VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI

A PROJECT REPORT ON

"DIABETES PREDICTION USING MACHINE LEARNING ALGORITHMS"

Submitted in partial fulfillment for the award of Degree of,

BACHELOR OF ENGINEERING

IN

COMPUTER SCIENCE & ENGINEERING

By

PRATHIKSHA	**4AL16CS070**
SHWETHA	**4AL16CS101**
SOUMYA MALLASARJA RUDRAPUR	**4AL16CS103**
VLEENA MASCARENHAS	**4AL16CS121**

Under the Guidance of

Mr. Vasudev S Shahapur

Associate Professor

ALVA'S
Education Foundation

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY

MOODBIDRI-574225, KARNATAKA

2019 – 2020

ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY

MIJAR, MOODBIDRI D.K. -574225, KARNATAKA

DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING

CERTIFICATE

This is to certify that the project entitled **"DIABETES PREDICTION USING MACHINE LEARNING ALGORITHMS"** has been successfully completed by

PRATHIKSHA	**4AL16CS070**
SHWETHA	**4AL16CS101**
SOUMYA MALLASARJA RUDRAPUR	**4AL16CS103**
VLEENA MASCARENHAS	**4AL16CS121**

the bonafide students of **DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING, ALVA'S INSTITUTE OF ENGINEERING AND TECHNOLOGY** of the **VISVESVARAYA TECHNOLOGICAL UNIVERSITY, BELAGAVI** during the year 2019–2020. It is certified that all corrections/suggestions indicated for Internal Assessment have been incorporated in the report deposited in the departmental library. The project report has been approved as it satisfies the academic requirements in respect of Project work prescribed for the Bachelor of Engineering Degree.

Mr. Vasudev S Shahapur **Project Guide**	**Dr. Manjunath Kotari** **Head of the department**	**Dr. Peter Fernandes** **Principal**

External Viva

Name of the Examiners	**Signature with Date**
1. HEMANTH KUMAR N P	
2. Dr. S. Mohideen Badhusha	

ABSTRACT

Diabetes is considered as one of the deadliest and chronic diseases which causes an increase in blood sugar. Many complications occur if diabetes remains untreated and unidentified. The tedious identifying process results in visiting of a patient to a diagnostic center and consulting doctor. But the rise in machine learning approaches solves this critical problem. The motive of this study is to design a model which can prognosticate the likelihood of diabetes in patients with maximum accuracy. Therefore six machine learning classification algorithms namely Decision tree, Gradient boosting, KNN, Logistic regression, Random forest and SVM are used in this experiment to detect diabetes at an early stage. Experiments are performed on Pima Indians Diabetes Database (PIDD) which is sourced from UCI machine learning repository. The performances of all the six algorithms are evaluated on various measures like Precision, Accuracy. Accuracy is measured over correctly and incorrectly classified instances. Results obtained show Gradient boosting outperforms with the highest accuracy of 81% comparatively other algorithms.