Project Design Phase
Data Flow Diagram and User Stories

Date: -	23th October, 2023
Team ID: -	Team-593068
Project Name: -	Genetic Classification of Individuals using Machine Learning
Maximum Marks: -	4 marks

Report:
Introduction:
The genetic variant classification project aims to enhance the understanding and categorization of genetic variants, thereby enabling more accurate disease prognosis and personalized treatment strategies. The data flow diagram illustrates the movement of data through various stages of processing and analysis in the project.

Data Collection:
The process begins with the collection of diverse datasets containing genetic variant information and corresponding clinical annotations. These datasets are sourced from various repositories and clinical databases, providing a comprehensive foundation for the subsequent stages of analysis.

Preprocessing and Feature Engineering:
The collected data undergoes preprocessing, involving the cleansing of datasets, handling missing values, and standardizing data formats. Then, relevant features are extracted from genetic data and clinical annotations, forming a well-organized and structured feature set for further analysis.

Model training and evaluation:
The pre processed data is then used to train a machine learning model that leverages features to accurately classify genetic variants. The model is rigorously evaluated using performance metrics to ensure its robustness and reliability in variant classification tasks.

Scoring system integration:
A custom scoring system is integrated into the classification process, incorporating relevant clinical annotations and genetic variant information. This scoring system improves the accuracy

and reliability of the overall classification process, allowing for more precise and nuanced classification of genetic variants.

Disease prediction and real-time analysis:
The trained model is used to predict diseases based on genetic variant classification. Real-time analytics capabilities are implemented to enable dynamic adjustments and updates to the classification system, ensuring it can adapt to changing genetic data and clinical information.

Conclusion:
The data flow diagram provides a comprehensive overview of the complex processes involved in the genetic variation classification project. It emphasizes the continuous flow of data across different stages, ultimately contributing to improved disease prognosis and appropriate treatment recommendations based on accurate classification of genetic variants.

Data Flow Diagram: -
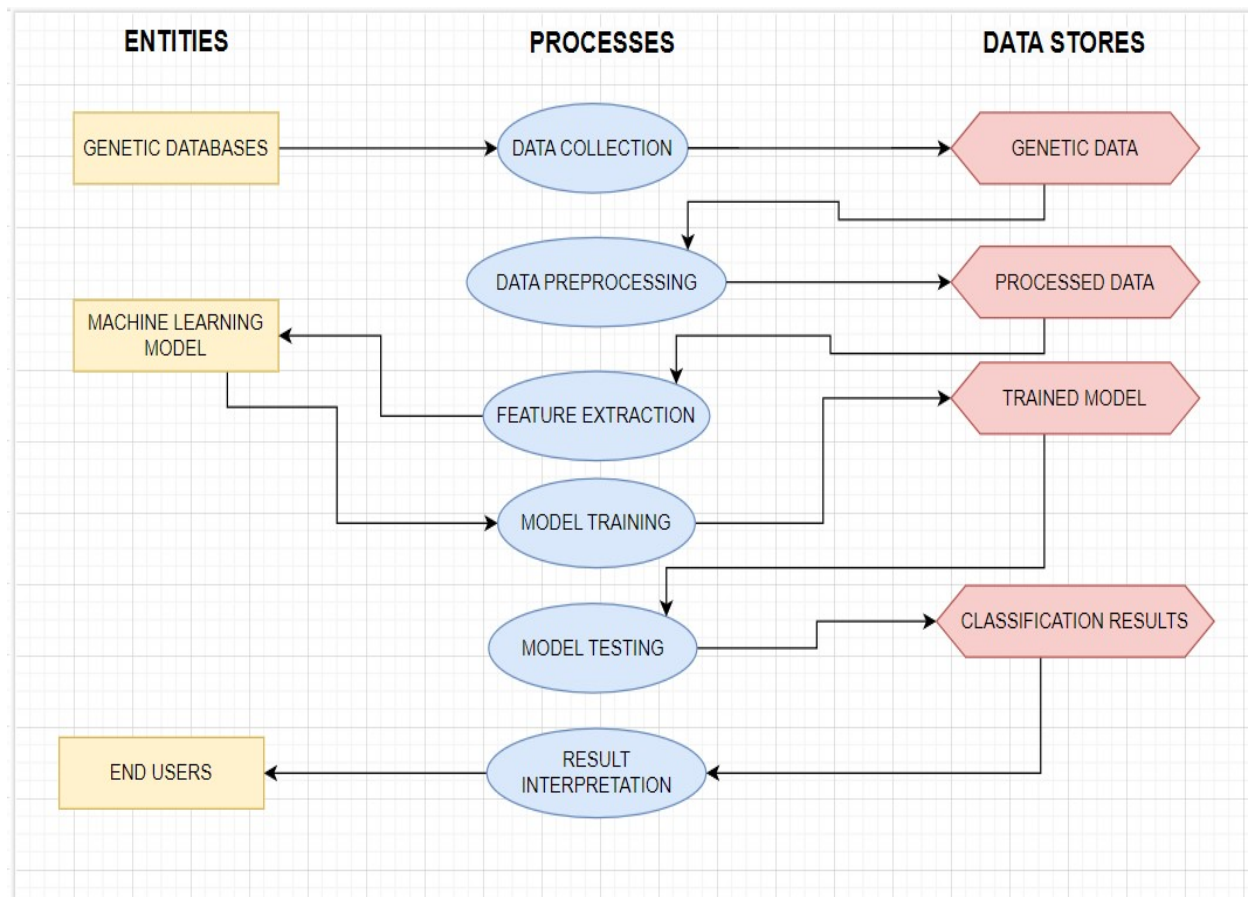
User Stories: -

User Type	Functional Requirement (Epic)	User Story Number	User Story/Task	Acceptance Criteria	Priority	Release
Clinical Laboratories	Setup & Infrastructure	USN-1	Configure the development environment with necessary libraries and frameworks for genetic variant classification.	Development environment is successfully set up with all required tools and frameworks.	High	Sprint 1
Genetic Researchers	Data Collection	USN-2	Gather a diverse dataset of genetic variants, including associated clinical data, for training and testing the machine learning models.	A comprehensive dataset comprising a range of genetic variants and relevant clinical data is obtained.	High	Sprint 1
General Public	User Interface Development	USN-3	Design an intuitive and user-friendly interface that allows the general public to access basic information about genetic variants and their potential health implications.	The user interface is visually appealing, easy to navigate, and provides understandable information about genetic variants.	Medium	Sprint 2
Industry Stakeholders	Product Integration	USN-4	Integrate the genetic variant classification tool into existing	The tool is successfully integrated, and it seamlessly	Medium	Sprint 3

			industry systems, ensuring compatibility and seamless operation.	interacts with the industry's existing infrastructure.		