YZV311E Project Proposal

1st Furkan Ünüvar
Istanbul Technical University
Artificial Intelligence and Data Engineering
150200334
unuvarf20@itu.edu.tr

2nd Furkan Öztürk
Istanbul Technical University
Artificial Intelligence and Data Engineering
150200312
ozturkahm20@itu.edu.tr

***Abstract*—This project will explore the relationship between social media usage and psychological well-being using a data mining analysis. The project will use a variety of methods, including machine learning and statistical analysis, to identify patterns and relationships in the data. The findings of the project will be communicated through data visualization. The project is expected to provide valuable insights for individuals, mental health professionals, and policymakers.**

I. Introduction

Our project aims to conduct a data mining analysis on the relationship between social media usage patterns and psychological well-being. We will explore how individuals' social media behavior, such as frequency of use, emotional responses, and interactions, may influence their mental health and overall well-being. This research is essential because of the growing concern about the impact of social media on mental health, and it can provide valuable insights for individuals, mental health professionals, and policy-makers.

II. Dataset

For this project, we will use a dataset containing responses to a set of questions related to social media usage and psychological well-being. The dataset is collected through surveys and interviews with a diverse group of participants. The dataset includes information such as age, gender, social media platform usage, daily time spent on social media, psychological well-being scores, and more. We received this data from Kaggle [1]. The data was collected through a survey. Questions such as gender, age, education level, relationship status were asked with multiple options. Other evaluation questions were collected on a rating scale between 1-5. Data was given as csv. The size of the data is 21 columns and 481 rows.

A. Potential Challenges

There may be some empty or invalid data. We will handle this issue by fill these data by deleting, filling on average, or filling in relation to other columns.

III. METHODOLOGY

A. Data Preprocessing

The first stage of our data analysis is to clean and prepare the dataset. During this process, we will undertake the following steps:

Handling Missing Data: We will address missing data, either by imputing missing values or removing observations with missing data using various techniques. Data Cleaning: We will eliminate outliers and meaningless information from the dataset. For example, we might remove negative values in columns like age.

Encoding Categorical Variables: Categorical variables such as gender will be converted into numerical values for modeling purposes.

B. Exploratory Data Analysis (EDA)

To better understand the dataset and extract information, we will perform exploratory data analysis.

Explore Relationships: We will examine relationships between variables through visualization and statistical analysis. For instance, we will investigate the relationship between age and daily time spent on social media.

C. Machine Learning Models

To work with the acquired data, we will build machine learning models.
We will use regression models to predict psychological well-being scores. Models such as logistic regression, decision trees, k-means and random forests will be evaluated.

D. Statistical Analysis

We will employ statistical analysis to identify relationships and patterns within the dataset.
Correlation Analysis: By examining correlations between variables, we will determine which factors have a greater impact on psychological well-being.
Hypothesis Testing: We will evaluate our results using statistical hypothesis tests. For example, we might test whether social media usage has an impact on age.

E. Data Visualization

To better communicate and understand our findings, data visualization will be employed.
Charts and Graphs: We will create various charts, tables, and graphs to describe the features of the dataset and our results. This approach has been formulated to ensure the successful progression of the project and the reliability of the results.

IV. EVALUATION METHODS

To assess the success of our project, we will employ the following evaluation methods:

Evaluation Metrics: We will use metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and accuracy to measure the performance of our predictive models for psychological well-being and sentiment analysis.

Cross-Validation: We will utilize cross-validation techniques to ensure that our models generalize well and do not overfit the data.

V. GITHUB REPOSITORY

Our GitHub repository: https://github.com/Unuvar59/Group_9

VI. TIME PLAN & DISTRIBUTION OF WORK

In the Table 1, we have indicated who is working on which issues and how many weeks it will take to complete.

TABLE I

	Furkan Ünüvar	**Furkan Öztürk**
3-4 week	Data Preprocessing: Handle missing data, data cleaning, and encoding categorical variables.	Machine Learning Models: Implement regression models for predicting psychological well-being scores.
1-2 week	Exploratory Data Analysis (EDA): Explore relationships between variables, conduct data visualizations.	Evaluation Methods: Implement evaluation metrics, such as MAE, MSE, and cross-validation.
1-2 week	Statistical Analysis: Perform correlation analysis and hypothesis testing.	Data Visualization: Create charts, graphs, and tables to visualize the dataset and results.
per week	GitHub Repository: Maintain and update the project repository on GitHub.	GitHub Repository: Maintain and update the project repository on GitHub.
1 week	Report Writing: Contribute to the writing of the project report.	Report Writing: Contribute to the writing of the project report.

VII. CONCLUSION

In conclusion, our project aims to delve into the intricate relationship between social media usage patterns and psychological well-being. By exploring how various aspects of social media behavior affect mental health, we hope to shed light on a topic of increasing concern. This research has the potential to provide valuable insights not only for individuals but also for mental health professionals and policymakers who are seeking to address the impact of social media on our well-being. Throughout the project, we will diligently follow the outlined methodology, which includes data preprocessing, exploratory data analysis, machine learning modeling, statistical analysis, and data visualization. We are confident that our systematic approach will yield valuable results and facilitate a comprehensive understanding of this complex issue.

REFERENCES

[1] https://www.kaggle.com/datasets/souvikahmed071/social-media-and-mental-health