



Course Project

Social Media & Smartphones Addiction



1. Project Description:

This project challenges students to apply data science methodologies to analyze different datasets. Students will explore, visualize, and model data using both classification and regression techniques. The project emphasizes the application of core data science principles, including exploratory data analysis, predictive modeling (regression and classification), and effective communication of insights.

2. Project Key Components:

2.1 Technical Requirements:

- **Data Visualization (MUST):** Create informative static and interactive visualizations to explore distributions, relationships, and trends within the dataset(s). This could involve boxplots, scatterplots, bar charts, or heatmaps to examine your data. **Dashboards** have to be used for the interactive visualization.
- **Regression Analysis (MUST):** Implement two regression models from lectures to predict key outcomes. Evaluate model performance using appropriate metrics (e.g., R-squared, RMSE).
- **Classification Analysis (MUST):** Use two classification models from lectures to categorize data. Evaluate performance using relevant metrics, and visualize results.
- **Model Expansion (BONUS):** Extend analysis by incorporating additional classification and/or regression models beyond those covered in lectures.
- **Communication (MUST):** Present findings through a technical report, a PowerPoint presentation, a poster, and a short video walkthrough.



2.2 Datasets:

Students are encouraged to preprocess the dataset(s), handle missing values, and ensure proper data formatting before analysis. **Students MUST select a dataset relevant to the given theme (Social media and smartphone addiction).**

Suggested datasets are available from:

- **Social Media Addiction SMA10 Dataset:**
<https://data.mendeley.com/datasets/9mxm455dfm/1>
- **Impact of Mobile Phones on Students:** <https://github.com/SyMasab/Impact-of-Mobile-Phones-on-Students>
- **Smartphone Usage and Behavioral Dataset:**
<https://www.kaggle.com/datasets/bhadramohit/smartphone-usage-and-behavioral-dataset/data>

Additional datasets are available from the following sources (ensure they are relevant to the theme of the project):

- **Kaggle:** <https://www.kaggle.com/datasets>
- **UCI Machine Learning Repository:** <https://archive.ics.uci.edu/ml/>
- **Google Dataset Search:** <https://datasetsearch.research.google.com/>

2.3 Deliverables:

- 2.3.1 **Code Repository (MUST to be eligible for discussion):** All Python code must be functional, well-documented, and tracked via GitHub.
- 2.3.2 **Dashboard Interface (MUST to be eligible for discussion):** A dynamic visualization dashboard showcasing key insights, is to be submitted as well.
- 2.3.3 **Technical Report (MUST to be eligible for discussion):** Your technical report should be a clear, concise, and well-organized document that effectively communicates your project's key aspects. For better organization, use headings, subheadings, and bullet points to structure the report logically. To enhance visual appeal and clarity, include relevant figures and tables to illustrate your findings and support your analysis. Make sure each visual is labelled clearly and includes captions to explain its relevance to your analysis. The report must include the following sections:
 - **Abstract:** A brief overview of your project, around 200 words. Its key elements are:
 - Briefly state the project's objective and the problem you addressed
 - Summarize the key findings and insights from your analysis.
 - Highlight the main conclusions and recommendations.
 - **Introduction:** It includes the following:
 - **Background and Context:** Provide background information on the topic and explain why it's relevant.



- **Literature Review:** Briefly review existing research and literature related to your project's theme.
- **Project Objectives:** Clearly state the specific goals and objectives of your project.
- **Materials and Methods:** It includes the following:
 - **Data Description:** Describe the dataset(s) used, including: Source of the data, Key variables and their types, Data size and structure, Any data cleaning or preprocessing steps performed
 - **Methodology:** Explain **ALL** methods and algorithms used for data analysis and modeling, including: Data visualization techniques, Regression models (with justification for their selection), Classification models (with justification for their selection), Model evaluation metrics.
- **Results:** It includes the following:
 - **Key Findings:** Present the main findings of your data analysis and modeling, including: Significant patterns and trends observed in the data, Results of regression analysis (e.g., coefficients, R-squared, RMSE), Results of classification analysis (e.g., accuracy, precision, recall, F1-score, confusion matrices)
 - **Visualizations:** Use various plots and visualizations to effectively communicate your results (e.g., scatter plots, histograms, box plots, ROC curves).
 - **Model Performance:** Clearly present and interpret the performance of your chosen models.
- **Conclusions:** It may include the following:
 - **Summary of Findings:** Concisely summarize the most important findings and insights from your analysis.
 - **Recommendations:** Provide actionable recommendations based on your findings.
 - **Limitations:** Discuss any limitations of your study or the data used.
 - **Future Work:** Suggest potential avenues for future research or extensions of the project.
- **Acknowledgements:** It includes the following:
 - **Team Contributions:** Acknowledge the contributions of each team member to the project.
 - **External Support:** Acknowledge any external support or resources used (e.g., libraries, datasets, mentors).
- **References:** List all cited sources (books, articles, websites, datasets) in a consistent citation format (e.g., APA, MLA).

2.3.4 PowerPoint Presentation: A clear and concise presentation visually communicating your project's key points.



- 2.3.5 Poster:** An engaging visual summary of the project suitable for a conference or exhibition setting.
- 2.3.6 Short Video Walkthrough (1-3 minutes):** A concise video explaining your project, its findings, any used packages, etc. The video must show how to run the code.

Students are encouraged to add bonus regression and/ or classification models of their choice (not discussed in lectures) for additional points.

The more features, the more bonus.

(Exact amount of bonus will be determined later).

Feel free to be creative with this!

3. General rules

- 3.1** Programming language used is Python.
- 3.2** A team can be formed from 4 to 6 students.

4. Evaluation Criteria:

- Correctness of code.
- Effectiveness and clarity of data analysis and visualization techniques.
- Justification for chosen models and performance of regression and classification models.
- Clarity, organization, and visual appeal of communication deliverables.

5. Deadline and late submission policy

- The deadline for project submission (including all deliverables) is **Saturday May 17th 2025** before **11:59 PM** (Cairo time zone).
- Late submissions are strictly prohibited.
- Tentative project discussion will be held on **Monday, May 26th 2025** and **Tuesday, May 27th 2025**.

Note 1:

All deliverables are to be submitted as a **single .zip** file named as “***Project_TeamLeaderID.zip***”, where “***TeamLeaderID***” is the EUI student ID of the leader of the team. Only one member of the team **has** to submit this .zip file on Canvas under the assignment section entitled “**Project**”.

Note 2

All members **MUST** attend the discussion, as you won't be graded unless you attend it.



6. Rubrics:

Deliverables (4 points)	Grading criteria
	Any missing deliverable will result in points being deducted from its equivalent weight in the project discussion grade. This means that even if you participate in the discussion, the absence of a deliverable will negatively impact your final grade.
Code + dashboard	<ul style="list-style-type: none">• Python code files are delivered and in correct format.• Code runs without error and fulfills all the requirements.• Excellent use of white space (indentation), naming of variables, functions, (unambiguous naming), etc.• Clear comments showing the purpose of each function or module, etc.• All codes including <u>dashboard part</u> have to be submitted on both Canvas and GitHub. All source codes must be tracked by GitHub.
Technical report (.doc)	<ul style="list-style-type: none">• Report must include the following sections: (abstract, introduction, materials and methods, results, conclusions, acknowledgments, references).
Presentation (.ppt)	<ul style="list-style-type: none">• Explanation of used data, methods, etc.• Individual contributions of each team member.• All references used must be included.• University and Faculty logos are included.• Readable presentation using appropriate fonts, size, colors.• Clear and organized presentation.
One-page Poster (.ppt)	<ul style="list-style-type: none">• One-page poster.• Clear and concise version of the presentation, keeping only necessary information.• All references used must be included.• University and Faculty logos are included.• Readable poster using appropriate fonts, size, colors.
Walkthrough video	<ul style="list-style-type: none">• Clear walkthrough video that showcase the functionality of your project (how to run it and how to use it).• Duration: 1 – 3 mins
Discussion (16 points)	Grading criteria
Discussion (oral presentation)	<ul style="list-style-type: none">• All team members <u>must</u> attend the discussion.• All team members <u>must</u> be involved in presenting their work.• Each student <u>should individually</u> demonstrate a comprehensive understanding of <u>the entire implementation, details, and concepts (including the bonus)</u>.