Project Initialization and Planning Phase

Date	15 July 2024
Team ID	739984
Project Title	Genetic Classification Of An Individual By Using Machine Learning
Maximum Marks	3 Marks

Project Proposal (Proposed Solution) template

Genetic classification involves categorizing individuals based on their genetic makeup. This classification can be used for various purposes, including medical diagnostics, ancestry tracking, and personalized medicine. Machine learning (ML) offers a powerful set of tools for analyzing large datasets and identifying patterns that are not easily discernible through traditional statistical methods. By applying ML to genetic data, we can automate and enhance the classification process, leading to more accurate and efficient outcomes.

Project Overview	
Objective	Develop a machine learning model to classify individuals based on genetic data, enhancing accuracy and efficiency in genetic classification.
Scope	•Data collection and preprocessing • Feature extraction • Model development and training • Validation, testing, and deployment • Documentation and results presentation
Problem Statement	
Description	Current genetic classification methods are manual and error-prone. An automated ML model can improve speed and accuracy
Impact	Facilitates personalized medicine, aids genetic research, reduces manual workload for geneticists.
Proposed Solution	
Approach	Collect and preprocess genetic data. Extract relevant genetic markers. Develop and train ML model (Random Forest) Validate and test models. Deploy the best-performing model.

	Monitor and evaluate performance.
Key Features	Automated classification High accuracy Scalability User-friendly interface

Resource Requirements

Resource Type	Description	Specification/Allocation
Hardware		
Computing Resources	CPU/GPU specifications, number of cores	e.g., 2 x NVIDIA V100 GPUs
Memory	RAM specifications	e.g., 8 GB
Storage	Disk space for data, models, and logs	e.g., 1 TB SSD
Software		
Frameworks	Python frameworks	e.g., Flask
Libraries	Additional libraries	e.g., scikit-learn, pandas, numpy
Development Environment	IDE, version control	e.g., Anaconda prompt, Jupyter Notebook, Git
Data		
Data	Source, size, format	e.g., Kaggle dataset, 10,000 images