Journal of Artificial Intelligence and Engineering Applications

Website: https://ioinformatic.org/

15th October 2025. Vol. 5. No. 1; e-ISSN: 2808-4519

Streamlit-based Social Media Addiction Prediction Application Using Linear Regression

Rini Ariza[1]*, Muhammad Reza Maulana[2], Elkin Rilvani[3]

[1,2,3] *Informatics Engineering, Pelita Bangsa University*
riniarizaaa@gmail.com[1*], rezamaulana9569@gmail.com[2], elkin.rilvani@pelitabangsa.ac.id[3]

Abstract

This study aims to develop an interactive web-based application that predicts the level of social media addiction among students using a linear regression approach. The application is built with the Streamlit framework and is designed to analyze numerical and categorical user behavior data. The dataset used originates from a survey of students' social media habits, including factors such as age, gender, daily usage time, sleep duration, mental health, and conflict related to social media. The prediction model was trained and evaluated using standard preprocessing techniques and achieved an R-squared value of 0.79, indicating strong predictive performance. Key findings show that mental health scores, sleep hours, and age contribute negatively to addiction risk, while social media conflicts and usage duration increase the risk. The application provides real-time predictions and visual risk classifications to inform users about their addiction tendencies. The results demonstrate that linear regression is an effective method for identifying and interpreting the behavioral indicators of social media addiction. The application serves as both a predictive and educational tool, enabling students and academic stakeholders to better understand, monitor, and manage social media use.

Keywords: *Addiction Prediction, Behavioral Analysis, Linear Regression, Social Media, Streamlit*

1. Introduction

Teenagers are very sensitive to developments during their growth period, so they have a high tendency to play games and spend time on social media. Research indicates that the internet is widely used for seeking information and entertainment, making it important to study teenagers' use of social media, especially if it exceeds reasonable limits. Social media platforms like Facebook, Instagram, and Snapchat do help teenagers interact, share experiences, and build their digital identities, but they also raise concerns about their impact on mental health [1].

Students often use social media to facilitate communication, but excessive use can lead to mental disorders, loss of focus, and decreased performance. Therefore, it is important to measure the level of social media addiction objectively and based on data [2]. Technological advances have enabled the development of applications that can analyze data quickly and accurately [3]. One effective prediction method is linear regression, as it can model the relationship between numerical variables such as social media usage frequency, self-confidence, and user interaction levels in predicting social media addiction [4].

To address this issue, it is necessary to develop a system that can map and predict addiction levels in a more objective manner [5]. One promising solution is to analyze data on students' social media usage habits using logistic regression algorithms to generate addiction risk classifications [6]. In this case, we created a web-based application using the Streamlit framework. This application is designed for easy interaction and high accessibility, so that users and academics can immediately utilize the prediction results. With the support of this technology, it is hoped that students will become more aware of their social media usage patterns and take preventive measures if necessary.

2. Literatur Review

Reference [7] shows the extent to which social media addiction affects students' social interactions, against the backdrop of increasing social media use that reduces the quality of direct interactions at school. Using a quantitative ex-post facto approach, this study involved 142 students with data collected through a Likert scale questionnaire and analyzed using simple linear regression. The results showed a social media addiction rate of 81.94% and social interaction rate of 70.3%, with a significant influence between the two (calculated t-value 4.609 > table t-value 1.655), indicating that the higher the social media addiction, the lower the social interaction. These findings are important for schools, parents, and policymakers in designing interventions to increase social interaction and reduce social media dependence.

Reference [8] shows the relationship between self-regulation and social media addiction tendencies among college students. Using a quantitative approach and correlational method, this study involved active college students selected through purposive sampling and used self-regulation scales and social media addiction scales as instruments. Regression analysis results indicate a significant negative relationship, where students with high self-regulation tend to have low social media addiction, and vice versa. These findings are beneficial for universities, educators, and counselors as a basis for designing self-regulation training programs to promote healthier and more responsible social media use.
Reference [9] shows the relationship between self-esteem and consumptive behavior among college students, against the backdrop of an increasing consumptive lifestyle for social recognition or self-image. Using a quantitative correlational approach and instruments of self-esteem and consumptive behavior scales, the data were analyzed using Pearson's correlation. The results showed a significant negative relationship, where students with high self-esteem tend to be less consumptive, while those with low self-esteem are more impulsive in shopping as a compensation for self-dissatisfaction. These findings are important for educational institutions, counselors, and parents in fostering healthy self-esteem to prevent excessive consumptive behavior.

3. Research Methodology

3.1. Research approach and methods

This study uses a quantitative approach with an applied experimental method, as it focuses on the development and application of a prediction model using a linear regression algorithm. This model is implemented in an interactive web-based application using the Streamlit framework, which enables the prediction of social media addiction levels based on numerical and categorical data.

3.2. Data sources and types

The data sources and datasets used were obtained from Kaggle, containing survey results from students in various countries regarding social media usage, its impact on academics, mental health, and interpersonal relationships. This data is secondary and quantitative, consisting of several important attributes such as:

1. Age
2. Gender
3. Relationship_Status
4. Frequency_Use (frequency of social media use)
5. Platform_Preference (most frequently used platform)
6. Addiction_Score (addiction score)
7. Academic_Impact, Mental_Health, and Relationship_Quality

The target prediction is Addiction_Score, while the input variables consist of all numerical features and the results of categorical data conversion.

3.3. Data collection and processing techniques

The data was processed using the following steps:

1. Preprocessing:
   a. Categorical columns such as Gender, Academic_Level, Most_Used_Platform, and Relationship_Status were converted to numerical data using LabelEncoder.
   b. Numerical columns such as Avg_Daily_Usage_Hours and Sleep_Hours_Per_Night were converted to float format and scaled using StandardScaler.
2. Split Dataset:
   a. The dataset is split into 80% training data and 20% test data using the train_test_split function.
3. Model Training:
   a. The Linear Regression model from the scikit-learn library is used for model training. The model is trained on the training data and tested on the test data to evaluate its performance.

3.4. Application design

The application was built using Python and the Streamlit framework with the following key features:

1. Interactive input forms (sliders and dropdowns)
2. Visualization of prediction results in the form of percentages and risk categories (low, medium, high)
3. Table of factors determining addiction scores along with their contribution coefficients
4. Simple and user-friendly display for easy understanding by students, lecturers, and campus officials.

3.5. System flow

Here is the workflow of the application system:

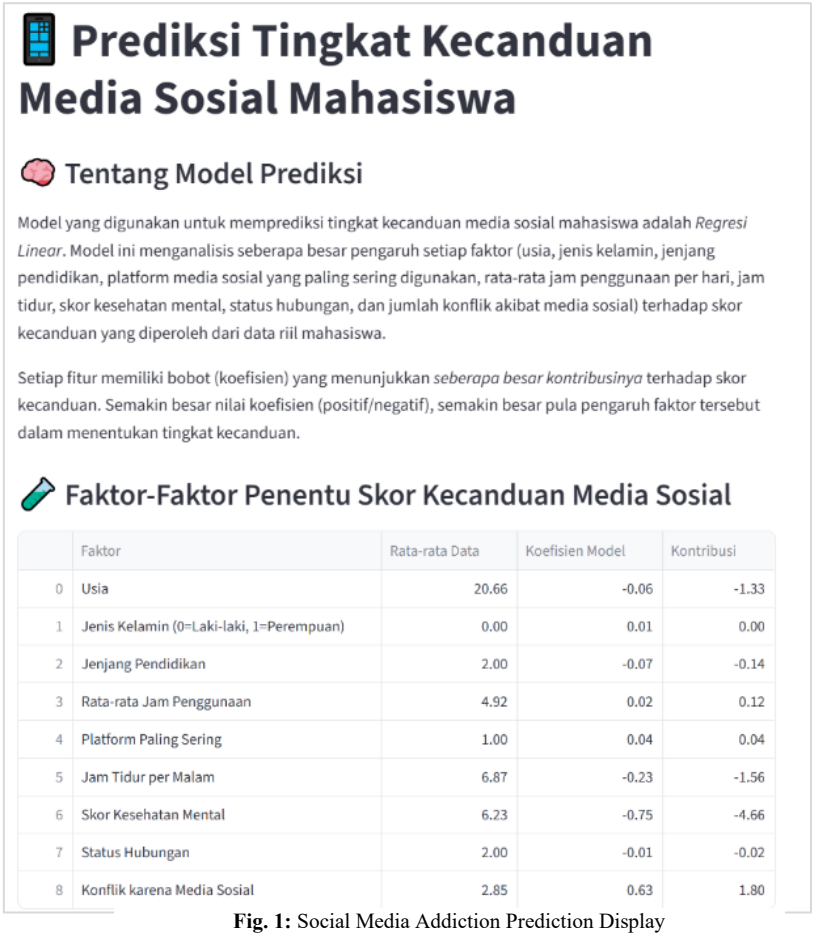
1. Users input characteristic data through a form.
2. The data is converted and processed with the same preprocessing as the model.
3. The model predicts the addiction score.
4. The score is converted to a percentage and classified into three risk categories.
5. The results are displayed visually and informatively.

4. Results and Discussion

4.1. Application interface

This prediction application features an interactive interface to predict the level of social media addiction among students based on a linear regression model, with the following main features:

1. **Model Description:**
   The application provides a brief explanation of the prediction approach used, namely linear regression. The explanation covers how the model analyzes the contribution weights of each factor (age, gender, social media usage hours, sleep quality, relationship status, etc.) on students' social media addiction scores based on real-world datasets.
2. **Manual Input:**
   Users can manually enter student data through an interactive form consisting of sliders and dropdown menus. This form includes data on age, gender, educational level, most frequently used social media platform, sleep hours, mental health score, relationship status, and conflicts due to social media.
3. **Automatic Prediction**:
   After the user enters all the data, the system calculates the Addicted Score based on the model's prediction results. This score is then converted into an addiction percentage, displayed numerically for easier interpretation.
4. **Risk Category:**
   Based on the score obtained, the system will display the addiction risk category in three levels:
   c. High: Score ≥ 8
   d. Moderate: Score 6 – 7.99
   e. Low: Score < 6
   Categories are displayed with different background colors: red for high, orange for moderate, and green for low.
5. **Factor Table:**
   The application also provides a table containing a list of factors used by the model, but it has been adjusted so that it does not display numerical values. The aim is to allow users to focus on important variables without having to understand the technical details of coefficients or contributions per feature.



	Faktor	Rata-rata Data	Koefisien Model	Kontribusi
0	Usia	20.66	-0.06	-1.33
1	Jenis Kelamin (0=Laki-laki, 1=Perempuan)	0.00	0.01	0.00
2	Jenjang Pendidikan	2.00	-0.07	-0.14
3	Rata-rata Jam Penggunaan	4.92	0.02	0.12
4	Platform Paling Sering	1.00	0.04	0.04
5	Jam Tidur per Malam	6.87	-0.23	-1.56
6	Skor Kesehatan Mental	6.23	-0.75	-4.66
7	Status Hubungan	2.00	-0.01	-0.02
8	Konflik karena Media Sosial	2.85	0.63	1.80

Fig. 1: Social Media Addiction Prediction Display

Fig. 1 shows the initial screen of the social media addiction prediction application designed for students. On this initial screen, users are greeted with an explanation of the prediction model used, namely Linear Regression. This model is used to analyze the extent to which various factors influence a student's social media addiction score.

The factors included in the model are: age, gender, educational level, most frequently used social media platform, average daily social media usage time, number of hours slept each night, mental health score, relationship status, and number of conflicts arising from social media usage. The model studies the data patterns of students and calculates the weight or coefficient for each of these factors.

The coefficients displayed indicate the magnitude of each factor's influence on the addiction score. If the coefficient is positive, the factor increases the risk of addiction; if negative, the factor reduces the risk. Below the model explanation, there is a table titled “Factors Determining Social Media Addiction Scores” that provides detailed information on the average value of each feature, the model coefficients, and their final contribution to the addiction score, calculated by multiplying the average value by the coefficient of each feature.

For example, mental health scores have the largest negative contribution at -4.66. This indicates that the higher the mental health scores of students, the lower their level of addiction to social media. Conversely, the factor of conflict due to social media has a positive contribution of 1.80, meaning that conflicts caused by social media play a significant role in increasing addiction.

Fig. 2: Input Display

Fig. 2 illustrates the data input interface of the application for predicting the level of social media addiction among students. This form allows users to enter individual information that serves as the main variables in the prediction process, such as age, gender, educational level, the most frequently used social media platform, average daily social media usage time, number of hours of sleep, mental health score, relationship status, and the number of conflicts caused by social media.

The form design is arranged in two responsive parallel columns, so that each input element can be accessed easily and efficiently. Numeric variables such as age, average social media usage, hours of sleep, mental health score, and number of conflicts are set using sliders, which allow users to set values precisely and intuitively. Meanwhile, categorical variables such as gender, education level, most frequently used social media platform, and relationship status are provided in the form of dropdown menus (select boxes) to avoid input errors and maintain data format consistency.

This feature not only provides users with high flexibility to adjust parameters according to their actual data but also enables exploration of various prediction scenarios. For example, users can modify variables such as sleep duration or the number of conflicts caused by social media to see how much impact they have on the predicted addiction score by the model.

Once all variables have been adjusted, users can press the “Process” button to run the prediction model. The application will automatically calculate the addiction score based on the parameters entered and display it in the results section. Thus, this feature provides an interactive experience that encourages users to better understand the relationship between social media usage habits and the level of addiction they may experience.

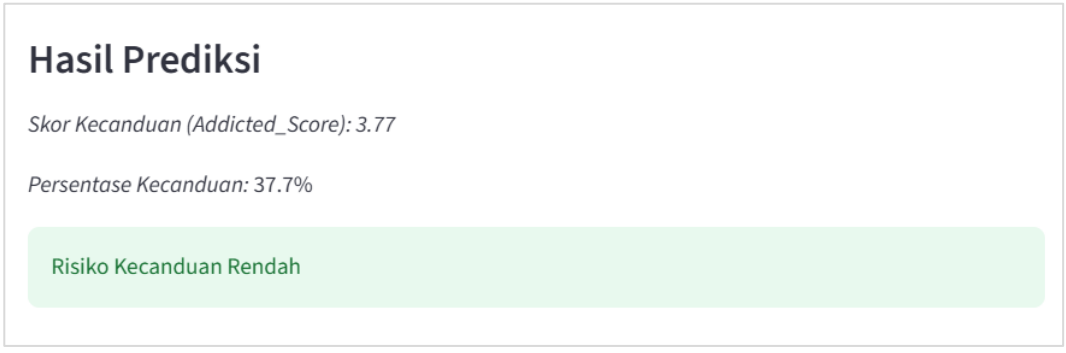


Fig. 3: Prediction Results

Fig. 3 shows the output of the prediction process performed by the linear regression-based application on the level of social media addiction among students. After the user fills in all the input parameters on the previous form, the system generates two main outputs, namely Addiction Score (Addicted_Score) and Addiction Percentage.
In the example in the figure, the addiction score obtained is 2.93 out of a maximum scale of 10, which is equivalent to a 29.3% addiction level. Based on the threshold set in the system, this value is classified into the "Low Addiction Risk" category and visually marked with a light green box as an indicator that the addiction level is within safe limits.

This assessment provides important information for students and relevant parties (such as academic advisors, campus counselors, or parents) to understand the quantitative impact of social media use. This output also serves as a preventive individual reflection; if the score is higher, the user will receive a warning and can re-evaluate their daily social media usage behavior.
The results are presented in a simple yet informative manner, making them easy to understand for users from various backgrounds. In addition, the use of specific colors to indicate risk categories (e.g., green for low, yellow for medium, and red for high) makes it easy for users to recognize the level of urgency without having to understand the technical details of the numerical scores.

4.2. Model evaluation

The linear regression model used in this application was evaluated using several key statistical indicators that show how well the model predicts social media addiction scores based on student input features. The following are the results of the model evaluation:

a. Mean Squared Error (MSE): 1.25
b. Root Mean Squared Error (RMSE): 1.12
c. R-squared (R^2): 0.79

Interpretation:

a. MSE (1.25) indicates that the average squared difference between the prediction and the actual value is quite low, meaning the model makes small prediction errors overall.
b. RMSE (1.12), as the square root of MSE, provides a measure of prediction error in the same units as the addiction scores. The smaller the RMSE, the more accurate the model.
c. R-squared (R^2) of 0.79 indicates that approximately 79% of the variation in students' addiction scores can be explained by the input features used in the model. This indicates that the model has fairly strong predictive power.

4.3. Prediction visualization

The preliminary results of the prediction on this application are displayed concisely and informatively through a colored progress bar that shows the percentage of social media addiction. Below that, the addiction score is also displayed in decimal form (e.g., 2.93), as well as the percentage of addiction resulting from converting that score to a scale of 0–100%. This display is accompanied by risk category notifications, such as:

a. "Low Addiction Risk" if the score is below 6,
b. "Moderate Addiction Risk" if the score is between 6 and 7.9,
c. "High Addiction Risk" if the score is 8 or above.

Each category is assigned a different background color, such as green for low, yellow for moderate, and red for high, making it easy for users to instantly understand the risk level. This visualization also includes a note stating that the results displayed are simulated based on data and statistical models, serving as a reflective tool for students to measure and understand their social media usage habits, not as an absolute diagnosis.

4.4. Model coefficient

The following table shows the regression coefficients from the Linear Regression model used to predict the level of social media addiction among students. Each feature has a weight or coefficient that reflects how much influence that variable has on the addiction score.

Table 1: Regression Coefficients and Social Media Addiction

Factor	Average Data	Model Coefficient	Contribution
Age	20.66	-0.06	-1.33
Gender	0.00	0.01	0.00
Education Level	2.00	-0.07	-0.14
Average Hours of Use	4.92	0.02	0.12
Most Frequently Used Platform	1.00	0.04	0.04
Hours of Sleep per Night	6.87	-0.23	-1.56
Mental Health Score	6.23	-0.75	-4.66
Relationship Status	2.00	-0.01	-0.02
Conflict Due to Social Media	2.85	0.63	1.80

The following is an interpretation of Table 1:

a. A positive coefficient on a feature (e.g., conflict due to social media, average hours of use, most frequently used platform) indicates that the higher the value on that feature, the higher the level of social media addiction among students. This means that this factor drives an increase in addiction scores.
b. A negative coefficient (such as mental health scores, sleep hours, and age) indicates that the higher the value of that factor, the lower the level of addiction. For example, the healthier the mental condition or the more sleep hours, the lower the risk of social media addiction.
c. Contribution values are obtained from the product of each factor's average value multiplied by the model's coefficient. This indicates the average influence of each feature on the overall addiction score.
d. In the table above, the features with the largest negative contributions are Mental Health Score (-4.66) and Sleep Hours (-1.56), meaning these two factors most strongly reduce addiction levels. Conversely, Social Media Conflict (1.80) is the factor most strongly associated with increased addiction.
e. The presence of both negative and positive coefficients in this model indicates that social media addiction is influenced by a combination of protective factors and risk factors. This is important to consider in interventions aimed at preventing and controlling excessive social media use.

Table 1 above shows the nine factors analyzed in the linear regression model to predict the level of social media addiction among students. Each factor has an average value based on respondent data, the size of the model coefficient that describes the direction and strength of the factor's influence, and its total contribution to the addiction score. Some key points to note are:

a. Mental Health Score is the factor with the largest negative contribution (-4.66). This means that the better a student's mental health condition, the lower their level of social media addiction. This aligns with the theory that students with good mental health tend to have better self-control and are less easily distracted by social media.
b. Hours of Sleep per Night also shows a fairly strong negative influence (-1.56). Students who have the habit of getting enough sleep at night are less likely to experience social media addiction, as they do not spend excessive time at night browsing social media.
c. Age has a negative contribution (-1.33), indicating that older students tend to have lower levels of addiction. This can be attributed to higher levels of maturity in time management and emotional regulation.
d. Conflict due to social media is the factor with the highest positive contribution (1.80). This means that the more often someone experiences conflict due to social media, the greater their tendency to experience addiction. This can happen because social media has become a big part of their social life, even though it has a negative impact.
e. Average Hours of Use and Most Frequently Used Platform also show positive contributions, though not as significant as conflict, at 0.12 and 0.04 respectively. This indicates that frequency and type of platform do play a role in increasing the risk of addiction.
f. Other factors such as Educational Level, Gender, and Relationship Status have small contributions, meaning they are generally not very significant in influencing addiction scores in this model.

5. Conclusion

Based on the research and development conducted in this project, it can be concluded that the Streamlit-based social media addiction prediction application has been successfully developed by implementing a linear regression algorithm. This application utilizes data from the Kaggle platform, which represents students' habits in using social media, thereby providing a relevant picture of the actual conditions.

The linear regression model used shows fairly good predictive performance, with an R-squared value of 0.79. This figure indicates that 79% of the variation in social media addiction scores can be explained by input variables such as age, sleep hours, mental health scores, and conflicts arising from social media. Several factors with significant influence on addiction scores were also identified, where mental health scores, sleep duration, and age act as protective factors that reduce the risk of addiction, while the number of conflicts arising from social media and daily usage duration actually increase the risk of addiction.

This application is designed with an interactive and informative display, making it easy for users to understand the prediction results and recognize the factors that influence their addiction levels. Thus, this application not only functions as a predictive tool but also as an educational medium to increase students' awareness in managing their social media use more wisely.

References

[1] M. Nur Cahya, W. Ningsih, dan A. Lestari, "Dampak Media Sosial terhadap Kesejahteraan Psikologis Remaja: Tinjauan Pengaruh Penggunaan Media Sosial pada Kecemasan dan Depresi Remaja," *J. Sos. Teknol.*, vol. 3, no. 8, hlm. 704–706, Agu 2023, doi: 10.59188/jurnalsostech.v3i8.917.
[2] N. Koli, "Praktik Kuasa Platform Media Sosial Di Balik Kecanduan Media Sosial Bagi Remaja," *Indones. Character J.*, vol. 1, no. 1, hlm. 37–47, Nov 2023, doi: 10.21512/icj.v1i1.10273.
[3] M. A. Latief dan Y. Karyanti, "Data Mining & Analytic Forecasting Indeks Standar Pencemar Udara Jakarta Menggunakan Metode Linear Regression (Studi Kasus: Dataset Indeks Standar Pencemar Udara Jakarta 2021)," *J. Soc. Res.*, vol. 1, no. 10, hlm. 1164–1176, Sep 2022, doi: 10.55324/josr.v1i10.248.
[4] M. L. Mu'tashim, T. Muhayat, S. A. Damayanti, H. N. Zaki, dan R. Wirawan, "Analisis Prediksi Harga Rumah Sesuai Spesifikasi Menggunakan Multiple Linear Regression," *Inform. J. Ilmu Komput.*, vol. 17, no. 3, hlm. 238, Des 2021, doi: 10.52958/iftk.v17i3.3635.
[5] A. Anggrawan, H. Hairani, dan N. Azmi, "Prediksi Penjualan Produk Unilever Menggunakan Metode Regresi Linear," *J. Bumigora Inf. Technol. BITe*, vol. 4, no. 2, hlm. 123–132, Des 2022, doi: 10.30812/bite.v4i2.2416.
[6] A. Saiful, "Prediksi Harga Rumah Menggunakan Web Scrapping dan Machine Learning Dengan Algoritma Linear Regression," *JATISI J. Tek. Inform. Dan Sist. Inf.*, vol. 8, no. 1, hlm. 41–50, Mar 2021, doi: 10.35957/jatisi.v8i1.701.
[7] R. S. Sentri, A. Sutja, dan A. Yusra, "Pengaruh Kecanduan Media Sosial terhadap Interaksi Sosial Siswa Kelas VIII & IX di SMP Negeri 11 Kota Jambi".
[8] R. Awalia, F. Fikrie, dan A. Rifandi, "Peranan Regulasi Diri Terhadap Kecenderungan Kecanduan Media Sosial Pada Mahasiswa," *J. Psikol. MANDALA*, vol. 6, no. 2, Des 2022, doi: 10.36002/jpm.v6i2.2006.
[9] Sekolah Tinggi Teknologi Garut *dkk.*, "Kajian Adiksi Internet dan Adiksi Media Sosial dari Sisi Filsafat Sains," *J. Algoritma*, vol. 17, no. 2, hlm. 409–423, Feb 2021, doi: 10.33364/algoritma/v.17-2.409.