**ICT 423 PROJECT DOCUMENTATION**

This project was undertaken by members of Group 1, Mechatronics 400 level students of Bells University of Technology for ICT 423. They are:

S/N Matric Number Fullname

1 2020/8975 OBOH JORDAN OGHENEFEGO

2 2020/8977 ODEY EMMANUEL OBULE

3 2020/8980 AKINLEKE FAVOUR OLAMIDE

4 2020/8988 ADEGBOLA FERANMI ALFRED

5 2020/8994 SULOLA LAWRENCE OWOMITOLA

6 2020/8995 OLORUNTOBA VICTOR AYOMIDEJI

7 2020/8996 LEKWA-UZO ONYEDIKACHI NAOMI

8 2020/9035 ADEGBOYEGA PRECIOUS OLUSEGUN

9 2020/9051 ILORI JOSHUA AYOMIDE

10 2020/9061 KEHINDE JOY ABIOLA

11 2020/9081 JESUSINA TEMILOLUWA DANIEL

12 2020/9082 OTTAILOBHEGBE SAMUEL OKHIRIA

13 2020/9089 BLESSING GIDEON OGHENEKOME

14 2020/9092 ODOCK NEPHI NCHOR

15 2020/9104 EJERE BESTON OJE

16 2020/9119 ADEBOWALE AYODEJI AUGUSTINE

17 2020/9122 OMO-ASIUGHWU ABRAHAM

18 2020/9126 ALABI OLUWAFOLAHAN DAVID

19 2020/9141 IFE ALPHA OLASUNKANMI

20 2020/9142 OBA HANEEFAH OLUWATOYIN

**PROJECT DEFINITION**

For the ICT 423 project, we will develop a binary classification model to predict whether a patient is diabetic based on relevant health features. This is to help healthcare professionals monitor and detect the disease as early as possible.

**DATA OVERVIEW**

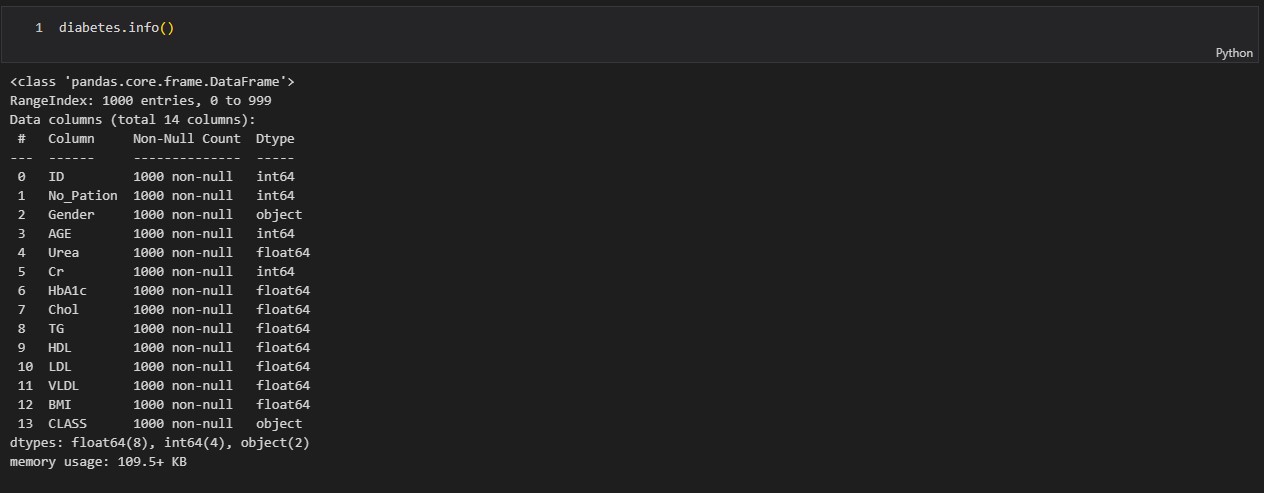
Diabetes is an opportunistic disease with several consequences and a multitude of available data. A more sophisticated and precise method of illness diagnosis is required.

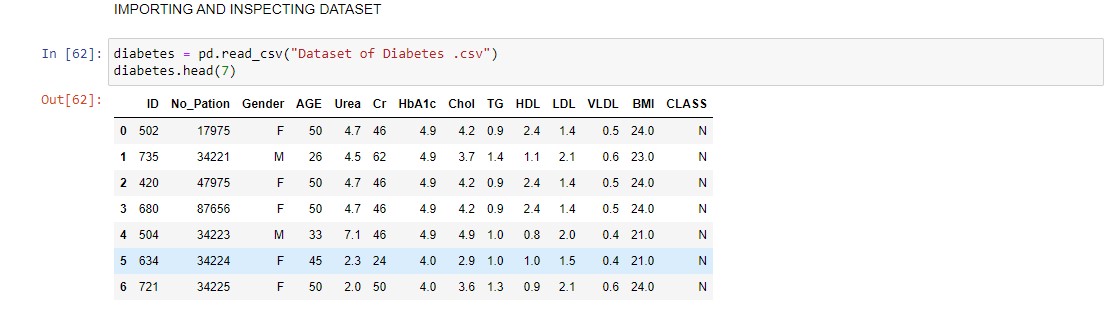
The data were collected from the Iraqi society, as the data were acquired from the laboratory of Medical City Hospital and (the Specializes Center for Endocrinology and Diabetes-Al-Kindy Teaching Hospital). Patients' files were taken and data was extracted from them and entered into the database to construct the diabetes dataset. The data consists of medical information and laboratory analysis. The data attributes are:

The data consist of medical information, laboratory analysis… etc. The data that have been entered initially into the system are The Patient, Sugar Level Blood, Age, Gender, Creatinine ratio(Cr), Body Mass Index (BMI), Urea, Cholesterol (Chol), Fasting lipid profile, including total, LDL, VLDL, Triglycerides(TG) and HDL Cholesterol, HBA1C, Class (the patient's diabetes disease class may be Diabetic(‘Y’), Non-Diabetic(‘N’), or Predict-Diabetic(‘P)).

The dataset contains:

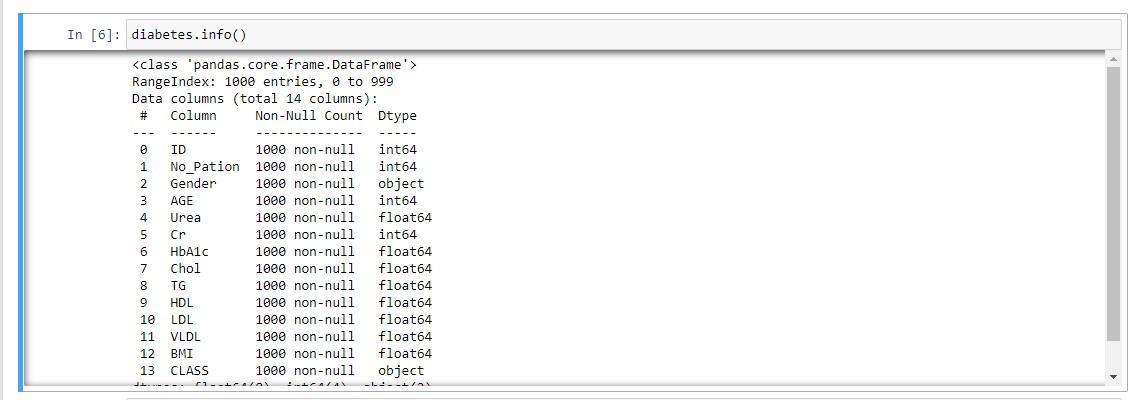
1. 1000 features from 14 feature categories: ID, Patient No, Age, Gender, Creatinine ratio(Cr), etc.
2. 1 target: Class





**DATA CLEANING/PREPROCESSING**

The data was inspected and it was observed that no duplicates were found and all the datatypes were as they should be.

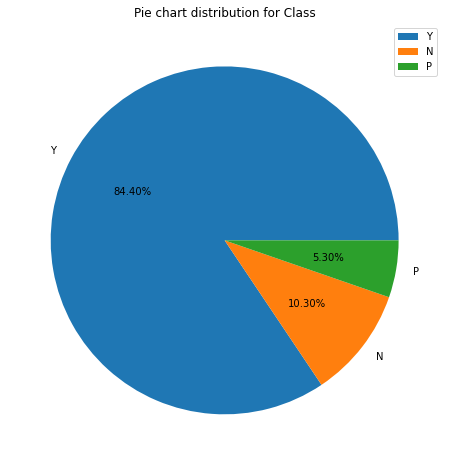
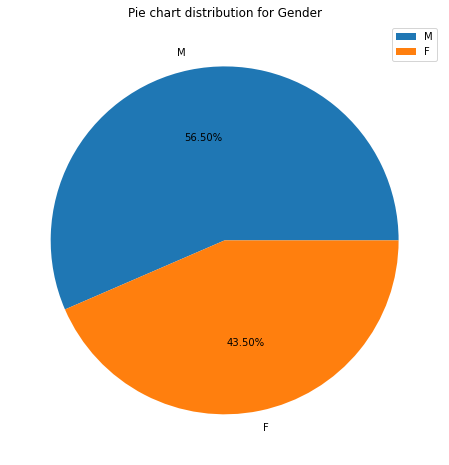
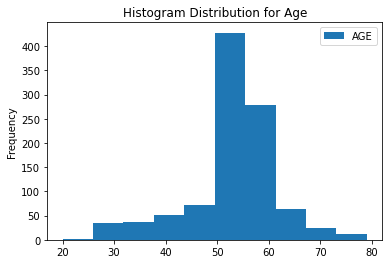


The columns, ‘Gender’ and ‘Class’ were inspected to ensure the correct information was there and there were three classes in Gender and 4 classes In the ‘Class’ column. These were corrected to have the appropriate information.



**EXPLORATORY DATA ANALYSIS**

The data was explored using visualizations to better understand the data. All categorical data was converted to numerical data. We were able to understand the proportion of the classes, gender and age in the dataset.

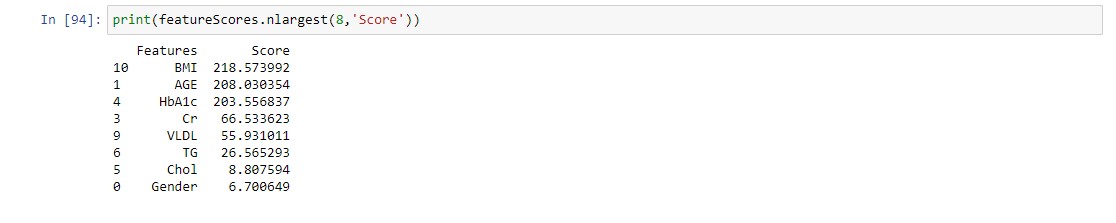


We also were able to see the correlation between the different features.

It was decided to remove the class of patients that had Predict-Diabetic(‘P’). The rows for this feature were removed from the dataset.

FEATURE SELECTION

The most important features were analyzed and had more impact on the model.



**MODEL: TRAINING AND VALIDATION**

The ML models used for this problem were Support Vector Machine and Random Forest Classifier.

**Random Forest Classifier**

Train Result:

================================================

Accuracy Score: 100.00%

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CLASSIFICATION REPORT:

0 1 accuracy macro avg weighted avg

precision 1.0 1.0 1.0 1.0 1.0

recall 1.0 1.0 1.0 1.0 1.0

f1-score 1.0 1.0 1.0 1.0 1.0

support 72.0 590.0 1.0 662.0 662.0

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Test Result:

================================================

Accuracy Score: 98.60%

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CLASSIFICATION REPORT:

0 1 accuracy macro avg weighted avg

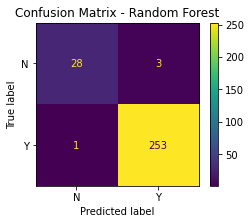
precision 0.965517 0.988281 0.985965 0.976899 0.985805

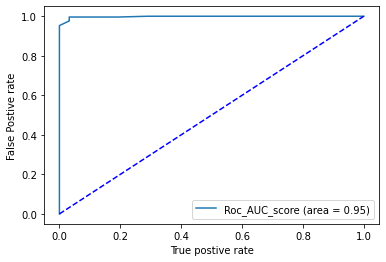
recall 0.903226 0.996063 0.985965 0.949644 0.985965

f1-score 0.933333 0.992157 0.985965 0.962745 0.985759

support 31.000000 254.000000 0.985965 285.000000 285.000000

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_





**Support Vector Machine Classifier**

Train Result:

================================================

Accuracy Score: 98.04%

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CLASSIFICATION REPORT:

0 1 accuracy macro avg weighted avg

precision 0.904110 0.989813 0.980363 0.946961 0.980492

recall 0.916667 0.988136 0.980363 0.952401 0.980363

f1-score 0.910345 0.988974 0.980363 0.949659 0.980422

support 72.000000 590.000000 0.980363 662.000000 662.000000

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Test Result:

================================================

Accuracy Score: 97.19%

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

CLASSIFICATION REPORT:

0 1 accuracy macro avg weighted avg

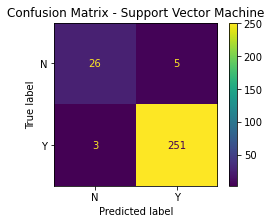
precision 0.896552 0.980469 0.97193 0.938510 0.971341

recall 0.838710 0.988189 0.97193 0.913449 0.971930

f1-score 0.866667 0.984314 0.97193 0.925490 0.971517

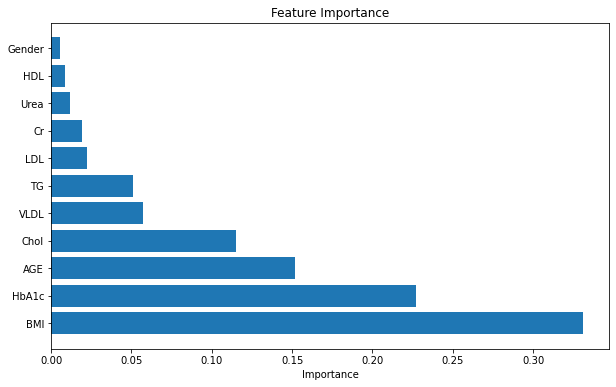
support 31.000000 254.000000 0.97193 285.000000 285.000000

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_



**RESULT INTERPRETATION**

Based on the results, we can get the features that were of the most importance.



The result of the project can be applied in the following ways:

These models can be deployed and applied to help the health sector. Healthcare can benefit from this project for diabetes prediction in several ways such as:

1. Early Detection and Diagnosis: The model can help in the early detection of diabetes by the analysis of noted important features. Early diagnosis allows for timely intervention and management, potentially preventing or delaying the onset of complications.
2. Risk Stratification: The models can stratify patients based on their risk of developing diabetes. This information can be used to prioritize interventions and resources for those at higher risk, leading to more efficient healthcare delivery.
3. Personalized Treatment Plans: The models can assist in developing personalized treatment plans based on individual patient characteristics. Tailored interventions may include lifestyle recommendations, medication choices, and monitoring strategies.
4. Resource Allocation: Healthcare providers can use these models to allocate resources more efficiently. For example, identifying patients at higher risk can help in allocating preventive services and targeted interventions to the individuals who need them the most.
5. Remote Patient Monitoring: These models can be integrated into remote monitoring systems, allowing continuous monitoring of patients' health parameters. This can help healthcare providers in assessing the effectiveness of treatment plans and making timely adjustments.
6. Patient Engagement and Education: The models can support patient engagement by providing personalized health insights and recommendations. Educational content and lifestyle advice can be tailored to individual patients based on their health status and risk factors.
7. Decision Support for Clinicians: The models can serve as decision-support tools for healthcare professionals. By analyzing patient data, these models can assist clinicians in making informed decisions regarding diagnosis, treatment, and follow-up care.
8. Public Health Planning: Aggregated data from the models can be used for public health planning and policy development. Identifying population-level trends and risk factors can inform public health strategies for diabetes prevention and management.
9. Clinical Trials and Research: The models can aid in patient selection for clinical trials and research studies. By identifying individuals with specific characteristics or at particular risk levels, researchers can recruit suitable participants for studies.
10. Continuous Improvement: The models can be updated and refined over time as more data becomes available. Continuous improvement ensures that the model stays relevant and accurate in predicting diabetes risk and outcomes.