**ACKNOWLEDGEMENT**

We are happy to present this **Project work (18AIP83)** report after completing it successfully. This final year project would not have been possible without the guidance, assistance and suggestions of many individuals. We would like to express our deep sense of gratitude and indebtedness to each and every one who has helped us make this a success.

We heartily thank our Principal**, Dr. Sanjay H A,** BMS Institute of Technology & Management, Autonomous Institute Affiliated to VTU for his constant encouragement and inspiration in taking up this Project Work.

We heartily thank our Professor & Head of the Department, **Dr. Anupama H S**, Department of Artificial Intelligence and Machine Learning, BMS Institute of Technology & Management, Autonomous Institute Affiliated to VTU, for her constant encouragement and inspiration in taking up this Project Work.

We gracefully thank our Project Work Guide, **NAME,** Designation, Dept. of AI&ML for his/her guidance, support and advice.

We also like to thank our **Student Project Assessment And Review Committee (SPARC)** Member’s **Dr. Bharathi Malakreddy A, Dr. Anupama H S, Dr. Pradeep K R, Dr. Rajesh I S, Dr. Niranjanamurthy M** , **Dr. Shanmuga Sundaram, Prof.Shobhit Tembhre Prof.Sanjay M Belgaonkar, Prof. Sachin A U, Prof. Pradeep Kumar G M** for their help and support provided to carry out the project and complete it successfully.

Special thanks to all the staff members of Artificial Intelligence and Machine Learning department for their help and kind co- operation.

Lastly, We thank our parents and friends for the support and encouragement given to us in completing this Project Work successfully.

II

**ABSTRACT**

Due to its continuously increasing occurrence, more and more families are influenced by diabetes mellitus. Most diabetics know little about their health quality or the risk factors they face prior to diagnosis. In this study, we have proposed a novel model based on data mining techniques for predicting type 2 diabetes mellitus (T2DM).

The main problems that we are trying to solve are to improve the accuracy of the prediction model, and to make the model adaptive to more than one dataset. Based on a series of preprocessing procedures, the model is comprised of two parts, the improved K-means algorithm, the Logistic regression algorithm, Support vector clustering and Decision tree. The Pima Indians Diabetes Dataset and the Waikato Environment for Knowledge Analysis toolkit were utilized to compare our results with the results from other researchers. The conclusion shows that the model attained a 3.04% higher accuracy of prediction than those of other researchers. Moreover, our model ensures that the dataset quality is sufficient. To further evaluate the performance of our model, we applied it to two other diabetes datasets. Both experiments' results show good performance. As a result, the model is shown to be useful for the realistic health management of diabetes.

III

**TABLE OF CONTENTS**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Chapter No.** | **Title** | | | **Page No.** |
|  | **ABSTRACT** | | | **III** |
| **1.** | **INTRODUCTION** | | | **01** |
| **1.1** | Over View | | | 01 |
| **1.2** | Purpose of the problem | | | 02 |
| **1.3** | Scope of the project | | | 03 |
| **1.4** | Definitions | | | 03 |
| **2.** | **PROBLEM DEFINITION** | | | **06** |
| **2.1** | Problem Statement | | | 06 |
| **3** | **LITERATURE SURVEY** | | | 08 |
| **3.1** | Behavioral Interventions for type 2 diabetes: 08  Background | | | 08 |
| **3.2** | Case Study: A 36-year-old woman with type 10  2 diabetes and pregnancy | | | 10 |
| **4.** | SYSTEM REQUIREMENT SPECIFICATION 12 | | | 12 |
| **4.1** | Hardware Requirements | | | 12 |
| **4.2** | Software Requirements | | | 12 |
|  | 4.2.1 Python | | | 12 |
|  | 4.2.2 SQLite | | | 13 |
|  | 4.2.3 Jupyter Notebook | | | 14 |
|  | 4.2.4 Anaconda | | | 15 |
|  | 4.2.5 Django | | | 16 |
| **4.3** | Non-Functional Requirements | | | 18 |
| 5. | | | **DESIGN** | **19** |
| 5.1 | | | System Architecture | 19 |
| 5.2 | | | Data Flow Diagram | 29 |
|  | | | 5.2.1 Constructing Data Flow Diagram | 31 |
| **5.3** | | | Sequence Diagram | 31 |
| **5.4** | | | Class Diagram | 33 |
| **5.5** | | | Activity Diagram | 34 |
| **5.6** | | | Use Case Diagram | 35 |
| **6.** | | | **IMPLEMENTATION** | **37** |
| **6.1** | | | Programming Language Selection | 37 |
| **6.2** | | | Factors to Consider | 37 |
|  | | | 6.2.1 Platform Selection | 38 |
|  | | | 6.2.2 Elasticity | 39 |
|  | | 6.2.3 Time to Production | | 39 |
|  | | 6.2.4 Performance | | 40 |
|  | | 6.2.5 Support and Community | | 40 |
| **6.3** | | Coding | | 40 |
|  | | 6.3.1 Code for Clustering of Dataset using k-means | | 41 |
|  | | 6.3.2 Code for comparing training and Testing data using LR | | 42 |
| **7.** | | **TESTING AND RESULTS** | | **44** |
| **7.1** | | Types of Tests | | 44 |
|  | | 7.1.1 Unit Testing | | 44 |
|  | | 7.1.2 Integration Testing | | 44 |
|  | | 7.1.3 Functional Testing | | 45 |
|  | | 7.1.4 System Testing | | 45 |
|  | | 7.1.5 White Box Testing | | 45 |
|  | | 7.1.6 Black Box Testing | | 46 |
|  | | 7.1.6 Acceptance Testing | | 46 |
| **7.2** | | Test Cases | | 46 |
| **7.3** | | Performance Evaluation | | 47 |
|  | | 7.3.1 Calculation of Precision | | 47 |
| **8.** | | **SNAPSHOTS** | | 49 |
| **8.1** | | Home Page | | 49 |
| **8.2** | | User Login Page | | 50 |
| **8.3** | | Enquiry | | 50 |
| **8.4** | | Patient Registration | | 51 |
| **8.5** | | Prediction | | 52 |
| **8.6** | | Analysis | | 53 |
| **9.** | | **CONCLUSION** | | 56 |
| **10.** | | **FUTURE ENHANCEMENTS** | | 57 |
| **11.** | | **REFERENCES** | | 58 |

V

**LIST OF FIGURES**

|  |  |  |
| --- | --- | --- |
| **Fig. No.** | **Figure Name** | **Page No.** |
| 5.1 | Architecture of proposed system | 20 |
| 5.2 | Flowchart for improved K-means | 23 |
|  | Algorithm |  |
| 5.3 | Flowchart for Logistic Regression |  |
|  | Algorithm | 25 |
| 5.4 | Flowchart for Support Vector |  |
|  | Clustering Algorithm | 26 |
| 5.5 | Flowchart for Decision Tree | 27 |
| 5.6 | General Notation of DFD | 30 |
| 5.7 | Sequence Diagram of proposed work | 32 |
| 5.8 | Class Diagram of proposed work | 33 |
| 5.9 | Activity Diagram of proposed work | 35 |
| 5.10 | Use Case Diagram of proposed work | 36 |
| 8.1 | Home Page | 50 |
| 8.2 | User Login Page | 50 |

VI

**LIST OF TABLES**

|  |  |  |
| --- | --- | --- |
| **Table. No.** | **Table Name** | **Page No.** |
| 2.1 | Architecture of proposed system | 20 |
| 3.2 | Flowchart for improved K-means | 23 |

**\**

VII