**DSAI 3201 Machine Learning Project: Indoor Localization Using WLAN Fingerprinting**

**Project Overview**

This project challenges teams to develop an indoor localization system that predicts a user’s location (building, floor, and geographical coordinates) based solely on WiFi fingerprint data. Students will work with the UCI Indoor Localization WiFi Dataset—a collection of WiFi signal strength (RSSI) readings gathered from hundreds of access points in multiple buildings—to design and evaluate machine learning models that solve both classification and regression tasks.

**Team Assignment**

Students will be organized into teams of two through random assignment. If there is an odd number of students, the remaining student will be randomly assigned to one of the existing teams.

**Note on the Use of AI Tools:**

The use of ChatGPT or other AI tools should be kept to a minimum throughout this project. Any use of AI tools must be clearly disclaimed at the beginning of your project documentation. Additionally, include a separate section in your final report explaining how you utilized AI tools—both individually and as part of your group efforts.

Please be aware that each student will be individually interviewed about their contributions and understanding of the project during the report submission and presentation sessions. A significant portion of your project grade will be based on the success of these interviews, so it is essential that you can clearly articulate your work and the methodologies you employed without relying solely on AI-generated content.

**Due Dates and Schedule**

* **13th March:** 1st Individual Report
* **3rd April:** 2nd Individual Report and Final Group Report Due
* **6th April – 10th April:** Project Presentations

**Project Objectives and Tasks**

1. **Data Understanding and Preprocessing**
   * **Learn WiFi Fingerprinting Concepts:** Study the basics of indoor localization and the role of RSSI in creating unique “fingerprints” for different indoor areas.
   * **Dataset Exploration:** Examine the structure and features of the UCI Indoor Localization WiFi Dataset (520 RSSI features, along with building, floor, and coordinate information).
   * **Data Cleaning and Preparation:**
     + Handle any missing values and normalize the RSSI measurements.
     + Scale features appropriately and encode any categorical variables if needed.
     + Optionally, apply feature reduction techniques (e.g., PCA) to reduce noise and improve model efficiency.
2. **Model Development**  
   The modeling phase is divided into two main tasks—a classification task for building and floor prediction and a regression task for predicting exact geographical coordinates (longitude and latitude).
   * **Basic Models:**
     + *Classification:* Develop a simple machine learning model (e.g., Decision Tree or SVM) to predict the Building ID and Floor based on WiFi fingerprints.
     + *Regression:* Build a baseline regression model (e.g., Linear Regression) to predict the exact coordinates.
   * **Advanced Models:**
     + *Classification:* Implement a deep learning model (e.g., a Convolutional Neural Network) to improve building and floor prediction accuracy.
     + *Regression:* Develop an advanced regression model (e.g., a neural network-based regressor) to enhance coordinate estimation.
   * **Work Distribution:**
     + Each team member is responsible for one basic model and one advanced model, ensuring that both classification and regression tasks are covered by every team.
3. **Model Evaluation and Optimization**
   * **Evaluation Metrics:**
     + **Classification:** Accuracy, precision, recall, and F1-score.
     + **Regression:** RMSE (Root Mean Squared Error) and MAE (Mean Absolute Error).
   * **Optimization:**
     + Apply hyperparameter tuning (e.g., using GridSearchCV) and cross-validation to improve model performance and robustness.
4. **Applied Analysis and Interpretation**
   * **Practical Implications:**
     + Analyze model performance in the context of real-world indoor navigation challenges.
     + Discuss applications such as navigation in large facilities (e.g., airports, malls), emergency response, or asset tracking.
   * **Stakeholder Relevance:**
     + Explain how your models’ outputs support the needs of stakeholders like facility managers, end-users, and technology developers.
   * **Documentation and Reporting:**
     + Document every stage—from data preprocessing to model deployment considerations—in a comprehensive final report.

**Project Deliverables**

1. **Individual Report 1: Exploratory Data Analysis (EDA) and General Project Understanding**  
   *Due Date: 13th March*
   * Provide an in-depth exploration of the UCI Indoor Localization WiFi Dataset with visualizations (histograms, scatter plots, correlation matrices) and descriptive statistics.
   * Summarize key findings from the data analysis, such as feature distributions, potential challenges, and insights.
   * Explain your understanding of the indoor localization problem and WiFi fingerprinting principles.
   * Outline initial ideas on potential modeling approaches.
   * Format your submission using the ACM standardized template, which you can find here: <https://www.acm.org/binaries/content/assets/publications/taps/acm_submission_template.docx>  
     **Rubric for Report 1:**
   * **Data Exploration & Visualization (40%):** Quality and clarity of plots, summary statistics, and discussion of dataset characteristics.
   * **Conceptual Understanding (30%):** Clear explanation of indoor localization, WiFi fingerprinting, and project goals.
   * **Clarity & Organization (20%):** Logical structure, clear language, and well-organized sections.
   * **Technical Accuracy (10%):** Correct interpretation of the data and relevant methodologies.
2. **Individual Report 2: Detailed Individual Reflection and Explanation of Your Work**  
   *Due Date: 3rd April (along with Final Group Report)*
   * Describe your specific contributions to the project.
   * Reflect on the challenges you encountered and how you addressed them.
   * Explain the technical details of your work (e.g., preprocessing choices, model selection, and optimization strategies).
   * Discuss the lessons learned and insights gained during the project.
   * Format your submission using the ACM standardized template, which you can find here: <https://www.acm.org/binaries/content/assets/publications/taps/acm_submission_template.docx>
   * **Rubric for Report 2:**
   * **Depth of Technical Detail (40%):** Clear explanation of methods, algorithms, and personal contributions.
   * **Reflection & Critical Analysis (30%):** Thoughtful discussion on challenges, decision-making, and learning outcomes.
   * **Clarity & Organization (20%):** Well-structured report with logical flow and clear writing.
   * **Supporting Evidence (10%):** Use of examples, figures, or code snippets to support explanations.
3. **Final Group Report: Comprehensive Project Report**  
   *Due Date: 3rd April*  
   The final report should cover the entire project lifecycle and include:
   * **Introduction:** Overview of the indoor localization problem, the significance of WiFi fingerprinting, and project objectives.
   * **Data Preprocessing:** Detailed description of data cleaning, normalization, feature reduction (e.g., PCA), and exploratory analysis.
   * **Model Development:** Documentation of both basic and advanced models for classification and regression, including the rationale behind model choices and optimization strategies.
   * **Model Evaluation:** Presentation of evaluation metrics and discussion of model performance.
   * **Applied Analysis:** Interpretation of how your model results address real-world challenges and support stakeholder needs.
   * **Conclusion:** Summary of key findings, limitations, and suggestions for future work.
   * **Appendices (if necessary):** Additional figures, tables, or code snippets.
   * Format your submission using the ACM standardized template, which you can find here: <https://www.acm.org/binaries/content/assets/publications/taps/acm_submission_template.docx>  
     **Rubric for Final Group Report:**
   * **Content and Completeness (40%):** Inclusion of all required sections with sufficient depth; integration of individual contributions into a cohesive report.
   * **Technical Quality (25%):** Accuracy and thoroughness in data preprocessing, model development, and evaluation.
   * **Analysis and Interpretation (20%):** Insightful discussion on practical implications and stakeholder relevance.
   * **Organization and Clarity (10%):** Clear structure, logical flow, and quality writing with well-formatted figures/tables.
   * **Presentation and Professionalism (5%):** Adherence to formatting guidelines, proper citations, and overall report polish.
4. **Project Presentation**  
   *Presentation Dates: 6th April – 10th April*
   * Deliver a concise, engaging presentation summarizing:
     + The problem statement and project objectives.
     + Key findings from the EDA, model development, and evaluation.
     + Practical implications and potential applications.
     + Lessons learned and future improvements.
   * The presentation should be professionally prepared, use visual aids effectively, and be well-coordinated among team members.

**DSAI3201 Rubric for Project Assessment**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **KPI** | **Assessment** | **Unsatisfactory (0-25)** | **Developing (25-59)** | **Satisfactory (60-84)** | **Exemplary (85-100)** |
| **Apply relevant theories, techniques, and tools according to the requirements of the data science lifecycle (50% of Project Grade)** | Based on the given case study and dataset, select and apply the most suitable machine learning theories, techniques, and tools throughout the data science lifecycle to address the problem. | Fails to select appropriate theories, techniques, or tools of the data science lifecycle, preventing their application to the problem. | Identifies some relevant theories, techniques, or tools of the data science lifecycle but does not apply them successfully to the problem. | Selects and applies relevant theories, techniques, and tools of the data science lifecycle that adequately address most of the problem’s requirements. | Selects and applies relevant theories, techniques, and tools of the data science lifecycle that fully address all the problem’s requirements. |
| **Analyze the results of applied theories, techniques, and tools according to stakeholders' needs (50% of Project Grade)** | Analyze the effectiveness of the applied theories, techniques, and tools to determine how the results meet the problem's requirements and stakeholders' needs. | Fails to analyze the effectiveness of the applied methods, missing how they meet the problem’s requirements. | Provides limited analysis on the effectiveness of the applied methods, meeting only some of the problem’s requirements. | Provides a complete analysis on the effectiveness of the applied methods, meeting the majority of the problem’s requirements. | Provides a comprehensive analysis on the effectiveness of the applied methods, fully meeting all the problem’s requirements. |