

**PREDICTING DIABETES USING MACHINE LEARNING: A  
COMPARATIVE ANALYSIS OF CLASSIFICATION  
ALGORITHMS  
Conception Phase (Portfolio Part 1)**

**Table of Contents**

Abstract..... 3

## **Abstract**

The research has underscored that the growing rates of diabetes have developed major challenges to the healthcare systems, as well as underscored the need for coherent tools for early diagnosis and prevention. The main agenda of this project is to use computational techniques in medical analysis by creating machine learning models that can predict diabetes using patient health indicators. The objective of using the Pima Indians Diabetes Dataset, which contains such attributes as glucose levels, BMI, blood pressure, age, and family history, is to see whether predictive models can be used to help healthcare professionals identify patients at risk. The researcher's main focus is to develop supervised machine learning algorithms, such as logistic regression as a baseline and advanced classifiers such as random forests and gradient boosting. The cogent methodology has been followed to develop the accurate implementation, such as the researcher will use data preprocessing, feature exploration, model training, and evaluation with the help of performance metrics, which are accuracy, precision, recall, and AUC. Also, for the implementation, the researcher will use Python libraries, such as scikit-learn, pandas, and matplotlib. The output has shown the comparative analysis of models, such as comparing the models' accuracy and precision to determine their effectiveness and limitations for medical prediction tasks. In addition to the performance aspect, the study will also consider how the application of such computational tools can be incorporated in the real-world decision-making processes towards better early detection and resource allocation in the healthcare setting.