Machine Learning Based Classification of Android Apps through Text Features

Mohamed Guendouz*, Abdelmalek Amine*
Reda Mohamed Hamou*

*GeCoDe Laboratory, Tahar Moulay University of Saida. Saida, Algeria
adresse@email,
http://www.une-page.html

1   Introduction

This paper deals with the problem of Android mobile apps classification using machine learning and text mining methods. Our approach consists in applying some machine learning methods on text characteristics that are extracted from app's description on Google Play Store.

Our proposed approach consists of two main phases. First we collect information about apps from the Google Play store using a web crawler. Then we extract some text information from this data. In our case, for each app we have extracted its description and its category. Second, we train different classifiers, this step involves pre-processing the text which includes removing URLs and digits, tokenization, and calculating TF*IDF to transfer text to a numeric vector which can be used as input for classifiers. Finally, to evaluate performance of our system, we have conducted various experiments on three real datasets using different evaluation metrics. Figure 1 illustrates the architecture of our app classification framework.

Finally, we evaluate our approach on three real datasets, obtained results shows that the use of text features in classifying Android apps can performs well.
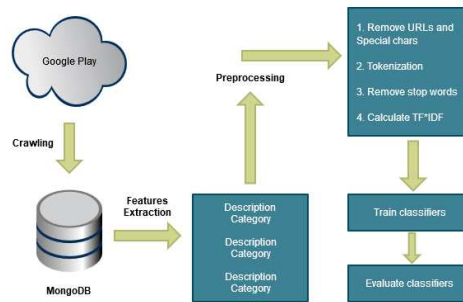

FIG. 1 – *Architecture of the proposed framework*