Prediction of Social Media Engagement Analysis Using Linear Regression

Shreya G (4AD21CS090)

Rakshitha M K (4AD21CS078)

Lamya Krushid (4AD21CS041)

Nausheen Fathima (4AD21CS056)

Aiyan Ahmed Khan (4AD21CS005)

Course Name: Python for data science

* Sowndarya
* 11-9-2024

# Abstract

This project aims to predict social media engagement, specifically the total likes a user might receive, based on their age and social media usage duration. Using a linear regression model, we analyze a dataset to determine trends in user engagement. Results demonstrate a measurable correlation between these factors and user engagement, providing insights for targeted content creation.

# Introduction

* **Background**: Social media platforms play a significant role in modern communication. Predicting user engagement can enhance targeted advertising and content personalization.
* **Objectives**: To predict the total likes based on usage duration and age, providing insights into factors driving social media engagement.
* **Scope**: This project limits its scope to basic regression analysis on two features (Usage Duration and Age).

# Literature Review

Previous studies show that user engagement is influenced by multiple factors, including demographic characteristics and activity levels. Linear regression is commonly used in similar studies to analyze relationships between variables.

# Methodology

* + **Data Collection**: The dataset contains information on UsageDuration, Age, and TotalLikes. This data was sourced to reflect patterns in social media engagement

and includes relevant demographic and activity features.

* + **Data Processing:** Initial preprocessing included renaming columns, addressing null values, and standardizing the data format. The UsageDuration and Age features were selected as predictors, while TotalLikes served as the target variable for prediction.
  + **Model Design**: A linear regression model was selected due to its ability to establish clear relationships between continuous variables. This model structure was deemed suitable for the problem as it allows for straightforward interpretation and effective trend identification.
  + **Model Training**: The data was split into training and testing sets, with 80% for training and 20% for testing. The model was then trained on the training set, learning the relationship between UsageDuration, Age, and TotalLikes.
  + **Evaluation**: The model’s performance was evaluated using Mean Squared Error (MSE) to assess prediction accuracy and R-squared (R²) to measure how well the model captured the variability in TotalLikes. Visualization of predicted versus actual values provided further insights into the model’s predictive capacity.

# GITHUB Link :

# <https://github.com/ShreyaGuruprasad5/Linear-Regression-Project-using-Machine-Learning>

# Experiments and Results:

# 

**Results**:

The project successfully analyzed and interpreted user data through visual plots, providing valuable insights into user behavior and demographics. Our analysis identified consistent user ID assignment, varied usage durations, and an age range that demonstrates demographic diversity, as well as highlighted peaks in engagement levels.