**TABLE OF CONTENTS**

**Content** **Page No**

**Abstractiii**

**Table of Contentvi**

**List of Figuresxi**

**Chapter 1: Introduction12**

1.1 Problem Statement12

1.2 Industry/Society Benefitted14

1.3 Proposed Solution14

1.4 Feasibility14

1.5 Literature Survey15

**Chapter 2: Machine Learning18**

2.1 What is Machine Learning?18

2.2 History of Machine Learning19

2.3 How does Machine Learning Work21

2.4 Literature Survey22

2.5 The Challenges Facing Machine Learning22

2.6 Features of Machine Learning23

2.7 Types of Machine Learning23

2.7.1 Supervised Learning24

2.7.2 Unsupervised Learning24

2.7.3 Reinforcement Learning25

2.8 Machine Learning Lifecycle25

2.9 Application of Machine Learning28

2.10 Future Scope29

**Chapter 3: Machine Learning Language30**

3.1 Python-The New Generation Language30

3.1.1 Features30

3.1.2 Why Python is a Suitable Language for Machine Learning31

3.3 Introduction to R33

3.3.1 Features34

3.3.2 Why R is a Suitable Language for Machine Learning35

**Chapter 4: SRS37**

4.1 Introduction37

4.1.1 Purpose37

4.1.2 Scope37

4.1.3 Objective38

4.2 Specific Requirements38

4.2.1 Hardware Requirements38

4.2.2 Software Requirements38

**Chapter 5: Machine Learning Algorithms43**

5.1 Decision Tree43

5.1.1 Importance of Decision43

5.1.2 How Decision Tree Works44

5.1.3 Advantages and Disadvantages of Decision Tree45

5.2 Random Forest47

5.2.1 Importance of Random Forest47

5.2.2 How Random Forest Works48

5.2.3 Advantages and Disadvantages of Random Forest49

5.3 Support Vector Machine (SVM)50

5.3.1 Importance of SVM50

5.3.2 How SVM Works51

5.3.3 Advantages and Disadvantages of SVM51

5.4 XGBoost52

5.4.1 Importance of XGBoost53

5.4.2 How XGBoost53

5.4.3 Advantages and Disadvantages of XGBoost54

**Chapter 6: Gathering Data56**

6.1 Data Collection56

6.1.1 Column Definitions56

6.2 Full Dataset Content57

**Chapter 7: Libraries Used and Their Installations58**

7.1 Numpy58

7.1.1 Description58

7.1.2 Function58

7.1.3 Installation59

7.1.4 Use in the Project59

7.2 Pandas59

7.2.1 Description60

7.2.2 Function60

7.2.3 Installation60

7.2.4 Use in the Project61

7.3 Matplotlib61

7.3.1 Description61

7.3.2 Function61

7.3.3 Installation62

7.3.4 Use in the Project62

7.4 Util62

7.4.1 Description62

7.4.2 Function63

7.4.3 Installation63

7.4.4 Use in the Project63

7.5 Seaborn63

7.5.1 Description63

7.5.2 Function63

7.5.3 Installation64

7.5.4 Use in the Project64

7.6 Keras64

7.6.1 Description65

7.6.2 Function65

7.6.3 Installation65

7.6.4 Use in the Project67

7.7 Warnings in Python67

7.7.1 Warning Categories68

7.8 Statsmodels68

7.8.1 Features70

7.9 Scikit-Learn70

7.9.1 Features71

**Chapter 8: Feature Extraction72**

8.1 Address Bar Based Features72

8.1.1 Using the IP Address72

8.1.2 Long URL to Hide Suspicious Part73

8.1.3 Using URL Shortening Service73

8.1.4 URL’s Having “@” Symbol74

8.1.5 Redirecting Using “//”74

8.1.6 Adding Prefix or Suffix Separated by - to the domain74

8.1.7 Sub Domain or Multi Sub Domain75

8.1.8 HTTPS 75

8.1.9 Domain Registration Length76

8.1.10 Favicon 76

8.1.11 Using Non-Standard Port77

8.1.12 The Existence of “HTTPS” Token in the Domain77

8.2 Abnormal Based Features77

8.2.1 Request URL77

8.2.2 URL of Anchor78

8.2.3 Links in <Meta>, <Script> and <Link> tags78

8.2.4 Server Form Handler78

8.2.5 Submitting Information to Email79

8.2.6 Abnormal URL79

8.3 HTML and JavaScript Based Features79

8.3.1 Website Forwarding79

8.3.2 Status Bar Customization79

8.3.3 Disabling Right Click79

8.3.4 Using Pop-Up Window80

8.3.5 IFrame Redirection80

8.4 Domain Based Features80

8.4.1 Age of Domain80

8.4.2 DNS Record80

8.4.3 Website Traffic81

8.4.4 Page Rank81

**Chapter 9: Dataset Preparation82**

9.1 Loading Libraries82

9.2 Descriptive Statistics83

9.3 Data Visualization84

9.4 Data Cleaning85

9.4.1 How do you Clean Data86

9.4.2 Components of Quality Data89

9.5 Feature Sampling90

9.6 Dimensionality Reduction90

9.6.1 The Curse of Dimensionality91

9.6.2 Feature Encoding91

**Result93**

**Future Work94**

**Conclusion95**

**References96**