A305 - Introduction to Machine Learning

Project Term

Project Overview:

The AI305 project is designed to apply machine learning techniques to real-world tasks. The project is structured into three cumulative phases, meaning each phase builds on the work done in the previous one. Students must complete and submit reports, presentations and demos for each phase, with the final submission combining all previous work into a comprehensive machine learning solution.
Dataset for the project: https://www.kaggle.com/competitions/isic-2024-challenge/data

Project Phases

Phase 1: Project Proposal

- **Description**: Form groups of **3-4 students** and submit a project proposal detailing the problem, dataset, and initial approach to solving it using machine learning. This phase sets the foundation for all subsequent work.
- **Deliverables**:
    - A proposal document that includes:
        - **Problem Definition**: Describe the problem and why it is significant.
        - **Dataset**: Identify the dataset you'll use and why it's appropriate for the problem.
        - **Planned Algorithms**: List and justify the machine learning algorithms you plan to apply.
        - **Cover Page**: Include a signed pledge of originality.
        - **Contributions**: A section describing each group member's role in the project.
    - A short group presentation summarizing your proposal.
- **Submission Deadline**: **October 25th.**
- **Dataset for the project: https://www.kaggle.com/competitions/isic-2024-challenge/data**

Phase 2: Data Preparation, Visualization, and Preliminary Model Implementation

- **Description**: Based on the proposal, you must now prepare and preprocess the dataset, perform data visualization, and implement initial machine learning models.
- **Deliverables**:

    - A report detailing:
        - **Data Preparation**: Describe how the dataset was preprocessed (e.g., cleaning, normalization, feature selection).
        - **Visualizations**: Provide insights and findings from data exploration.
        - **Initial Models**: Implement at least two models and evaluate their performance.
        - **Contributions**: Specify each member's contribution.
        - **Cover Page**: Include a signed pledge of originality.
    - A group presentation showing the findings, initial models, and evaluation.
- **Submission Deadline**: November 8th.

Phase 3: Final Submission – Model Optimization and Full Project Report

- **Description**: This final phase builds upon the previous phases, where you will optimize the models, refine the results, and complete the project. This phase should show how your model evolved through experimentation and refinement, ending in a final comprehensive solution.
- **Deliverables**:
    - **Final Report**: A full report summarizing the entire project. The report structure should include:
        1. **Cover Page**: Signed pledge of originality.
        2. **Contributions**: Describe each group member's role.
        3. **Motivation**: Define the problem you are solving and its relevance.
        4. **Introduction**: Provide context and background.
        5. **Methodology**: Explain the machine learning techniques applied.
        6. **Dataset**: Describe the dataset and the preprocessing steps.
        7. **Experiments**: Detail the models used, experiments run, and testing procedures.
        8. **Evaluation**: Compare the performance of different models.
        9. **Discussion**: Reflect on the results, challenges faced, and the improvements made.
        10. **Conclusion**: Summarize your findings and propose future work.
    - **Source Code**: Submit all source code in a zip file (well-commented).
    - **Final Presentation & Demo**: A final group presentation to summarize the project and a live demo to showcase the results.
- **Submission Deadline**: **December 7th.**

Key Requirements

- **Cumulative Work**: Each phase builds upon the previous phase. For example, Phase 2's models must be based on the dataset identified and processed in Phase 1. Phase 3 should refine and optimize the models created in Phase 2.
- **Multiple Algorithms**: Implement and compare at least **two algorithms** based on the dataset's nature. Results must be compared using appropriate evaluation metrics.

- **Originality**: Do not use code or algorithms directly applied to the same dataset from public sources. Declare all external resources used, or face penalties for undisclosed usage.

Submission and Presentation

For each phase:

1. **Report**: Submit a detailed report covering the work done in that phase. Include a **Cover Page** with a signed pledge and a **Contributions** section detailing each member's work.
2. **Presentation**: Prepare a presentation to explain the methods and results for each phase.
3. **Source Code**: Submit the corresponding source code in a zip file.

Rules

1. The project must be completed by groups of **3-4 students**.
2. A signed pledge of originality must be included in each phase report.
3. **Group Contributions**: Each report must include a section specifying the contributions of each group member.
4. All resources must be listed, and external code must be declared.
5. Each group will give a presentation and demo following the final submission.
6. All the group members must understand each part in the code, report and presentation and be able to present any part.
7. **No Public Code**: The use of public code applied to the same dataset is prohibited unless properly declared. Penalties will be applied, up to a **zero grade** for the project.
8. **Late Submissions**: A 5% penalty will be applied for each day the submission is late.

Final Evaluation Criteria

Projects will be evaluated based on the following:

- **Technical Quality**: Depth, correctness, and implementation of the machine learning techniques used. The models should be appropriate for the problem and demonstrate the students' ability to apply theory to practice.
- **Cumulative Development**: Clear progression across all phases, where each phase builds upon the previous one (from proposal, data preparation to model optimization). The final submission should reflect refinement and improvement over time.
- **Originality**: Whether the team introduced original methods or extended beyond publicly available solutions. The project must avoid replicating existing work without significant contributions.
- **Model Comparison**: The quality and depth of analysis when comparing multiple algorithms. Proper use of evaluation metrics to demonstrate the strengths and weaknesses of each model.
- **Comprehension of Processes and Components**: The team’s understanding of all project components (data preparation, model selection, evaluation, error analysis, etc.)

must be evident in the report, presentations and demo. Each member must demonstrate a clear understanding of the project's processes.

- **Communication**: Clarity and structure of reports and presentations. The team must effectively communicate their methods, findings, and insights. Presentations should be concise but thorough, and the final report must be comprehensive.
- **Timeliness**: Submitting all deliverables on time.
- **Teamwork**: Effective collaboration with clear task division and contributions from all members.

Deadlines

- **Phase 1: October 25th.**
- **Phase 2: November 8th.**
- **Final Submission: December 7th.**
- **Demo and Presentation (within next week of project final's submission)**