

MACHINE LEARNING

TOPIC

*Diabetes
prediction*



INTRODUCTION

Diabetes is one of the most common diseases in the world, when detected early, it is possible to stop the progression of the disease and prevent further complications. in this work, we design a predictive model that predicts whether a patient has diabetes or not, based on certain diagnostic measures contained in the dataset, and explore different techniques to improve performance and accuracy. logistic regression is the main algorithm used in this article and the analysis was performed using Jupiter notebook.

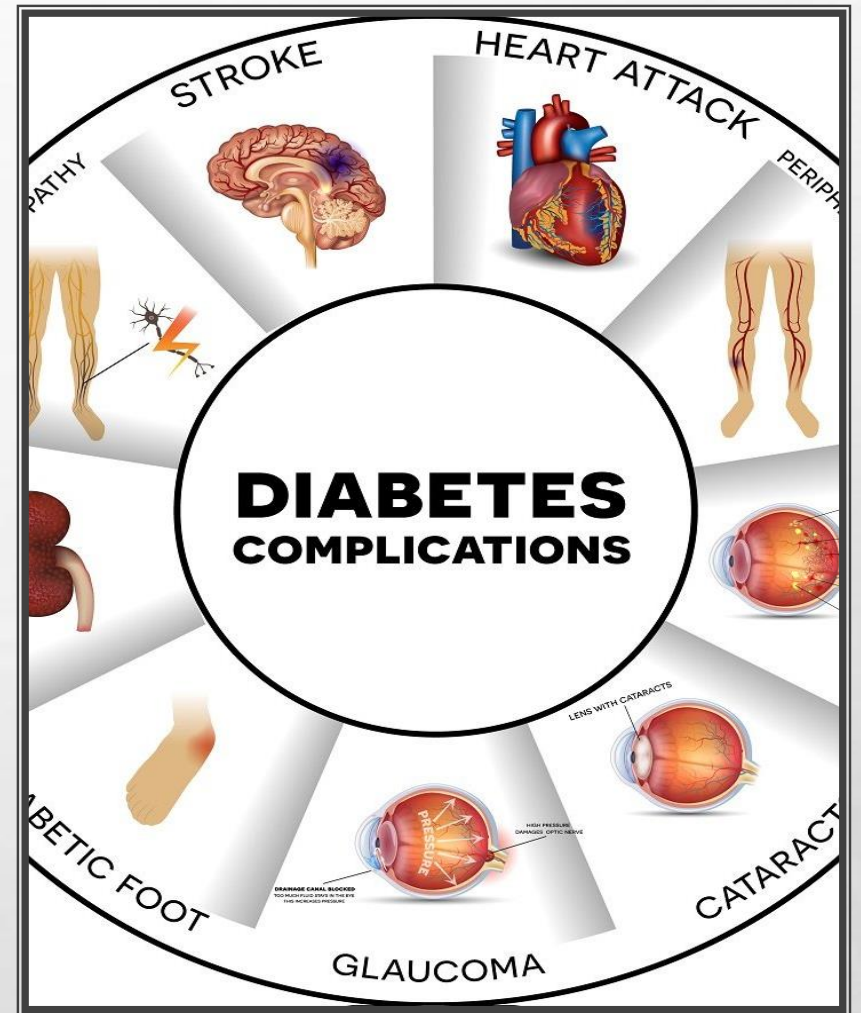
ABOUT RESEARCH PAPER AND DATASET

According to International Journal of Engineering Research and Technology (IJERT) which conducted a research on diabetes in 2020, Diabetes also known as diabetes mellitus affected about 463 million adults in the age bracket of 20-76 years in the world in 2019 and it was predicted that this number will increase to about 700 million in 2045. The research paper reveled that there are three types of diabetes. The type 1 diabetes which cause by the inability of the human body to produce enough insulin. The type 2 diabetes which occurs as results of the body cells not using enough insulin and the type 3 diabetes which occurs as a results of diabetes not detected early in the human body.

ABOUT RESEARCH PAPER AND DATASET

The paper further goes on to talk about the effect of diabetes on the human body. Some of these effects are;

1. Blindness
2. Heart attack
3. Stroke
4. Kidney failure etc.



ABOUT RESEARCH PAPER AND DATASET

According to the paper, 768 records of patients(all females with age at least 21 years) were collected and test was conducted on them to determine the possibility of them having diabetes. Some features was taken into consideration before the test was made and an outcome was given based on these features.

ABOUT RESEARCH PAPER AND DATASET

Some variables or features used for the prediction are:

1. **PREGNANCY:** SEVERAL TIMES THE PATIENT HAS BEEN PREGNANT.
2. **BLOOD PRESSURE:** DIASTOLIC BLOOD PRESSURE (MM HG).
3. **AGE:** IN YEARS.
4. **BMI:** $703 \times \text{WEIGHT (LBS)} / [\text{HEIGHT (IN)}]^2$
5. **GLUCOSE:** PLASMA GLUCOSE CONCENTRATION OVER TWO HOURS ORALLY GLUCOSE TOLERANCE TEST.

ABOUT RESEARCH PAPER AND DATASET

Some variables or features used for the prediction are:

6. **DIABETES PEDIGREE FUNCTION**

7. **SKIN THICKNESS:** TRICEPS SKINFOLD THICKNESS (MM).

8. **INSULIN:** TWO-HOUR SERUM INSULIN (MU U/ML).

*THERE IS AN **OUTCOME** VARIABLE WHICH IS CATEGORICAL AND A TARGET VARIABLE AND IT IS (0 IF NOT DIABETIC AND 1 IF DIABETIC)

ABOUT RESEARCH PAPER AND DATASET

*



CONCLUSION

IT HAS BEEN DEMONSTRATED THAT AMONG THE FIVE MODELS USED, THE EFFECTIVE TECHNIQUES FOR CREATING PREDICTION MODELS IS RANDOM FOREST, WHICH HAD THE HIGHEST TEST ACCURACY OF 74% COMPARED TO THE OTHER MODELS. THE METHOD SELECTED IS JUST ONE OF SEVERAL VARIABLES THAT AFFECT THE MODEL'S ACCURACY. PREPROCESSING OF DATA IS ONE SUCH ELEMENT. IN ORDER TO IMPROVE EFFICIENCY, REDUNDANT AND NULL VALUES MUST BE REMOVED. WHEN CHARACTERISTICS DIVERGE ON A BROAD SCALE, NORMALIZING THE VALUES ALSO HAS A SIGNIFICANT IMPACT

CONCLUSION

IN CONCLUSION, MACHINE LEARNING STANDS AS A TESTAMENT TO OUR ABILITY TO HARNESS THE POTENTIAL OF DATA AND COMPUTATIONAL PROWESS. ITS IMPACT EXTENDS FAR BEYOND THE REALMS OF TECHNOLOGY, INFLUENCING HOW WE MAKE DECISIONS, SOLVE PROBLEMS, AND SHAPE THE WORLD AROUND US. AS WE STAND AT THE CROSSROADS OF INNOVATION, LET US TREAD FORWARD WITH MINDFULNESS, CURIOSITY, AND A COMMITMENT TO BUILDING A FUTURE WHERE MACHINE LEARNING ENRICHES, EMPOWERS, AND SERVES HUMANITY.