

Project Title	Decoding Phone Usage Patterns in India
Skills take away From This Project	Python scripting, data cleaning, Exploratory Data Analysis (EDA), Machine Learning, Multi class classification, Clustering, Streamlit,
Domain	User Behavior Analysis

Problem Statement

Design a system to analyze mobile device usage and user behavior by using a dataset containing user information and device statistics. The project aims to preprocess and clean the data, apply machine learning and clustering techniques, and build models to classify primary use and identify distinct usage patterns. The final application will be an interactive interface deployed with Streamlit, which will include EDA visualizations and model results.

Business Use Cases

1. **Behavioral Insights:** Provide insights into user behavior patterns based on device usage metrics.
2. **Device Optimization:** Assist manufacturers in optimizing device performance based on usage data.
3. **Personalized Services:** Enable businesses to personalize offerings based on identified user segments.
4. **Energy Efficiency:** Help users understand battery drain patterns and optimize device usage.

Approach

1. Data Preparation:

- Use a dataset containing user IDs, device models, operating systems, and usage statistics.
- Merge and preprocess the dataset to ensure consistency and accuracy.

2. Data Cleaning:

- Handle missing values using imputation techniques.
- Standardize formats for features like operating systems and device models.
- Remove outliers based on statistical thresholds.

3. Exploratory Data Analysis (EDA):

- Analyze trends in mobile app usage, screen-on time, and battery consumption.
- Visualize correlations between features like data usage and battery drain.
- Identify patterns in Primary use Class.

4. Machine Learning and Clustering:

- Implement classification models to predict **Primary use Class**:
 - Logistic Regression
 - Decision Trees
 - Random Forest
 - Gradient Boosting (e.g., XGBoost or LightGBM) etc.
- Apply clustering techniques to group users based on device usage patterns:
 - K-Means
 - Hierarchical Clustering
 - DBSCAN
 - Gaussian Mixture Models
 - Spectral Clustering etc.
- Evaluate classification models using metrics like precision, recall, and accuracy.
- Analyze clustering performance using silhouette scores and visualization techniques.

5. Application Development:

- Build a user-friendly interface using Streamlit to:
 - Display visualizations and insights from EDA.

- Allow users to input data for primary use classification.
- Present clustering results and user segmentation.

6. Deployment:

- Deploy the Streamlit application for accessibility and user interaction.

Data Flow and Architecture

1. Data Preparation:

- Clean and preprocess the dataset using Python libraries like Pandas and NumPy.

2. Processing Pipeline:

- Perform EDA and feature engineering for machine learning models.

3. Model Training:

- Train classification and clustering models using Scikit-learn, TensorFlow, or PyTorch.
- Save trained models for deployment.

4. Deployment:

- Develop a Streamlit-based interface.

Dataset

[phone_usage_india.csv](#)

Dataset Description

The dataset comprises user demographic and device usage statistics:

- **User ID:** Unique identifier for each user.
- **Age:** Age of the user.
- **Gender:** Gender of the user (Male, Female, Other).
- **Location:** City of the user.
- **Phone Brand:** Brand of the user's smartphone.
- **OS:** Operating system of the device (iOS or Android).
- **Screen Time (hrs/day):** Average daily screen-on time in hours.
- **Data Usage (GB/month):** Monthly data consumption in gigabytes.
- **Calls Duration (mins/day):** Average daily call duration in minutes.
- **Number of Apps Installed:** Total number of apps on the device.
- **Social Media Time (hrs/day):** Average daily time spent on social media in hours.

- **E-commerce Spend (INR/month):** Monthly expenditure on e-commerce platforms in INR.
- **Streaming Time (hrs/day):** Average daily time spent on streaming platforms in hours.
- **Gaming Time (hrs/day):** Average daily time spent on gaming in hours.
- **Monthly Recharge Cost (INR):** Monthly mobile recharge expenditure in INR.
- **Primary Use:** Primary purpose of the mobile device (Education, Gaming, Entertainment, Social Media).

Exploratory Data Analysis (EDA)

Key Analyses:

- Distribution of app usage time across different user classes.
- Relationship between screen-on time and battery consumption.
- Trends in data usage and app installations by age group.

Key Visualizations:

- Histograms for app usage time and battery consumption.
- Heatmaps to visualize feature correlations.
- Box plots to compare screen-on time across user classes.

Results

1. Cleaned and preprocessed dataset ready for modeling.
2. Insights from EDA visualizations and feature correlations.
3. Trained classification models to predict primary use.
4. Identified user clusters based on device usage patterns.
5. A Streamlit-based interactive application showcasing EDA and model results.

Project Evaluation Metrics

1. Data Cleaning Process: Evaluate how missing data, duplicates, and outliers are handled.
2. Model Performance: Assess classification accuracy and clustering quality.
3. Application Functionality: Ensure user interface is intuitive and responsive.
4. EDA Visualization: Verify the clarity and relevance of EDA results displayed in Streamlit.

Technical Tags:

Python, Machine Learning, Multi class classification, Clustering, User Behavior Analysis, Streamlit

Deliverables:

1. Data Preparation:

- Clean and merge the dataset for analysis.

2. Machine Learning Models:

- Train and evaluate classification and clustering models.

3. Application Development:


- Develop a Streamlit interface to display results and predictions, including EDA.

Timeline:

To be defined based on project milestones and deadlines.

References:

Project Live Evaluation Metrics	Project Live Evaluation
EDA Guide	Exploratory Data Analysis (EDA) Guide
Capstone Explanation Guideline	Capstone Explanation Guideline
GitHub Reference	How to Use GitHub.pptx
Streamlit recordings (English)	Special session for STREAMLIT(11/...

Streamlit recordings (Tamil)	https://us06web.zoom.us/rec/share/JTr7DywhE1-SarjylHBSCn4qnI7_uvJH6IGk06qAlkE0Ny1o_rqcq5FRFKuo93dm.iyM2o6l0h9aTUKNE
Streamlit documentation	Install Streamlit
Project Orientation (English)	 mobile_usage_orien_english.mp4
Project Orientation (Tamil)	 phone_usage_orien_tamil.mp4

PROJECT DOUBT CLARIFICATION SESSION (PROJECT AND CLASS DOUBTS)

About Session: The Project Doubt Clarification Session is a helpful resource for resolving questions and concerns about projects and class topics. It provides support in understanding project requirements, addressing code issues, and clarifying class concepts. The session aims to enhance comprehension and provide guidance to overcome challenges effectively.

Note: Book the slot at least before 12:00 Pm on the same day

Timing: Monday-Saturday (4:00PM to 5:00PM)

Booking link : <https://forms.gle/XC553oSbMJ2Gcfug9ss>

LIVE EVALUATION SESSION (CAPSTONE AND FINAL PROJECT)

About Session: The Live Evaluation Session for Capstone and Final Projects allows participants to showcase their projects and receive real-time feedback for improvement. It assesses project quality and provides an opportunity for discussion and evaluation.

Note: This form will Open only on Saturday (after 2 PM) and Sunday on Every Week

Timing:
For DS and AIML
Monday-Saturday (05:30PM to 07:00PM)

Booking link : <https://forms.gle/1m2Gsro41fLtZurRA>

Evaluation Metrics : [Project Live Evaluation](#)

Project Created By	Verified By	Approved By
Vinodhini	Nehlath Harmain	Nehlath Harmain

