M.Sc. Data Science and Analytics

**Detailed Project Proposal**

7COMxxxxxxxxx- Advanced Computer Science Masters Project

Student Name – Suresh

Student ID –

Supervisor –

--------------------------------------------------------------------------------------------------------------------------

**Diabetes Detection Using Machine Learning**

**Problem Statement –**

Diabetes is amongst the most common causes of chronic illnesses in the United States, affecting millions of people each year and putting a considerable financial strain on the country's financial resources. Chronically high amounts of sugar staying in the circulation for persons with diabetes are related with complications such as heart disease, eyesight loss, lower-limb amputation, and kidney illness. While there is currently no cure for diabetes, efforts such as decreasing weight, eating healthfully, being physically active, and obtaining medical treatment may help many individuals reduce the negative effects of the illness. The ability to detect diabetes early may lead to lifestyle modifications and more effective treatment, making prediction models for diabetes risk useful tools for the general population and public health professionals.

**Aim –**

To address the above problem statement, the primary aim of this project is to develop Machine Learning algorithms to diagnose the diabetes in early stages based on the patients’ health inputs as the data.

**Objectives –**

1. Understand the trends and patterns among the diabetic patients using the Exploratory Data Analysis.
2. Analyze the most important health inputs that directly impact the diabetic conditions among the patients.
3. Use Machine Learning algorithms to develop a system that could identify the diabetes among the patients at early stages.
4. Evaluate the performance of the algorithms and draw interpretations from the results that were achieved.
5. For each algorithm that is being used in this project, find the optimal parameters that could achieve higher accuracies.
6. Balance between the underfitting and overfitting to achieve the appropriate fitting models.

**Research Questions –**

1. What are the most correlated features among the health inputs used in the data?
2. What is the role of the Machine Learning in the health industry and to what extent we can rely on the Machine Learning algorithms for detecting diseases in humans?
3. What is the most suitable Machine Learning algorithm that achieves high accuracy in this project?

**Project Plan –**

In the early phases of the project, my goal is to get a thorough understanding of the issue I'm working on and to do research on related topics that are readily accessible on the internet. In addition to this, I want to study the algorithms that will be employed in this project in order to have a better grasp of their usefulness.

After performing thorough research for my project, I will begin by using EDA on the data and plotting several visuals for the features in order to better comprehend the data and its implications. As a last step, I'll train models on data and then verify them using validation sets before evaluating their performance on test sets, which will be the last stage of the process. After I have completed the training of the algorithms, my next step will be to begin working on a thesis that will be based on my results and understandings gained from this research.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | February | March | April | May | June | July |
| Background research |  |  |  |  |  |  |
| Project Proposal |  |  |  |  |  |  |
| Analysing Legal and Ethical risks |  |  |  |  |  |  |
| Literature Review |  |  |  |  |  |  |
| Exploratory Data Analysis |  |  |  |  |  |  |
| Experiments |  |  |  |  |  |  |
| Results evaluation |  |  |  |  |  |  |
| FPR |  |  |  |  |  |  |
| Submission |  |  |  |  |  |  |

**References**

Sisodia, D. and Sisodia, D., 2018. Prediction of Diabetes using Classification Algorithms. Procedia Computer Science, 132, pp.1578-1585.

Yahyaoui, A., Jamil, A., Rasheed, J. and Yesiltepe, M., 2019. A Decision Support System for Diabetes Prediction Using Machine Learning and Deep Learning Techniques. 2019 1st International Informatics and Software Engineering Conference (UBMYK).

Mujumdar, A. and Vaidehi, V., 2019. Diabetes Prediction using Machine Learning Algorithms. Procedia Computer Science, 165, pp.292-299.