

Machine learning for data mining and performance optimization at the CERN Large Hadron Collider

Progress Report

Marc Ferriggi

Supervisor: Dr. Gianluca Valentino



L-Università
ta' Malta

Faculty of Science
University of Malta
December 2018

*Submitted in partial fulfillment of the requirements for the degree of B.Sc.
(Hons.) Computing Science AND Statistics and Operations Research*

Contents

1	Abstract	2
---	--------------------	---

1 Abstract

The CERN particle accelerator complex generates around 2 TB of data per week from almost 1 million signals. In this dissertation, unsupervised machine learning techniques for applications such as clustering and anomaly detection shall be used to analyze past LHC data in order to visualize correlations, determine data driven models and identify opportunities for improving the LHC machine availability and performance reach.

2 Introduction & Motivation

3 Why is the Problem non-Trivial

4 Background Research and Literature Review

5 Aims and Objectives

6 Methods and Techniques Used or Planned

7 The Evaluation Strategy and Technique being Proposed

8 Deliverables

9 Progress

10 References