**Project Proposal**

**Candidate Number:**

246576

**Supervisor's Name:**

Peter Wijeratne

**Working Title:**

**Machine Learning-Based Star Classification Using GAIA Satellite Data**

**Aims and Objectives**

**Aims:**

The purpose of this project is to develop a machine learning model capable of classifying stars based on properties derived from GAIA satellite data. My motivation for undertaking this topic stems from a passion for astrophysics and data science, something I hope to take into the future. I want to develop and compare multiple machine learning approaches, whilst also seeing if my methods are comparable to traditional ways to handle this data.

**Primary Objectives:**

1. **Research and Evaluate:**
   * Research GAIA to understand the full picture of the data it collects and where it can be accessed. Gain a full understanding of the topic I am building my project around.
2. **Compare Available Techniques:**
   * Analyse and compare existing star classification techniques, including those that do not rely on machine learning. This will provide a baseline for comparing the effectiveness of machine learning models against traditional approaches.
3. **Data Extraction and Preparation:**
   * Extract relevant features (luminosity, colour indices, effective temperatures, parallax) from GAIA data.
   * Handle missing values and outliers effectively.
4. **Data source and storage**
   * Find the most suitable way to store the vast amounts of data this project requires; I will look into solutions such as a database or data lake.
5. **Model Implementation:**
   * Implement and compare classification algorithms, specifically Decision Trees, Random Forest, Support Vector Machines (SVM), and Neural Networks.
6. **Data Splitting:**
   * Split the dataset into training, validation, and test sets to ensure robust evaluation.
7. **Model Evaluation:**
   * Evaluate the models using metrics such as accuracy, precision, recall, and F1-score.
8. **Conduct User Research and Interviews**
   * Interview a wide range of both peers and members of the public to understand what is expected of the user interface.
9. **User Interface Development:**
   * Create a user-friendly interface that allows users to input star data and receive classification results.
10. **Testing**
    * Develop a comprehensive test plan and conduct testing with each iteration of code, this will include, but not be limited to, unit and user testing.

**Extensions:**

1. Implement advanced feature selection techniques to enhance model performance.
2. Investigate the potential for real-time data classification.
3. Develop interactive visualisations for better understanding of results.
4. Use the fully developed models to predict future data.

**Relevance**

This project will summarise the last three years of my degree, requiring a combination of many aspects of computer science, focusing on machine learning and data science. As a subset of AI, machine learning is a branch of computer science that is rapidly growing and expanding into every detail of everyday life. It also utilises my learning from my studies abroad last year where I took a course on data science in astronomy.

**Resources Required**

* Access to GAIA satellite data (available through [specify source]
* Python programming environment with libraries such as pandas, scikit-learn and Matplotlib for data analysis and visualisation.
* A computer with adequate processing power to handle the machine learning task

**Timetable**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
|  | **Monday** | **Tuesday** | **Wednesday** | **Thursday** | **Friday** | **Saturday** | **Sunday** |
| **9 – 10** |  |  |  |  |  |  |  |
| **10 – 11** | Intro to Computer Security |  | Intro to Computer Security |  |  |  |  |
| **11 – 12** |  |  |  |  |  |  |  |
| **12 – 1** |  |  |  | Human-Computer Interaction |  |  |  |
| **1 – 2** |  | Project Lecture |  |  |  |  |  |
| **2 – 3** | Comparative Programming Lab |  |  |  |  |  |  |
| **3 - 4** | Human-Computer interaction |  |  | Intro to Computer Security Lab | Compa |  |  |
| **4 - 5** |  |  |  |  |  |  |  |
| **5 - 6** |  |  |  |  |  |  |  |

**Detailed Bibliography**

**Interim Log**

**Meetings with Supervisor:**

**Material Consulted:**