

CHAPTER 7

ANDROID MALWARE ANALYSIS TOOL

7.1 INTRODUCTION

Malware analyzer is a platform independent light weight android malware detection web-based application. In addition to malware detection Malware Analyzer also provides android malware analysis and predictive modelling. Our tool mainly provides the following modules/functionalities:

1. Analyze module: examine android applications for malicious activity using our pretrained classification models.
2. Extraction module: Extract features like permissions, opcodes etc. from given android applications.
3. Reduction module: Reduce the feature space i.e dimensionality of a dataset using two popular topic modelling techniques LSI and HDP.
4. Training module: Train classification models on user's dataset.

7.2 INSTALLATION

Malware analyzer is a cross-platform tool. It has been configured to work on any platform with python 3+ support. We have used HTML, CSS and Python for its development stack and developed it in windows 10 environment.

7.3 SYSTEM REQUIREMENTS

- Operating System: Windows 7/8.1/10, Ubuntu 16.04.6 - 64-bit
- RAM: 4GB
- Disk space: At least 512 MB

7.4 ANALYSIS MODULE

Malware analyzer allows users to detect malicious nature if present in android applications using the pretrained classification models. It comprises of two components. First, an Input component shown in Figure 7.1 that accepts an android application and prompts the user to select the reduction method and classification models to be used in evaluation of application. It also provided an analyze button to start the processing of android application based on user input and second, an Output Component is responsible for generating the final report on the basis of classification models. It gives an overall decision about the application's nature as well as the model specific result. Figure 7.2 and Figure 7.3 shows the analysis report generated in the output component for both malicious and benign application.

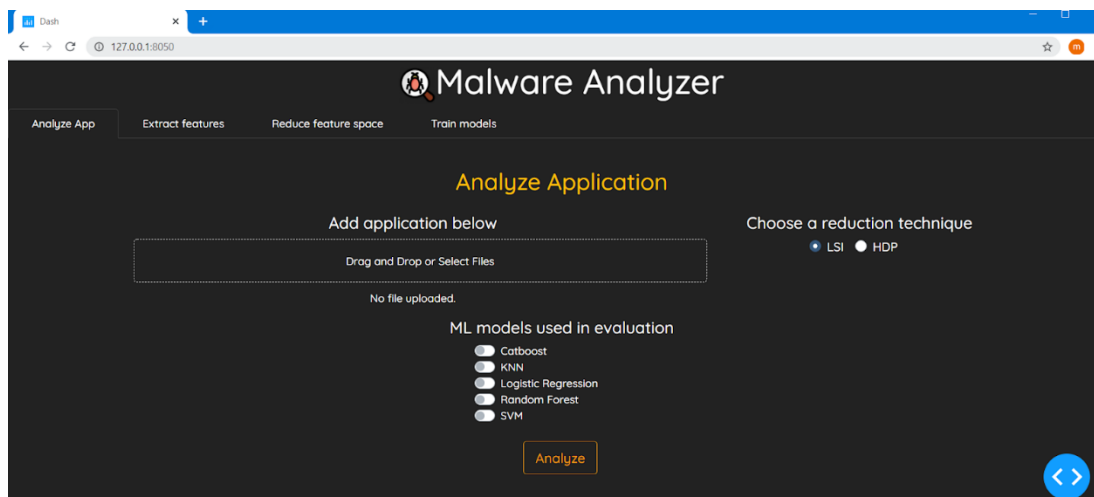


Figure 7.1. Input component of analyze module

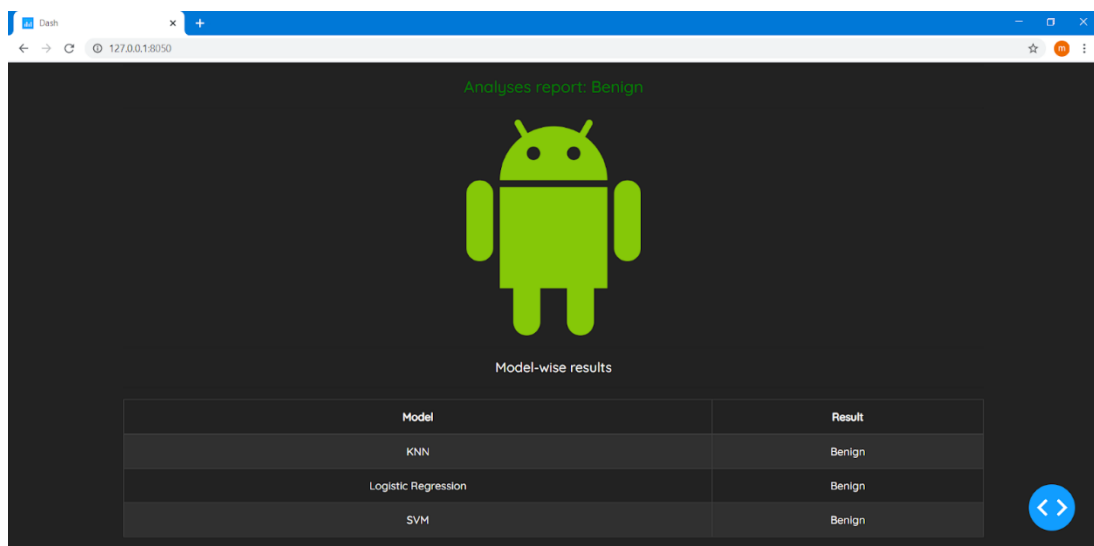
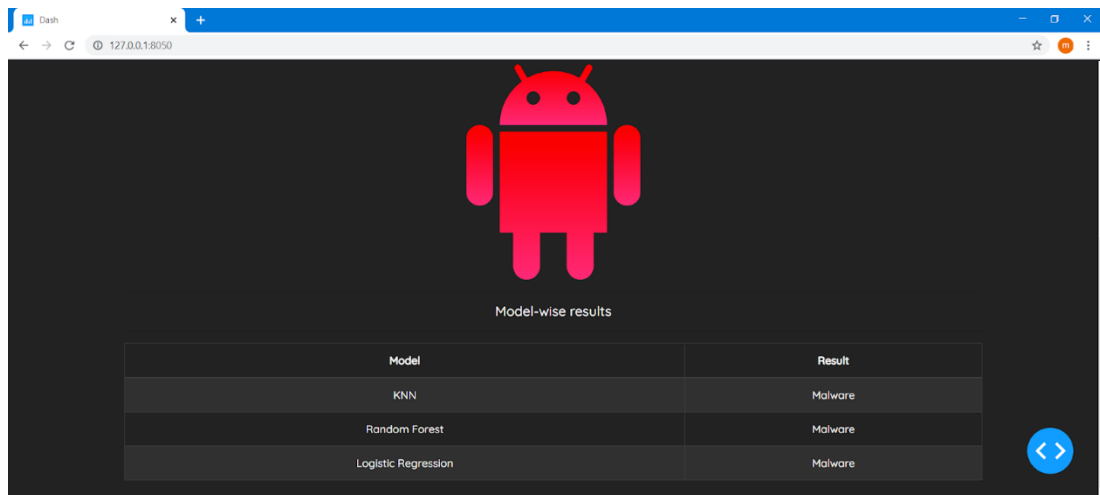


Figure.7.2 Output component of analyze module for benign application.



The screenshot shows a web browser window with a dark-themed dashboard. At the top, there's a red Android robot icon. Below it, the text "Model-wise results" is displayed. A table with two columns, "Model" and "Result", is shown. The table contains three rows of data. A blue navigation button with left and right arrows is located at the bottom right of the table area.

Model	Result
KNN	Malware
Random Forest	Malware
Logistic Regression	Malware

Figure 7.3 Output component of analyze module for benign application.

7.5 EXTRACTION MODULE

Extraction module forms a key part of the analysis functionality of our tool. It enables users to extract feature(s) from android applications. It supports feature extraction of multiple applications at once. Currently, It is capable of extracting application's permissions, intents, opcodes and api calls. A user can extract any subset of features from an application. The input user interface consists of a dialog box to upload APK files and checklist allows user to select which features to extract from the given APK files as demonstrated in Figure 7.4. The output of the extraction process is saved in a csv file and a report is shown to the user in output section shown in Figure 7.5.



Figure 7.4 Input user interface of Extraction module.

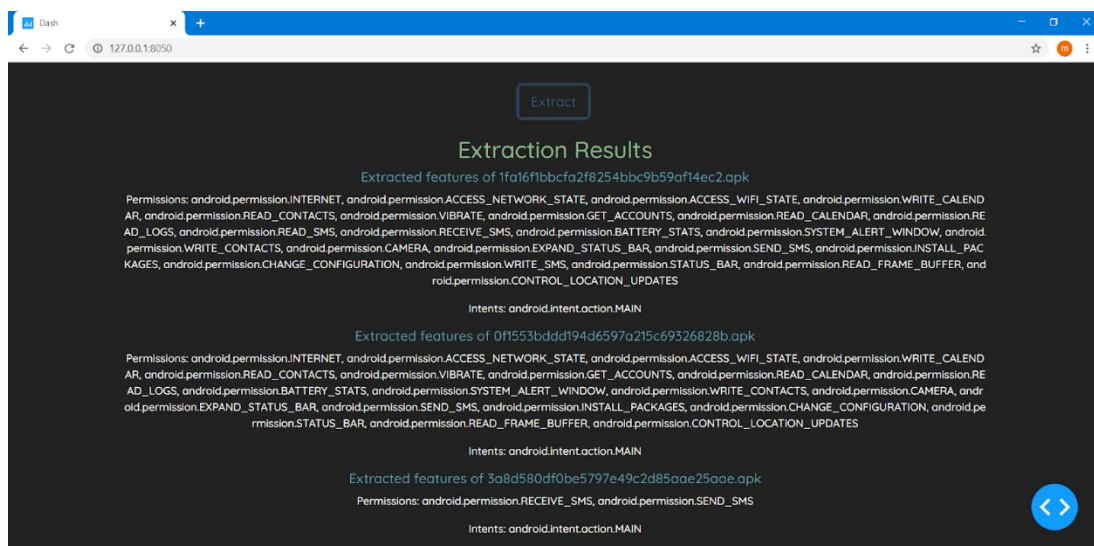


Figure 7.5 Output of Extraction module.

7.6 REDUCTION MODULE

The feature space of a dataset has a huge impact on the run time of classification models and android application's feature space can be in order of thousands. Therefore, reduction is an important step in android malware analysis. Malware analyzer's reduction module provides two reduction techniques: Latent semantic indexing and Hierarchical Dirichlet Process. The module accepts a csv file consisting of the original feature dataset and then prompts the user to select a reduction method and reduced feature space's dimensionality through radio buttons and input field as shown in Figure 7.6. After the feature space is reduced reduction module saves the reduced dataset to a csv file and presents the reduced dataset in a tabular format as illustrated in Figure 7.7 to the user.

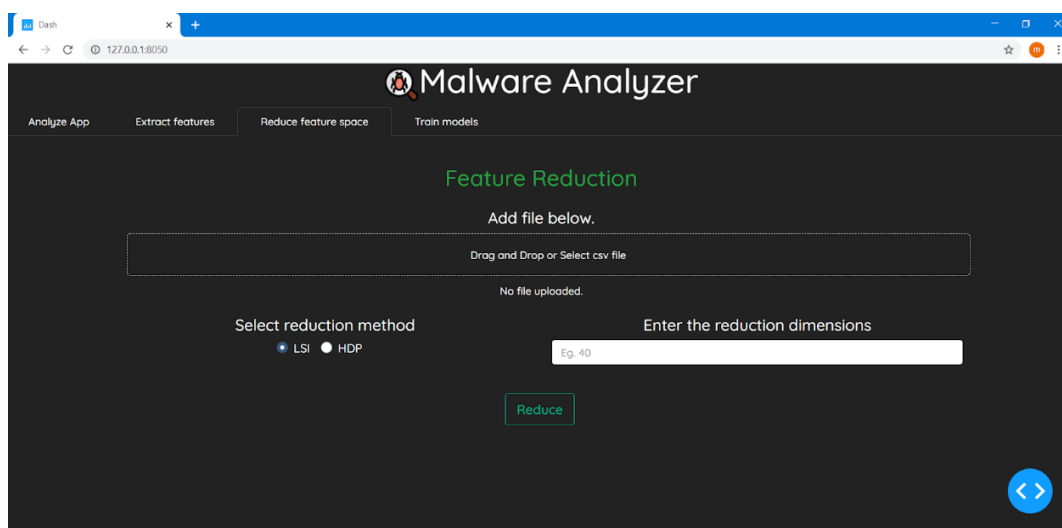


Figure 7.6 Input part of Extraction module window.

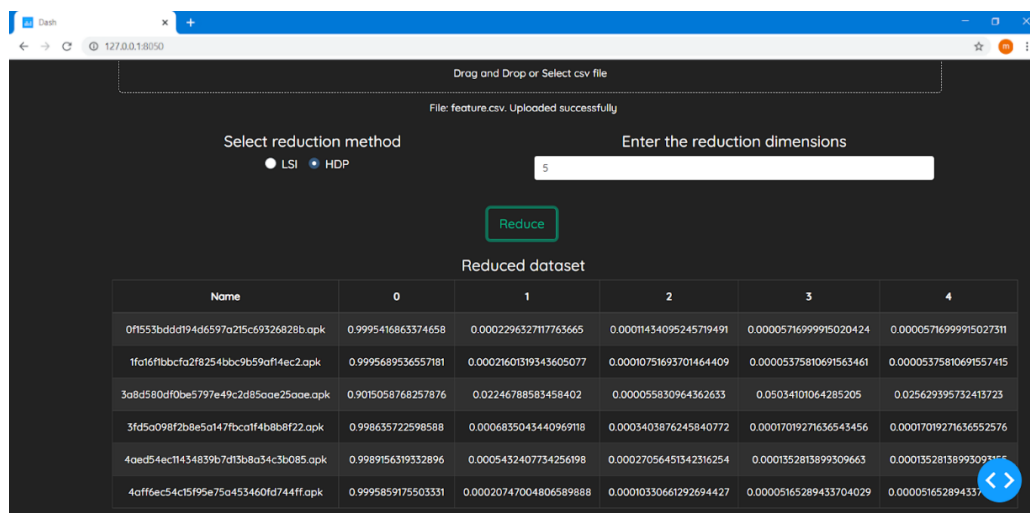


Figure 7.7 Output of Extraction module window.

7.7 TRAIN MODULE

The training module forms the last and most important part of the malware analysis. Its key functionalities include training various classification models using the user specified dataset, saving of trained models and combining of datasets. Its user interface comprises of an input form as in Figure 7.8 that includes two dialog boxes one for feature dataset to combine different dataset partitioned on feature space and other for the class labels of applications in feature dataset. Input form also includes a checklist for classification models and an output section where the trained model evaluation metrics involving accuracy, precision, recall and F1-score is published in a tabular form as shown in Figure 7.9.

Malware Analyzer

Analyze App Extract features Reduce feature space **Train models**

Train Models

Add Dataset file(s) below.

Drag and Drop or Select csv file

No file uploaded.

Add Labels file.

Drag and Drop or Select csv file

No file uploaded.

0.2 Dataset split ratio

Select ML models

- ☐ Catboost
- ☐ KNN
- ☐ Logistic Regression
- ☐ Random Forest
- ☐ SVM

Train

Figure 7.8 Input Form of Train module.

Select ML models

- ☒ Catboost
- ☒ KNN
- ☒ Logistic Regression
- ☐ Random Forest
- ☐ SVM

Train

Trained Model Metrics

name	accuracy	precision	recall	f1-score
Catboost	93.673966	0.941176	0.909091	0.924855
Logistic Regression	89.537713	0.880000	0.875000	0.877493
Random Forest	93.917275	0.963190	0.892045	0.926254

Figure 7.9 Output section of Train module.