**Diabetes Prediction System; Data Mining and Bioinformatics**

**Synopsis of Data Mining**

Data mining simply is put is the process of examining massive amounts of data to identify patterns and trends (Alexandra Twin, 2022). Data mining is a very crucial component of analytics as it is able to detect underlying patterns and correlation between attributes of data, there are several algorithms/methods that can be used for data mining such as clustering, association rule mining, e.t.c which will be used in the process of diabetes prediction.

**Synopsis of Bioinformatics**

Bioinformatics is defined as the use of computational and analytical tools to gather and interpret biological data. With the growing amount of biological data and computer experts, data mining in conjunction of bioinformatics has made it possible to study data and generate actionable insights which is continually helping the medical industry thrive.

**Applications of Bioinformatics:**

* Gene therapy.
* Microbial applications.
* Discovery of new drugs, e.t.c

**Briefing of Diabetes**

Diabetes, one of the most dreaded terminal illnesses in the world is a disease that occurs when the body is unable to use the glucose gotten from digested food (CDC, 2022). Diabetes can also occur when the pancreas doesn’t produce enough insulin.

**Types of Diabetes:**

* *Type 1 Diabetes*: This type of diabetes is caused by an autoimmune reaction where the body mistakenly fights its own self therefore causing the body not to produce insulin.
* *Type 2 Diabetes:* This type of diabetes occurs when the body is not able to properly make use of insulin which causes the blood sugar levels to rise above usual.
* *Gestational Diabetes:* This type of diabetes is more common in pregnant women and goes away as soon as the child is born. Although it seems harmless, it is known to increase the risk factor of having type 2 diabetes later in life. (CDC, 2022)
* *Prediabetes:* In type 2 diabetes, the blood sugar level of an individual is higher than usual but not high enough to be classified as type 2 diabetes.

**Synopsis of the Diabetes Prediction System**

**Brief description:**

Diabetes patients used to have to take tests in order to detect if they are positive or negative but due to the advancement of Machine Learning with focus on Bioinformatics, Data Mining; expert systems can now look into past data and through heuristic search determine the status of an individual. A prediction system will be built to collect user data which will be processed by the expert system and the result will be returned to the user after inference with the inference engine.

**AIM:**

To build a diabetes prediction system that will predict an individual’s diabetes status.

**Objectives:**

* To develop a diabetes classification model.
* To design and integrate a web app.

**Methodology:**

* A classification model will be built using data gotten from Kaggle.com. The following steps will be taken in order to ensure the data is clean and model ready; data cleaning, EDA (Exploratory Data Analysis) will commence soon after which will encompass a broad range of analytical operations including outlier detection and removal, feature visualizations, feature extraction, selection and transformation will be carried out on the data in order to prepare it for model training. After the previous steps are complete, several attempts will be made to train models using various algorithms such as; Logistic regression, XGB classifier, Decision tree classifier followed by model testing and hyper-parameters tuning. As soon as these have all been concluded, the final model will be saved as a pickle file.
* The web app will be built using HTML, CSS & JavaScript for the front-end which will serve as an interface for users to get educated and tested. The data will be collected and sent to the backend which will be built with python (Flask). The backend through the aid of the model will infer, come up with the test result and return it to the front-end.

**REFERENCES**

C. (2022, July 7). *What is Diabetes?* Centers for Disease Control and Prevention. Retrieved November 12, 2022, from

<https://www.cdc.gov/diabetes/basics/diabetes.html>

Twin. (2022, August 2). *What Is Data Mining? How It Works, Benefits, Techniques, and Examples*. Investopedia. Retrieved November 12, 2022, from

<https://www.investopedia.com/terms/d/datamining.asp>

Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. *BMJ : British Medical Journal*, *324*(7344), 1018-1022.

<https://doi.org/10.1136/bmj.324.7344.1018>