Final project

45

Diabetes Analysis

Made by:

Youssef Ahmed

Abdelrahman Emad Ahmed

Abdelrahman Mohamed

Table of Contents

[ **Introduction to AI:** 3](#_Toc153223797)

[ **Types of artificial intelligence** 3](#_Toc153223798)

[ **Introduction to data set:** 4](#_Toc153223799)

[ **Data set attributes:** 4](#_Toc153223800)

[ **Data Analysis:** 5](#_Toc153223801)

[ **Image dataset** 9](#_Toc153223802)

[ **Training process:** 9](#_Toc153223803)

[ **Prediction process:** 9](#_Toc153223804)

Table of figures

[Figure 1 6](#_Toc153226130)

[Figure 2 7](file:///C:\Users\Abdelrhman%20emad\Desktop\orange%20word%205.docx#_Toc153226131)

[Figure 3 8](#_Toc153226132)

[Figure 4 8](file:///C:\Users\Abdelrhman%20emad\Desktop\orange%20word%205.docx#_Toc153226133)

[Figure 5 9](file:///C:\Users\Abdelrhman%20emad\Desktop\orange%20word%205.docx#_Toc153226134)

[Figure 6 9](file:///C:\Users\Abdelrhman%20emad\Desktop\orange%20word%205.docx#_Toc153226135)

[Figure 7 10](file:///C:\Users\Abdelrhman%20emad\Desktop\orange%20word%205.docx#_Toc153226136)

[Figure 8 10](file:///C:\Users\Abdelrhman%20emad\Desktop\orange%20word%205.docx#_Toc153226137)

# **Introduction to AI:**

Artificial intelligence is a machine that's able to learn, make decisions, and take actions even when it encounters a situation it has never come across before. (AI) is the hottest buzzword now, and almost every major company is adding some kind of AI features to its product or service.

## **Types of artificial intelligence**

1. Machine learning
2. Deep learning
3. Theory of mind
4. Expert system
5. Neural network
6. Computer vision

* **Machine learning**
* Machine learning is a branch of artificial intelligence  and computer science which focuses on the use of data and algorithms to imitate the way that humans learn, gradually improving its accuracy.
* Machine learning is an important component of the growing field of data science. Through the use of statistical methods, algorithms are trained to make classifications or predictions, and to uncover key insights in data mining projects. These insights subsequently drive decision making within applications and businesses, ideally impacting key growth metrics. As big data continues to expand and grow, the market demand for data scientists will increase. They will be required to help identify the most relevant business questions and the data to answer them.

# **Introduction to data set:**

The Diabetes prediction dataset is a collection of medical and demographic data from patients, along with their diabetes status (positive or negative). This dataset can be used to build machine learning models to predict diabetes in patients based on their medical history and demographic information. This can be useful for healthcare professionals in identifying patients who may be at risk of developing diabetes and in developing personalized treatment plans. Additionally, the dataset can be used by researchers to explore the relationships between various medical and demographic factors and the likelihood of developing diabetes.

## **Data set attributes:**

The data includes features such as age, gender, body mass index (BMI), hypertension, heart disease, smoking history, HbA1c level, and blood glucose level.

**Csv files processing:**

Figure

## A diagram of a software company Description automatically generated with medium confidence**Data Analysis:**

**Here we will show the steps we follow to process our data:**

1. Import the dataset as csv file.
2. We select our target.
3. Using impute help for substitutes any missing value using several methods to avoid losing any record.
4. Preprocess help for increasing the efficiency of analyzing the dataset and increasing the accuracy of prediction by making the data normalize from 0 to 1.
5. Outliers it removes the un useful data that affects the training and the prediction.
6. Data sampler it makes the dataset two groups one of them is the sample dataset and the other one is the remaining one is the remaining dataset.

* **A screenshot of a graph

  Description automatically generatedDistribution**

Figure 3

1. **It represents the featuers in charts**

**ThePercentge of the women and men**

**having diabetes.**

1. **It shows that women are higher than**

**men in having diabetes .**

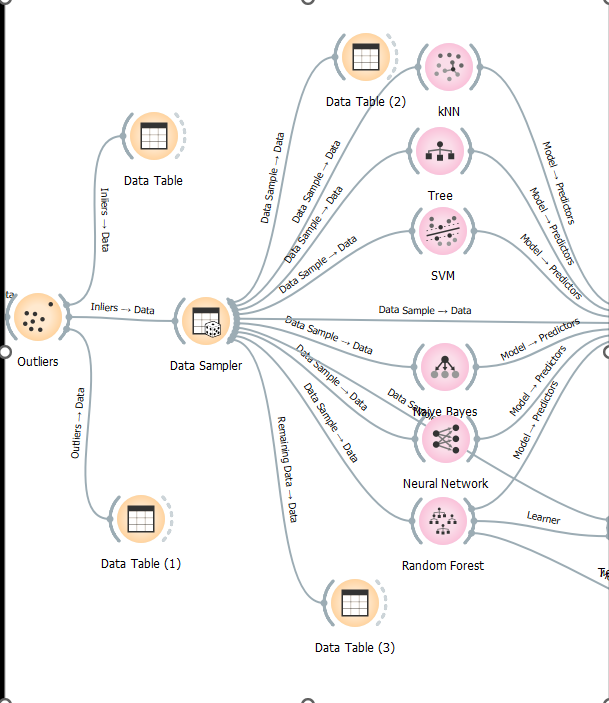
**A graph of blue and red lines

Description automatically generated**

1. **Here it show the ages that having diabetes.**
2. **The higher age that has diabetes is age of 80 years old.**
3. **The lower age that has diabetes is the age between 5 and 11.**

Figure 2

* **Data modelling:**

****

* **Training the dataset:**
* It’s important to train the dataset to different models to take the most accurate model to make the training on.
* Here random forest has the highest CA with 0.978 so it’s the best model to train our dataset on for more accurate prediction.

Figure

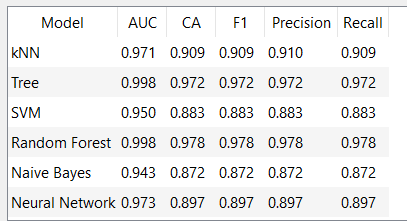
****

Figure 4

* **Data prediction:**

After finishing the training and saving the model, we have to test a new dataset. So, we uploaded a new dataset which we can predict using the file that we have trained.

A diagram of a model

Description automatically generatedA calendar with numbers and a few days of the week

Description automatically generated

Figure 5

Figure 6

# **Image dataset**

## **Training process:**

A diagram of a diagram

Description automatically generatedAfter finding a good dataset to be used in training process I used the Import Image widget to import the images to orange and then used image embedding widget to evaluate the images and after this I used different methods (SVM, Random Forest ,kNN, Logistic Regression, Neural Network) to train the data set and used test and score to know the best training method in them by looking at the highest CA which was for Logistic Regression method by score 0.894 in CA and in the end I used the save model widget to save the trained data.

Figure 7

## **Prediction process:**

After Training the data and saving the training model comes the prediction process, I used import images widget to put uncategorized images and connected it to image embedding widget to evaluate the images and then used load model to load the saved train model and connected it and the image embedding widget to prediction widget to predict them and connected the it to image grid widget to see the images and its prediction

A diagram of a diagram

Description automatically generated

Figure 8

* References:

1. <https://zapier.com/blog/what-is-ai/?utm_source=google&utm_medium=cpc&utm_campaign=gaw-row-nua-evr-search_nb_desktop_blog_prospecting_developing1_developing2-ads&utm_adgroup=DSA-BestApps-Form_Builder&utm_term=&utm_content=21468&gad_source=1&gclid=Cj0KCQiA4NWrBhD-ARIsAFCKwWvNKYf0hZHH38vZQ6ERMCVZGC1fKT93VXaoJXfBRy_Im-CezKAmfXMaAnaiEALw_wcB#what>
2. <https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset>
3. <https://www.ibm.com/topics/machine-learning>