

2 Related Work

Mental health prediction, especially related to depression, has been a significant area of research. Many studies have explored the relationship between social media use and depression symptoms. Previous work has utilized Natural Language Processing (NLP) techniques to analyze posts and messages on platforms like Twitter or Reddit, and psychological surveys to identify patterns correlated with depressive behaviors.

Commonly used techniques in the field include:

- Text mining and sentiment analysis on social media posts to capture emotional tone.
- Feature engineering from user profiles (e.g., activity patterns, number of followers, post frequency).
- Supervised machine learning models such as Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Gradient Boosting, to classify users based on risk factors.
- Neural networks and deep learning approaches, especially for text-based data, leveraging embeddings and pre-trained language models (e.g., BERT).

Feature selection methods such as model-based selection with Random Forests, Recursive Feature Elimination (RFE), or Lasso regression are commonly used to enhance model performance and reduce dimensionality. Studies have shown that integrating psychological self-assessment data with demographic and behavioral information can improve the predictive power of machine learning models.

Numerous studies have explored the link between social media usage and mental health, especially depression. Research has shown that excessive use of social media platforms can lead to increased feelings of loneliness, anxiety, and depressive symptoms.

Primack et al. (2017) found a significant correlation between high social media usage and perceived social isolation in young adults. Similarly, Keles et al. (2020), through a systematic review, concluded that social media use was consistently associated with higher levels of depression and anxiety, particularly among adolescents.

From a computational perspective, studies like Reece et al. (2017) utilized Instagram data to detect markers of depression, using image processing and machine learning. Another notable study by Guntuku et al. (2019) examined language use in social media posts, finding linguistic indicators strongly associated with depression.

These works illustrate the potential of using digital footprints for mental health assessment, but also highlight challenges such as data privacy, self-report bias, and generalizability. Our approach focuses on a structured survey dataset rather than passive data collection, aiming for interpretability and ethical transparency.

3 Proposed Method

This project follows the **CRISP-DM** framework, with the following phases:

3.1 Data Understanding

The "smmh.csv" dataset contains 481 rows and 21 columns, including:

- Demographic data: age, gender, relationship status, occupation.
- Social media behavior: time spent on social media, platforms used, validation-seeking behavior.
- Mental health indicators: self-reported feelings of depression, sleep issues, difficulty concentrating.

Initial exploration revealed missing values in the `Affiliate_organization` column (approximately 6%), which were handled by imputing the mode.

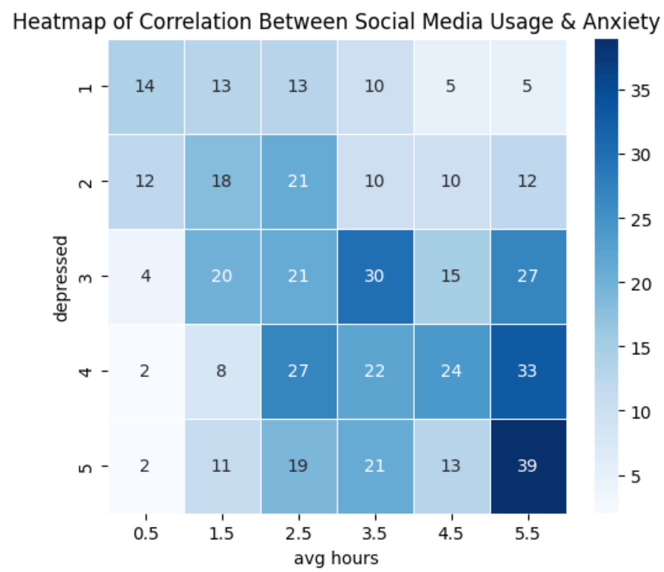

Fig. 2. Heatmap of Correlation Between Social Media Usage - Anxiety

3.2 Data Preparation

The dataset was first inspected for missing values, outliers, and inconsistencies. Columns with textual open-ended responses were excluded from the analysis (e.g., "platforms"). A binary target variable risk was created, labeling individuals with a depression score greater than or equal to 3 as "at risk" (1), and others as not at risk (0).

Categorical features were one-hot encoded to convert them into a numerical format suitable for machine learning. Numerical features were scaled using StandardScaler to standardize their distribution. Data preparation involved several key steps:

- Feature selection using a Random Forest model and `SelectFromModel`, keeping features with importance above the median.
- Splitting the dataset into 80% training and 20% testing using stratified sampling.

3.3 Feature Selection

To improve model performance and reduce overfitting, feature selection was applied using the feature importance scores from a Random Forest classifier. Only features with importance above the median were retained. Selected features included demographic information (e.g., age, gender, relationship status), behavioral patterns (e.g., distraction, sleep issues), and social media usage (e.g., average time spent online).

3.4 Modeling

Two classification models were implemented:

- **Random Forest:** chosen for its robustness, ability to handle mixed data types, and feature importance insights. Hyperparameters were optimized via GridSearchCV.
- **Decision Tree:** implemented for interpretability, with a maximum depth of 3 to prevent overfitting.

3.5 Evaluation

Performance was assessed using:

- Accuracy, precision, recall, F1-score for class-specific analysis.
- ROC AUC to evaluate overall discriminative power.

ROC curves and decision tree visualizations were created for interpretation.

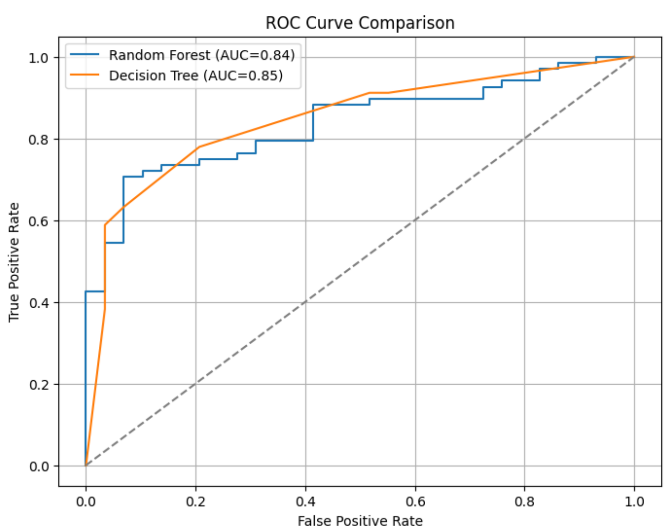

Fig. 4. ROC curve comparison

4 Results

The evaluation results are summarized as follows:

TABLE 1
Performance Comparison of Models

Model	Accuracy	ROC AUC	Precision (1)	Recall (1)
Random Forest	0.77	0.84	0.81	0.88
Decision Tree	0.78	0.85	0.81	0.91

Key insights:

- Time spent on social media and age were the most predictive features.
- The Decision Tree revealed interpretable rules: for example, users spending more than a certain threshold of hours on social media were more likely to be at risk of depression.
- Class imbalance affected the models, prioritizing recall for the depression class over precision.

5 Conclusions

This project demonstrates the feasibility of using structured social media and demographic data to predict depression risk. Tree-based models such as Random Forest and Decision Tree achieved good performance, with Random Forest offering robustness and Decision Tree providing interpretability.

Future work could include applying resampling techniques to address class imbalance, exploring more advanced models (e.g., Gradient Boosting, Neural Networks), and validating the findings on external datasets. This analysis highlights the importance of model interpretability and comprehensive evaluation when applying machine learning to sensitive domains like mental health.

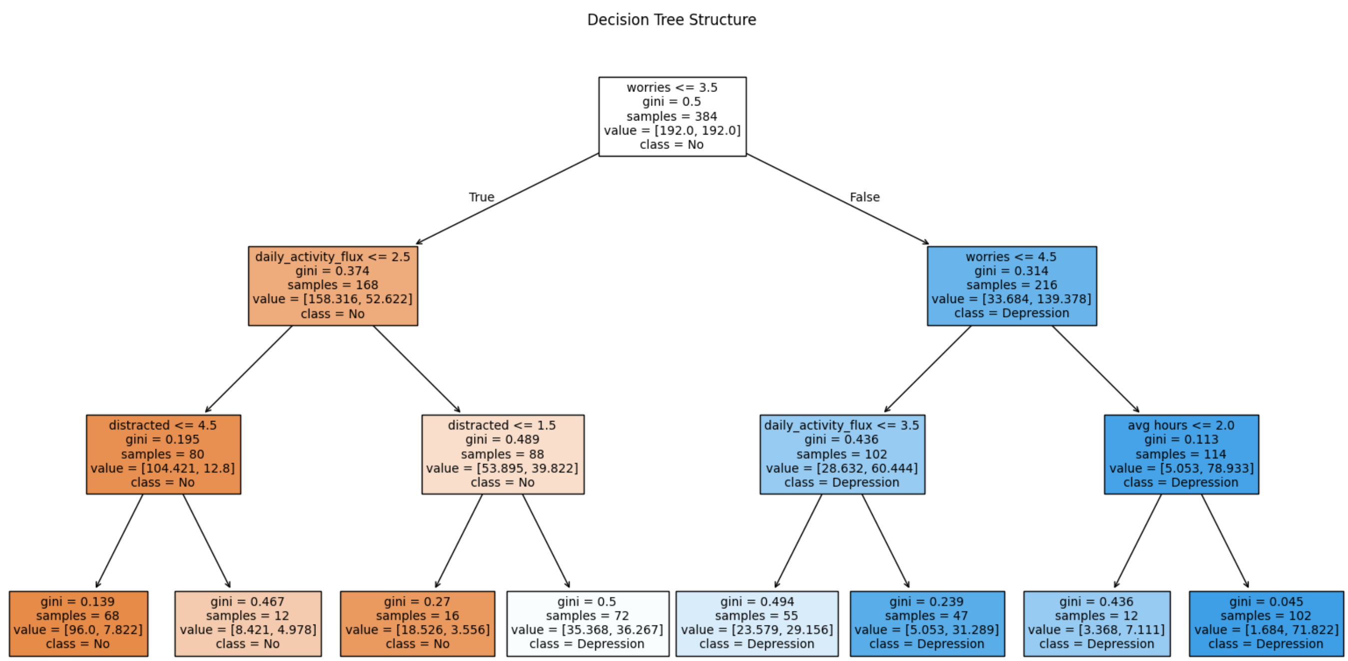

Fig. 3. Decision Tree Structure