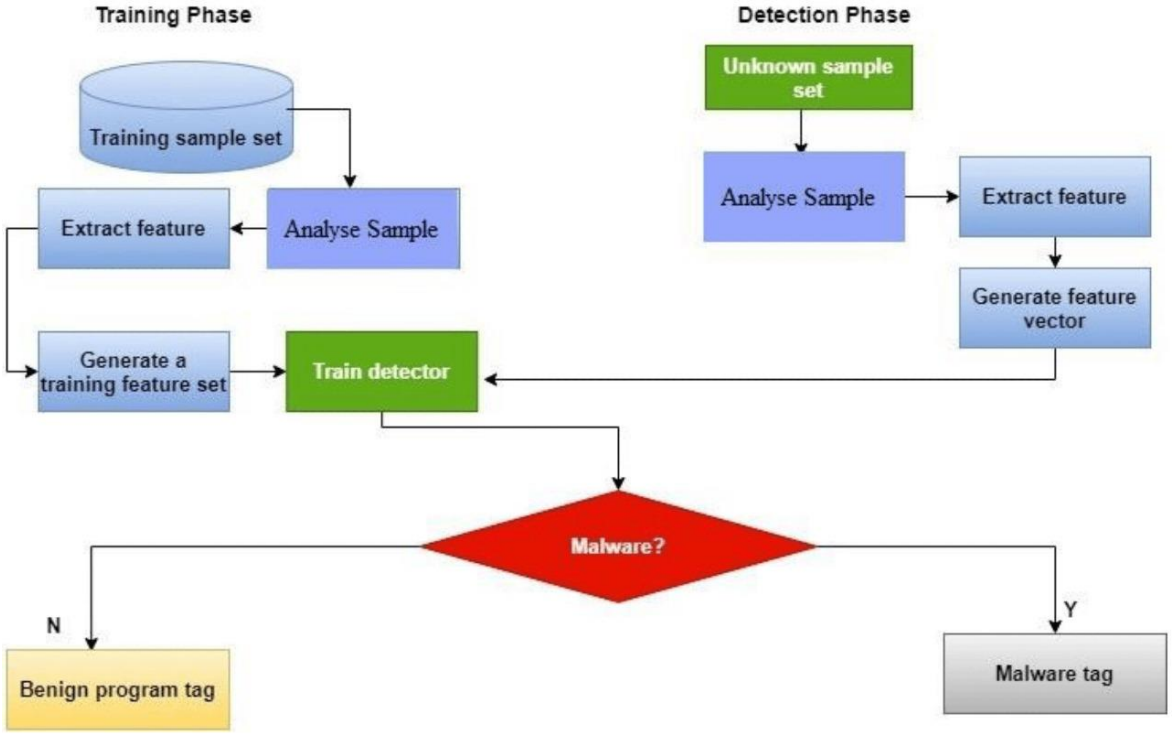
Project Design Phase-II Data
Flow Diagram & User Stories

Date	27-10-2023
Team ID	2.3
Project Name	Malware detection and classification
Maximum Marks	4 Marks

Data Flow Diagrams:

A Data Flow Diagram (DFD) is a traditional visual representation of the information flows within a system. A neat and clear DFD can depict the right amount of the system requirement graphically. It shows how data enters and leaves the system, what changes the information, and where data is stored.



User Stories

User Type	Functional Requirement (Epic)	User Story Number	User Story / Task	Acceptance criteria	Priority	Release
Software based companies	Project setup & Infrastructure	USN-1	Set up a robust development environment with the necessary tools and frameworks to initiate the project for malware detection and classification.	successfully configured with all necessary tools and frameworks	High	Sprint 1
Government organisations	development environment	USN-2	Acquire a diverse dataset of malicious software samples, encompassing various types of malware such as viruses, trojans, and worms. This dataset is essential for training our machine learning model.	Gathered a diverse dataset of files depicting various types of malware	High	Sprint 1
Households and Individuals	Data collection	USN-3	Preprocess the collected malware dataset by extracting pertinent features, applying appropriate labels, and dividing it into training and testing subsets.	preprocessed the dataset	High	Sprint 2
Developers and testers	data preprocessing	USN-4	Investigate and assess a variety of machine learning and deep learning algorithms, including options such as Random Forest and Convolutional Neural Networks (CNNs). The objective is to determine the most suitable model for malware detection and classification.	we could explore various DL models	High	Sprint 2
Non-Governmental Organizations (NGOs)	model development	USN-5	Train the chosen machine learning model with the preprocessed dataset, and closely monitor its performance on the testing set to ensure accuracy and effectiveness.	we could do validation	High	Sprint 3
Educational Institutions	Training	USN-6	implement data augmentation techniques (e.g., rotation, flipping) to improve the model's robustness and accuracy.	we could do testing	medium	Sprint 3
	model deployment & Integration	USN-7	Deploy the trained model as a RESTful API or web service, making it accessible for malware detection purposes. Additionally, integrate the model's API into a user-friendly web interface, enabling users to upload files and obtain malware classification results.	we could create web application	medium	Sprint 4

	Testing & quality assurance	USN-8	Conduct intensive testing of the model and web interface to uncover and report any issues, including false positives and false negatives. Further, fine-tune the model's hyperparameters and optimize its performance based on feedback and results.	We could do post release service	medium	Sprint 5