



Exploratory study of Machine Learning Tools on Big Data platforms

Filipe Ferreira da Ascensão, Prof. Nuno Lopes, Prof. Joaquim Silva

Instituto Politécnico do Cávado e do Ave

BACKGROUND: There are today techniques and tools for analyzing Big Data. These tools leverage the large processing capabilities available, and can deliver accurate results by allowing machines to learn with the data processed. Using algorithms, machines can create knowledge and aid in Data Driven Decisions. Nowadays, people and organizations can identify profitable opportunities and/or avoid risks based on the machine aid provided.

OBJECTIVES: The objective is to study the Machine Learning tools for Big Data platforms. The study will be based on the analysis of tools and will be made by characterizing the functionalities, advantages, disadvantages and possibilities of implementing of them in Big Data platforms.

METHODOLOGY: To explore the Machine Learning tools for Big Data platforms. The methodology adopted will be based on literature review, identification and selection of featured tools, evaluation of selected tools, and the redaction of the dissertation.

RESULTS AND CONCLUSIONS: The results extracted from the technical aspects of tools/platform will be complemented by the application of a possible use case. This dissertation aim to be a document that will serve as reference for an introduction of the Machine Learning tools in Big Data platforms.

Keywords: Machine Learning, Big Data, Distributed Computing, MapReduce, Hadoop, Spark.