

Project Design Phase-II
Technology Stack (Architecture & Stack)

| | |
|---------------|--------------------------------------|
| Date | 23 October 2023 |
| Team ID | 2.11 |
| Project Name | Malware Detection and Classification |
| Maximum Marks | 4 Marks |

Technical Architecture:

The Deliverable shall include the architectural diagram as below and the information

Example: Order processing during pandemics for offline mode

Reference:

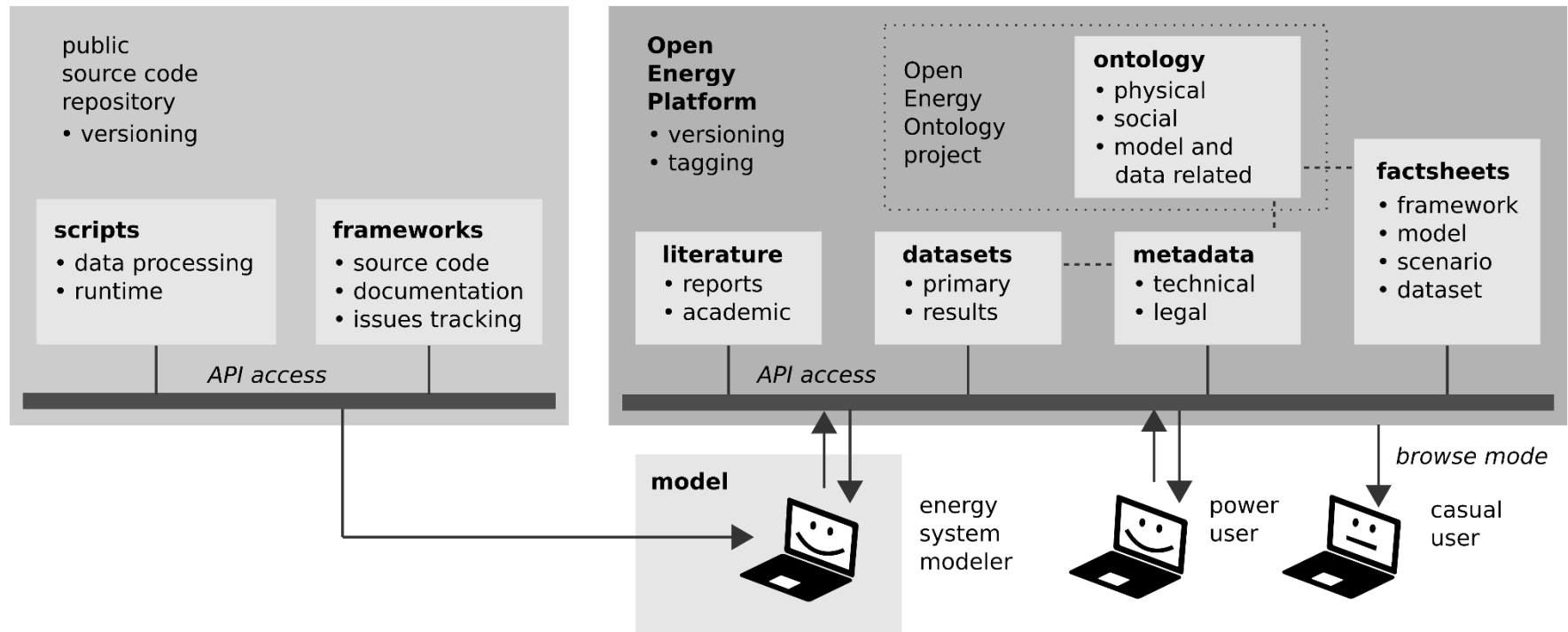


Table-1 : Components & Technologies:

| S.No | Component | Description | Technology |
|------|-------------------------|--|-------------------------|
| 1. | Library | used for data preprocessing, data exploration, and data cleaning in the project. | Pandas |
| 2. | Library | machine learning library provides a wide range of machine learning algorithms, tools for model selection, and evaluation. It is used for model building, training, and evaluation. | SciKit |
| 3. | Library | package for scientific computing with Python. used for numerical operations, mathematical functions, and creating arrays of data, especially when working with data preprocessing. | Numpy |
| 4. | Development Environment | interactive development environment that allows you to write and execute code in a web-based interface. It's commonly used for data exploration, analysis, and report generation. | Google Colab |
| 5. | LDA | LDA is a dimensionality reduction technique used in the code to transform and reduce the dimensionality of the data while preserving class separability. | LDA |
| 6. | Regression | Regression is a machine learning algorithm for classification tasks. It's used in the code to build a classification model. | Logistic Regression |
| 7. | Scaler | Scaler is used for feature scaling, ensuring that all features have the same scale and are centred around zero. It's crucial for many machine learning algorithms. | Standard Scaler |
| 8. | Algorithm | Feature selection algorithms are techniques used to select the most relevant features from a dataset. | SelectKBest, RFE |
| 9. | Validation tool | Cross-validation is a method used to assess a machine learning model's performance by dividing | RepeatedStratifiedKFold |

| | | | |
|-----|-----------------------------|--|--------------------------------|
| | | the dataset into multiple subsets for training and testing. | |
| 10. | Model | Machine learning models are algorithms that learn patterns and relationships in data to make predictions. | Decision Trees, Random Forest |
| 11. | Model Interpretability Tool | Model interpretability tools are used to explain and interpret the predictions of complex machine learning models. | Decision Trees, Random Forest. |

Table-2: Application Characteristics:

| S.No | Characteristics | Description | Technology |
|------|--|--|----------------------------|
| 1. | Programming Language | Versatile programming language used for data preprocessing, machine learning model development, and data analysis in the project. It offers a wide range of libraries and tools for these tasks. | Python |
| 2. | Confusion Matrix and Performance Metrics | Confusion matrices and various performance metrics, such as accuracy, precision, recall, F1-score, and ROC-AUC, are used to evaluate the model's performance and assess its strengths and weaknesses. | ROC-AUC |
| 3. | Hyperparameter Tuning | Hyperparameter tuning is the process of optimizing the parameters of machine learning models to improve their performance | Grid search, random search |
| S.No | Characteristics | Description | Technology |
| 4. | Feature Engineering Techniques | Feature engineering involves creating new features or transforming existing ones to enhance the model's ability to capture patterns in the data. Various techniques are applied in the project to improve model performance. | Technology used |

| | | | |
|----|-------------------|--|-------------------------|
| 5. | Ensemble Learning | Ensemble learning combines multiple machine learning models to improve predictive performance. Methods like bagging and boosting are used to create more robust models. | Random forest, AdaBoost |
| 6. | Version Control | Version control systems are used for tracking changes in the codebase, collaborating with team members, and maintaining a history of modifications and improvements made to the project. | Git |

References:

<https://c4model.com/>

<https://developer.ibm.com/patterns/online-order-processing-system-during-pandemic/>

<https://www.ibm.com/cloud/architecture> <https://aws.amazon.com/architecture>

<https://medium.com/the-internal-startup/how-to-draw-useful-technical-architecture-diagrams-2d20c9fda90d>