Social Media Addiction Predictor

Session: 2023 – 2027

Submitted by:

Muhammad Tayyab 2023-CS-101

Supervised by:

Sir. Kabir Ahmad

Course:

CSC-200 Artificial Intelligence Lab

Table of Contents

1. Executive Summary

The Social Media Addiction Prediction System is a comprehensive machine learning application designed to analyze and predict social media addiction levels among students aged 15-30. This innovative solution combines advanced machine learning algorithms with AI-powered personalized recommendations to provide users with detailed insights into their social media usage patterns.

1.1 Key Achievements:

• Developed an ensemble machine learning model achieving high prediction accuracy
• Integrated Google's Gemini AI for personalized recommendations and analysis
• Created an intuitive web interface using Streamlit framework
• Implemented PDF report generation for comprehensive user analysis
• Designed a scalable architecture supporting multiple deployment options

The system serves as both a diagnostic tool and an educational platform, helping users understand their digital consumption patterns while providing actionable insights for maintaining healthy social media habits.

2. Project Overview

2.1 Problem Statement

Social media addiction has become a significant concern among young adults and students, affecting academic performance, mental health, and overall well-being. Traditional assessment methods are often subjective and lack the precision needed for effective intervention. This project addresses the need for an objective, data-driven approach to identify and analyze social media addiction patterns.

2.2 Solution Approach

Our solution leverages machine learning algorithms to analyze behavioral patterns and predict addiction levels on a 0-10 scale. The system incorporates multiple data points including usage hours, platform preferences, academic impact, sleep patterns, and mental health indicators to provide comprehensive addiction assessment.

2.3 Target Audience

2.3.1 Primary Users:

• Students and young adults (ages 15-30)
• Educational counselors and advisors
• Mental health professionals
• Researchers in digital wellness

2.3.2 Secondary Users:

• Parents and guardians
• Educational institutions
• Digital wellness organizations

3. Technical Architecture

3.1 System Components

The application follows a modular architecture with clear separation of concerns:

Core Modules:
• app.py - Main application controller and Streamlit configuration
• modelHandler.py - Machine learning model management and prediction logic
• userIO.py - User input handling and results display
• ui.py - User interface components and styling
• config.py - Configuration management and constants
• Gemini_integration.py - AI-powered analysis and recommendations
• AnalyzeAddiction.py - Core addiction analysis algorithms

3.2 Data Flow Architecture

1. User Input Collection: Multi-step form collects demographic and behavioral data
2. Data Preprocessing: Input validation, encoding, and scaling
3. Model Prediction: Ensemble model generates addiction score
4. AI Analysis: Gemini AI provides personalized insights
5. Results Display: Interactive visualizations and recommendations
6. Report Generation: PDF export with comprehensive analysis

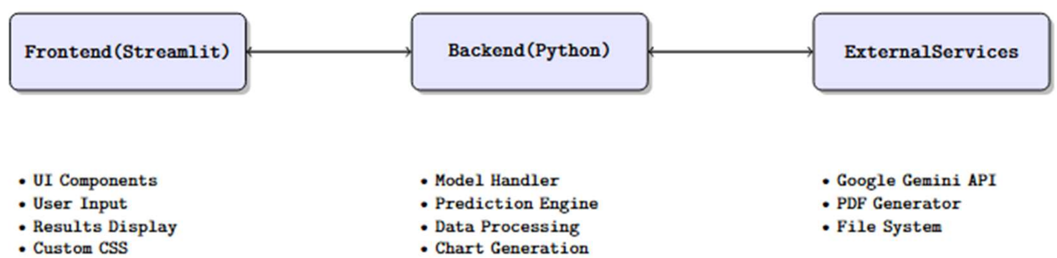

Fig. System Architecture

3.3 Technology Stack

Technology	Purpose
Python 3.12+	Core programming language
Streamlit	Web application framework
Scikit-learn	Machine learning algorithms
Pandas/NumPy	Data manipulation and analysis
Google Gemini AI	AI-powered recommendations
FPDF2	PDF report generation
Plotly	Interactive data visualizations
Python-dotenv	Environment configuration

4. Feature Analysis

4.1 Input Features

The system analyzes ten key behavioral and demographic features:

Feature	Type	Description	Values/Range
Age	Numeric	Student age	15-30 years
Gender	Categorical	Student gender	Male, Female
Academic_Level	Categorical	Education level	High School, Undergraduate, Graduate
Country	Categorical	Country of residence	195+ countries
Avg_Daily_Usage_Hours	Numeric	Daily social media usage	0.5-12.0 hours
Most_Used_Platform	Categorical	Primary platform	Instagram, TikTok, Facebook, YouTube, WhatsApp
Affects_Academic_Performance	Categorical	Academic impact	Yes, No
Sleep_Hours_Per_Night	Numeric	Sleep duration	4.0-10.0 hours
Mental_Health_Score	Numeric	Mental health rating	1-10 scale
Conflicts_Over_Social_Media	Numeric	Conflict frequency	0-5 incidents

4.2 Addiction Level Categories

The following categories are derived based on the predicted addiction score, which is calculated using the machine learning model and input features such as age, gender, academic level, daily usage, and more as detailed in the previous section.

Score Range	Level	Description
0-3	Low	Well-controlled usage with healthy balance
3-6	Moderate	Moderate usage requiring monitoring
6-8	High	Concerning patterns suggesting intervention
8-10	Very High	Severe addiction requiring professional support

5. Machine Learning Models

5.1 Models Tested

This system uses five machine learning algorithms to predict social media addiction scores. Each model is evaluated to choose the best one based on performance.

5.1.1 Linear Regression

LinearRegression()

- **Pros**: Fast, simple
- **Cons**: Only works well with linear data
- **Use**: Quick predictions, basic trend analysis

```
Training Linear Regression...
Results for Linear Regression:
  RMSE: 0.3303
  MAE: 0.2405
  R² Score: 0.9564
  Accuracy: 95.64%
  CV Score: 0.9509 (±0.0102)
```

Fig. Training LR

5.1.2. Random Forest Regressor

RandomForestRegressor(n_estimators=100, random_state=42)

- **Pros**: Accurate, stable
- **Cons**: Hard to interpret
- **Key Settings**:
  - n_estimators=100
  - random_state=42

```
Training Random Forest...
Results for Random Forest:
  RMSE: 0.2127
  MAE: 0.0801
  R² Score: 0.9819
  Accuracy: 98.19%
  CV Score: 0.9800 (±0.0034)
```

Fig. Training RF

5.1.3 K-Nearest Neighbors

KNeighborsRegressor(n_neighbors=5)

- **Pros**: Easy to understand
- **Cons**: Slow with large data
- **Key Setting**: n_neighbors=5

- **Note**: Needs scaling with StandardScaler

```
Training KNN...
Results for KNN:
  RMSE: 0.2989
  MAE: 0.1433
  R² Score: 0.9643
  Accuracy: 96.43%
  CV Score: 0.9639 (±0.0130)
```

Fig. Training KNN

5.1.4 Support Vector Machine

SVR(kernel='rbf', C=1.0, gamma='scale')

- **Pros**: Works in complex data
- **Cons**: Needs feature scaling
- **Key Settings**:
  - kernel='rbf'
  - C=1.0
  - gamma='scale'
- **Note**: Use StandardScaler

```
Training SVM...
Results for SVM:
  RMSE: 0.2805
  MAE: 0.1623
  R² Score: 0.9686
  Accuracy: 96.86%
  CV Score: 0.9645 (±0.0109)
```

Fig. Training SVM

5.1.5 Gradient Boosting Regressor

GradientBoostingRegressor(n_estimators=100, random_state=42)

- **Pros**: Very accurate
- **Cons**: Can overfit, slow
- **Key Settings**:
  - n_estimators=100
  - random_state=42

```
Training Gradient Boosting...
Results for Gradient Boosting:
  RMSE: 0.2236
  MAE: 0.1326
  R² Score: 0.9800
  Accuracy: 98.00%
  CV Score: 0.9780 (±0.0047)
```

Fig. Training XGBoost

5.2 Model Selection Process

The system trains all five machine learning models on the dataset. After training, it evaluates each model using the R^2 score to measure how well it explains the variance in the data. The model with the highest R^2 score is automatically selected as the best performer. This ensures optimal prediction accuracy for the addiction score.

```
============================================================
MODEL COMPARISON SUMMARY
============================================================
Linear Regression    | R²: 0.9564 | RMSE: 0.3303 | Accuracy: 95.64%
Random Forest        | R²: 0.9819 | RMSE: 0.2127 | Accuracy: 98.19%
KNN                  | R²: 0.9643 | RMSE: 0.2989 | Accuracy: 96.43%
SVM                  | R²: 0.9686 | RMSE: 0.2805 | Accuracy: 96.86%
Gradient Boosting    | R²: 0.9800 | RMSE: 0.2236 | Accuracy: 98.00%
```

Fig. Comparison Table

5.3 Best Model

```
🏆 BEST MODEL: Random Forest
   R² Score: 0.9819
   RMSE: 0.2127
   Accuracy: 98.19%
   Cross-validation: 0.9800 (±0.0034)
```

Fig. Best Model

6. User Interface Design

6.1 Design Philosophy

The interface prioritizes simplicity, accessibility, and user engagement:

Design Principles:
• Clean, modern aesthetic with professional blue color scheme
• Responsive layout adapting to different screen sizes
• Progressive disclosure of information to prevent overwhelm
• Clear visual hierarchy guiding user attention
• Accessibility compliance for inclusive user experience

6.2 Wireframes

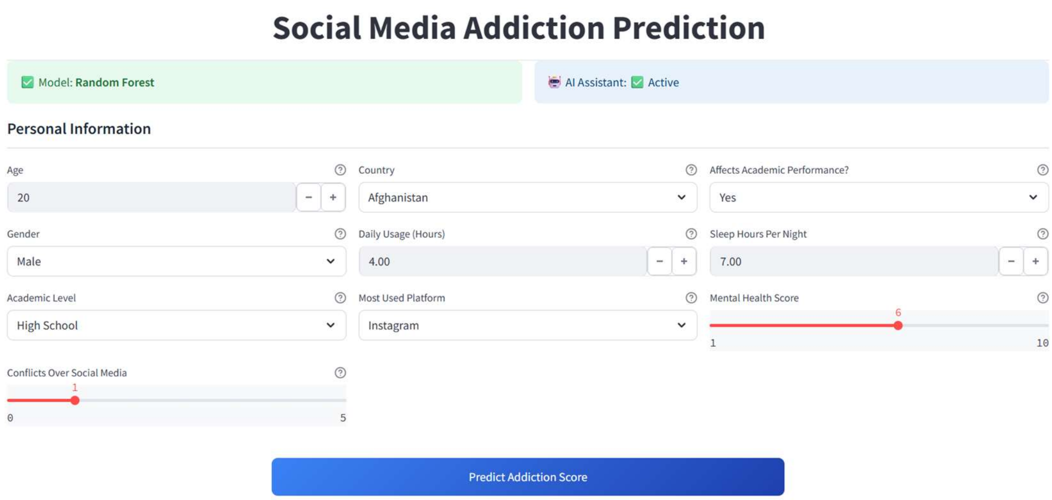

Fig. User Input Features

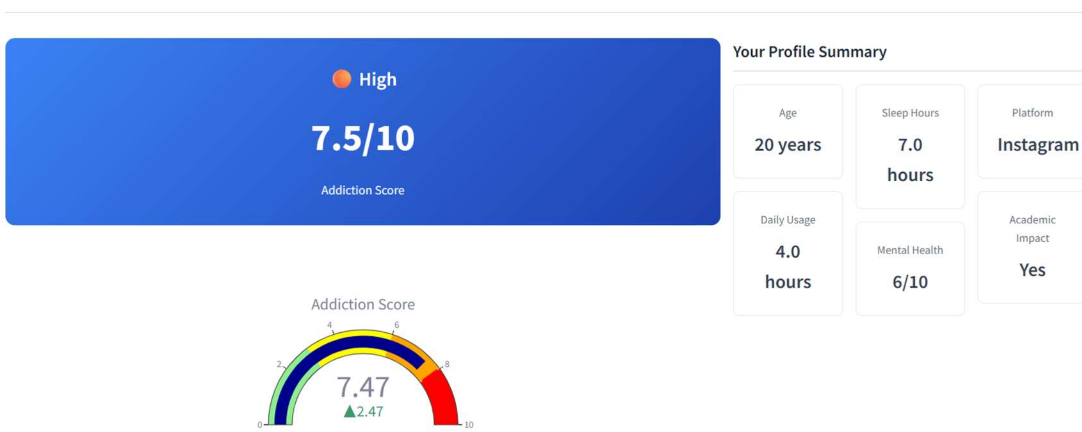

Fig. Predicted Addiction

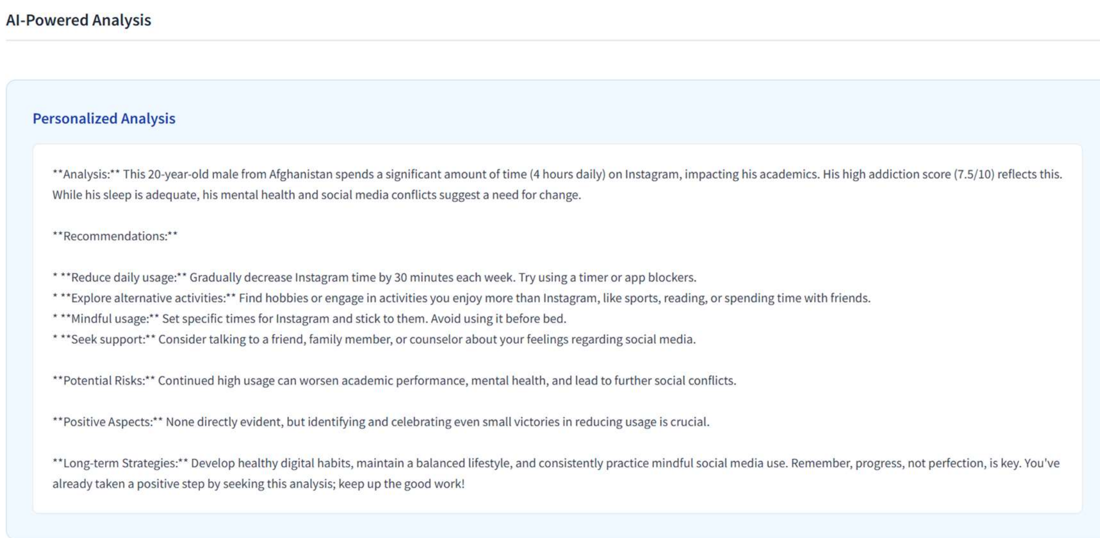

Fig. AI Assistance

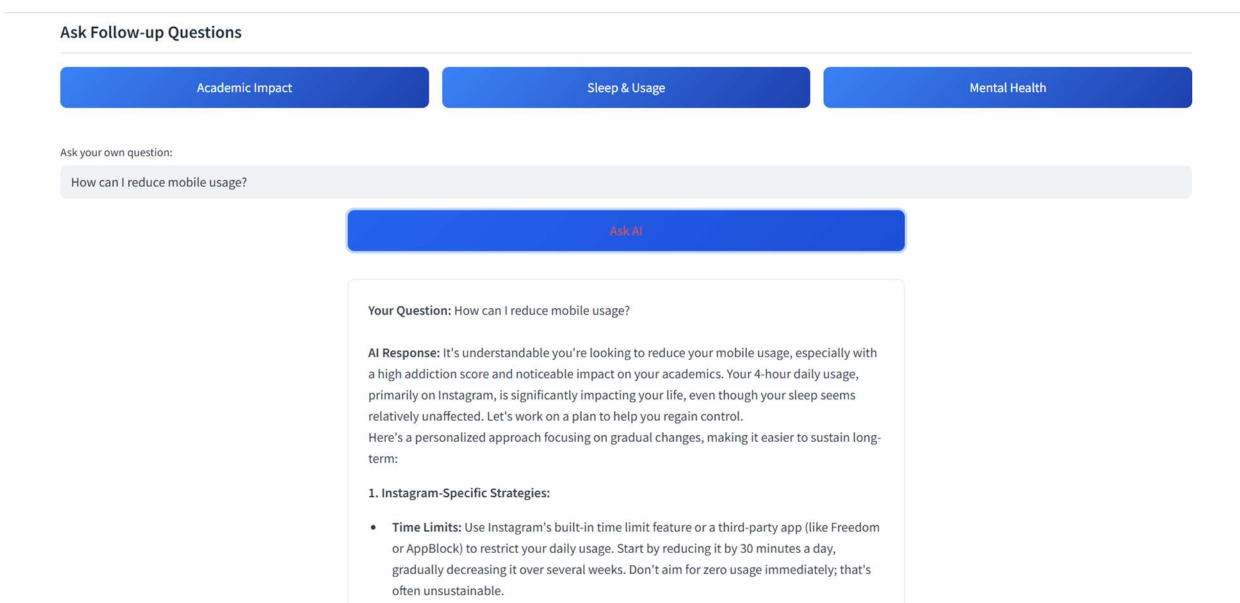

Fig. Follow-Up Questions

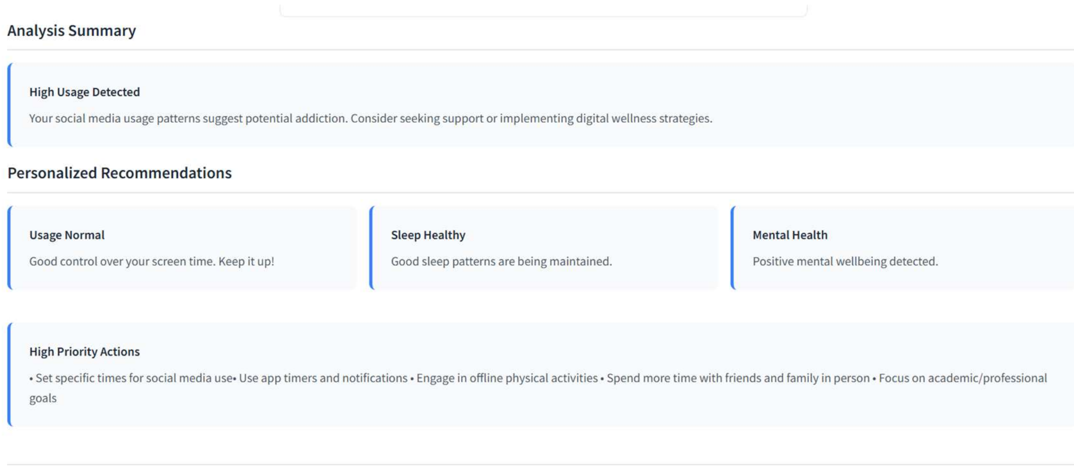

Fig. Summary and Recommendations

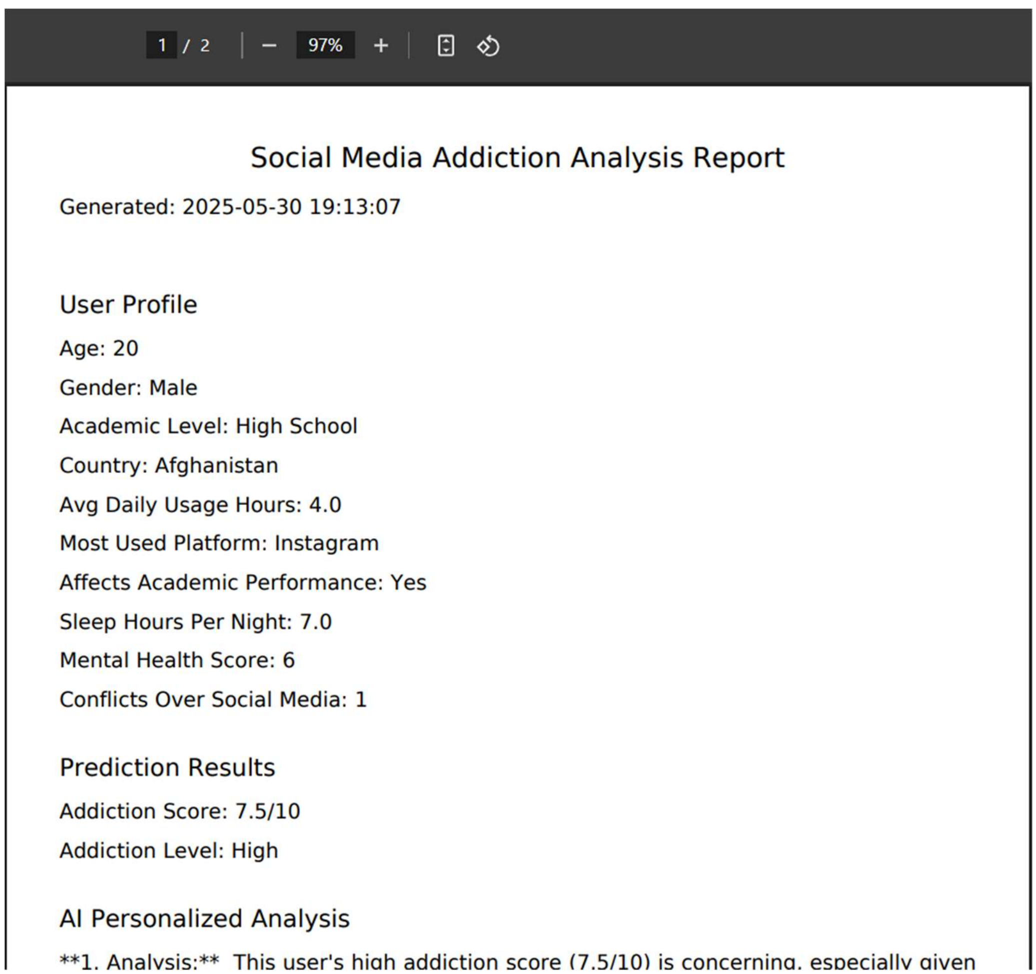

Fig. PDF Report

7. AI Integration (Google Gemini)

7.1 Gemini AI Implementation

Google's Gemini AI significantly enhances the system by enabling intelligent and personalized analysis capabilities. It supports features such as personalized recommendation generation, behavioral pattern analysis, and risk factor identification. The system also offers coping strategy suggestions tailored to individual needs and can respond effectively to follow-up questions. Furthermore, it provides contextual insights based on the user's profile, making the overall experience more intuitive and user-centric.

8. System Performance

The system is optimized for real-time performance, with prediction generation averaging less than two seconds and personalized AI analysis completed within five seconds. Comprehensive PDF reports are generated in under three seconds, ensuring quick access to insights. Memory usage is carefully optimized to minimize resource consumption, and the platform is designed to support scalability for concurrent user sessions. The caching strategy further enhances speed and responsiveness, with model loading handled via Streamlit's @st.cache_data decorator, static UI components cached for quicker rendering, and API responses fine-tuned to reduce latency.

9. Deployment & Configuration

The system supports flexible deployment options suitable for various environments. For local development, setup is streamlined with a simple pip install from the requirements.txt file, environment configuration through a .env file, and a one-command launch using streamlit run app.py. For production, the platform integrates with Streamlit Cloud and supports Docker containerization. It is compatible with major cloud platforms such as AWS, GCP, and Azure, and accommodates environment-specific configuration management. The configuration requirements include a Python 3.12+ runtime, a valid Gemini API key for AI functionality, and a minimum of 1GB RAM to ensure optimal performance.

10. Future Enhancements

Several future improvements and new features are planned to enhance the system's capabilities. On the technical side, there are plans to integrate deep learning models for improved accuracy, enable real-time data collection via API integrations, support multiple languages for wider accessibility, and implement advanced interactive visualizations. Mobile application development is also on the roadmap. In terms of feature additions, the system will support longitudinal tracking for progress monitoring, group analysis functionality tailored for educational institutions, integration with wearable devices for health data, gamification elements to boost user engagement, and community features to encourage peer support.

11. Conclusion

The Social Media Addiction Prediction System represents a significant advancement in digital wellness technology. By combining machine learning precision with AI-powered personalization, the system provides users with actionable insights into their social media consumption patterns.

Key Accomplishments:

• Successfully developed and deployed a production-ready machine learning application
• Integrated cutting-edge AI technology for enhanced user experience
• Created an accessible, user-friendly interface for complex data analysis
• Implemented robust security and privacy protection measures
• Established a foundation for future research and development

Impact and Value:

The system serves as both a diagnostic tool and educational platform, empowering users to make informed decisions about their digital consumption. By providing objective, data-driven insights, it bridges the gap between subjective self-assessment and professional intervention.