**Analyzing Social Media Usage Behavior & Predicting Addiction Risk Among Indian Users**

**🧠 Problem Statement:**

With the explosive rise in smartphone usage and social media consumption in India, there's a growing concern around behavioral addiction, attention span decline, and sleep disruption, especially among the younger population.

Despite increasing screen time and app engagement, **there’s limited understanding of how specific patterns like binge usage, session frequency, and content preference correlate with app addiction levels and lifestyle disruption**.

**📍 Key Questions:**

* Which demographic groups are most at risk for app addiction?
* What usage patterns indicate higher addiction levels?
* Can we predict addiction risk or distraction levels based on behavioral indicators?
* How does social media use impact sleep, productivity, and mental health?

**🎯 Project Objectives:**

1. **Create a comprehensive dataset** combining user demographics, app behavior, and psychological impact indicators.
2. **Identify key behavioral features** influencing high app addiction levels.
3. **Predict addiction risk scores** using ML models.
4. **Segment users** based on addiction risk and social validation needs.
5. **Suggest potential interventions** for high-risk users (e.g., screen time alerts, content type control).

**🔧 Tasks:**

|  |  |  |
| --- | --- | --- |
| # | Task | Description |
| 1 | **Data Cleaning & Preprocessing** | Handle missing values, invalid entries (e.g., negative ages), convert units (time formats), and standardize labels (like gender, region) |
| 2 | **Exploratory Data Analysis (EDA)** | Analyze patterns in age vs addiction, peak usage times, break frequency, etc. Use plots for correlation and outliers |
| 3 | **Feature Engineering** | Derive new fields like Total\_Weekly\_Screen\_Time, Engagement\_Rate, Addiction\_Index |
| 4 | **Clustering Users** | Use K-Means or DBSCAN to group users into behavior clusters (e.g., binge users, productive users, social seekers) |
| 5 | **Predictive Modeling** | Train ML models (Random Forest, XGBoost) to predict App\_Addiction\_Level, Distraction\_Level, or Sleep\_Disruption |
| 6 | **Behavioral Profiling** | Create user personas (e.g., Addicted Teen, Productive Professional, Casual User) using clustering insights |
| 7 | **Model Evaluation** | Use ROC, accuracy, F1-score, and AUC for classification models; RMSE for regression (if predicting continuous addiction risk) |
| 8 | **Deployment (Optional)** | Build a lightweight Streamlit dashboard for insights, predictions, or user profiling |

**🧮 Potential Features for Modeling:**

* **Numerical:** Age, Screen\_Time, Session\_Count, App\_Addiction\_Level, Engagement\_Level, Sleep\_Disruption
* **Categorical (One-Hot Encode):** Gender, Device\_Type, App\_Name, Region, App\_Purpose
* **Derived Features:**
  + Engagement\_Rate = Social\_Interaction / Session\_Count
  + Total\_Notifications\_Per\_Week
  + Active\_Use\_Percentage = Average\_Interaction\_Time / Time\_Spent\_Per\_Session

**🧩 Insight-Driven Questions (Grouped by Focus Area)**

**📌 1. User Demographics & App Usage Behavior**

* Which **age group** shows the highest screen time and session frequency?
* Does **gender or marital status** influence app addiction or sleep disruption?
* Is there a relationship between **income or education level** and app usage patterns?
* Do **retired or unemployed** individuals show higher social media engagement?

**📱 2. Device & Platform Trends**

* Do users on **Android vs iOS** differ in terms of addiction level or engagement?
* Is there a trend in usage patterns based on **App Download Source** (Google Play, Apple Store, Other)?
* Which **regions** (North, South, East, West) have the highest average screen time?

**🔄 3. Usage Patterns & Addiction**

* How does the **frequency of breaks** affect app addiction levels?
* Are users who use apps during **evening or night** more prone to sleep disruption?
* What is the correlation between **notifications received** and distraction levels?
* Can binge users be distinguished from regular users based on **session count** and **average interaction time**?

**🧠 4. Behavioral Impact**

* Is **sleep disruption** strongly correlated with app usage at night?
* How does **distraction level** vary across different occupational groups?
* Do users with higher **engagement levels** also report higher **social validation needs**?
* Is **emotional impact** more severe among users who use **multiple apps** frequently?

**🤖 5. Predictive Insights**

* Can we **predict app addiction risk** based on session count, screen time, and engagement?
* What are the top **5 features** most important in predicting Distraction\_Level or Sleep\_Disruption?
* Can clustering reveal hidden **user personas** like “Night Owl Addict” or “High-Engagement Professional”?
* Can we build a model to classify users into **Low**, **Moderate**, or **High Risk** for addiction?

**🧭 6. Engagement & Content Type**

* Which **content types** (videos, text posts, photos, stories) lead to the highest addiction?
* What’s the average interaction time per **content type** and does it vary by age?
* Does the **App Purpose** (Work, Socializing, Entertainment, etc.) influence addiction risk?

**🔐 7. App-Specific Analysis**

* Which apps (Instagram, Facebook, WhatsApp, TikTok, Twitter) are associated with **binge usage**?
* How does addiction risk vary across **different app versions** (Beta, v1.0, Latest)?
* Which app has the **highest screen time-to-engagement ratio**?