**ABSTRACT**

Distributed machine learning (DML) can realize massive dataset training when no single node can work out the accurate results within an acceptable time. However, this will inevitably expose more potential targets to attackers compared with the non-distributed environment. In this paper, we classify DML into basic-DML and semi-DML. In basic-DML, the center server dispatches learning tasks to distributed machines and aggregates their learning results. While in semi-DML, the center server further devotes resources into dataset learning in addition to its duty in basic-DML. We firstly put forward a novel data poison detection scheme for basic-DML, which utilizes a cross-learning mechanism to find out the poisoned data. We prove that the proposed cross-learning mechanism would generate training loops, based on which a mathematical model is established to find the optimal number of training loops. Then, for semi-DML, we present an improved data poison detection scheme to provide better learning protection with the aid of the central resource. To efficiently utilize the system resources, an optimal resource allocation approach is developed. Simulation results show that the proposed scheme can significantly improve the accuracy of the final model by up to 20% for support vector machine and 60% for logistic regression in the basic-DML scenario. Moreover, in the semi-DML scenario, the improved data poison detection scheme with optimal resource allocation can decrease the wasted resources for 20-100%.

INDEX

Title page

Certificate

Acknowledgement

Declaration

Abstract

**CHAPTER 1:** INTRODUCTION……………………………………………………………1

1.1 Introduction……………………………………………………………......1

**CHAPTER 2:** LITERATURE SURVEY…………………………………………………….3

**CHAPTER 3:** SYSTEM ANALYSIS………………………………………………………..5

3.1 Existing System............................................................................................5

3.2 Propose System............................................................................................5

3.3 Feasibility Study…………………………………………………………..6

3.4 Input Design………………………......…………………………………...7

3.5 Output Design…………………………...………………………………...8

**CHAPTER 4:** SYSTEM REQUIREMENTS SPECIFICATION…………………………...10

4.1 Functional Requirements…………………………………………………10

4.2 Non Functional Requirements……………………………………………11

**CHAPTER 5:** SYSTEM DESIGN………………………………………………………….15

5.1 System Specifications…………………………………………………....15

5.2 System Architecture.............………………………………………..……15

5.3 UML Diagrams…...…………………………………………………...…16

5.3.1 Use Case Diagram………………………………………………......17

5.3.2 Class Diagram……………………………………………………....18

5.3.3 Sequence Diagram………………………………………………......19

5.3.4 Activity Diagram.............................................................................20

5.3.5 Collaboration Diagram ..........………………........…………….…21

**CHAPTER 6:** SYSTEM IMPLEMENTATION…………………………………………..22

6.1 Modules Description…………………………………………………....22

6.2 Sample Code…………………………………………………………....23

6.3 System Testing………………………………………………………….27

6.4 Types of Testing………………………………………………………..28

6.5 Testing Methodologies…………………………………………………30

**CHAPTER 7:** SCREEN SHOTS…………………………………………………………..36

**CHAPTER 8:** CONCLUSION…………………………………………………. ………..44

**APPENDIX-A :** BIBLIOGRAPHY………………………………………………………..45

**B :** TECHNOLOGY USED………………………………………………....46

**LIST OF FIGURES**

Figure No Figure Name Page No

Fig 5.2 System Architecture 15

Fig 5.3.1 Use Case Diagram 17

Fig 5.3.2 Class Diagram 18

Fig 5.3.3 Sequence Diagram 19

Fig 5.3.4 Activity Diagram 20

Fig 5.3.5 Collaboration Diagram 21