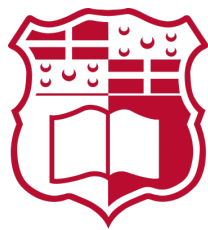


# Machine learning for data mining and performance optimization at the CERN Large Hadron Collider

## 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
2	Introduction & Motivation . . . . .	2
3	Why is the Problem non-Trivial . . . . .	2
4	Background Research and Literature Review . . . . .	2
5	Aims and Objectives . . . . .	2
6	Methods and Techniques Used or Planned . . . . .	2
7	The Evaluation Strategy and Technique being Proposed . . . . .	2
8	Deliverables . . . . .	2
9	Progress . . . . .	2
10	References . . . . .	2

## **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.

## **Introduction & Motivation**

### **Why is the Problem non-Trivial**

### **Background Research and Literature Review**

### **Aims and Objectives**

### **Methods and Techniques Used or Planned**

### **The Evaluation Strategy and Technique being Proposed**

### **Deliverables**

### **Progress**

### **References**