Introduction

This assignment is initiated because of the Week 3 ANN workshop at Fontys UAS.

Objective

Construct, train and test an artificial neural network using a dataset of your own choice. Try different settings for two or more hyperparameters and investigate the effect on learning. Write a Jupyter notebook which contains your python code and in which you describe your approach and results. In your notebook, you should describe your dataset and add a reference to the source of your dataset. Also, include references to any source code or tutorials that you used to write your code. If your neural network is aimed at classification, you should create a confusion matrix and discuss the results. Also reflect on the knowledge and skills you acquired on artificial neural networks.

Assignment Idea

The goal of this assignment is to test how to work with Artificial Neural Networks (ANN). The dataset chosen for this is taken from the Kaggle's website.

Assignment Goal

Diabetes is one of the major diseases of the population across the world. Diabetes is a chronic disease that occurs either when the pancreas does not produce enough insulin or when the body cannot efficiently use the insulin it produces. In 2014, 8.5% of adults aged 18 years and older had diabetes. In 2012, diabetes was the direct cause of 1.5 million deaths and high blood glucose was the cause of another 2.2 million deaths. Over the time, diabetes can damage the heart, blood vessels, eyes, kidneys, and nerves. Early diagnosis can be made through a relatively inexpensive method of computation. In this notebook the ANN model is used to analyze and make the diabetes prediction model.



The Dataset

In this dataset, all patients are females at least 21 years old of Pima Indian heritage. The total number of instances is 768, which is completely used in this study. It contains 8 attributes plus one class (Label) column. Each attribute is numeric-valued; attributes of this dataset are as follows:

- 1. Number of times pregnant
- 2. Plasma glucose concentration at 2 h in an oral glucose tolerance test
- 3. Diastolic blood pressure (mm Hg)
- 4. Triceps skinfold thickness (mm)
- 5. 2-hour serum insulin (mu U/ml)
- 6. Body mass index (weight in kg/(height in m2)
- 7. Diabetes pedigree function

```
8. Age (years)
```

9. Class variable (0 or 1).

(Class value 1 is interpreted as "tested positive for diabetes").

Import Libraries

```
import pandas as pd
import numpy as np
import tensorflow as tf
import seaborn as sns
import matplotlib.pyplot as plt
import warnings;
warnings.simplefilter('ignore')
from tensorflow import keras
from tensorflow.keras import layers
from tensorflow.keras.layers import Input, Dense
from tensorflow.keras.models import Sequential
from sklearn.model_selection import train_test_split
import matplotlib.pyplot as plt
%matplotlib inline
```

Import the data

```
In [2]: df = pd.read_csv("diabetes.csv")
In [3]: df.head(10)
```

Out[3]:		Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	вмі	DiabetesPedigreeFunction	Age	Outcome
	0	6	148	72	35	0	33.6	0.627	50	1
	1	1	85	66	29	0	26.6	0.351	31	0
	2	8	183	64	0	0	23.3	0.672	32	1
	3	1	89	66	23	94	28.1	0.167	21	0

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	ВМІ	DiabetesPedigreeFunction	Age	Outcome
4	0	137	40	35	168	43.1	2.288	33	1
5	5	116	74	0	0	25.6	0.201	30	0
6	3	78	50	32	88	31.0	0.248	26	1
7	10	115	0	0	0	35.3	0.134	29	0
8	2	197	70	45	543	30.5	0.158	53	1
9	8	125	96	0	0	0.0	0.232	54	1

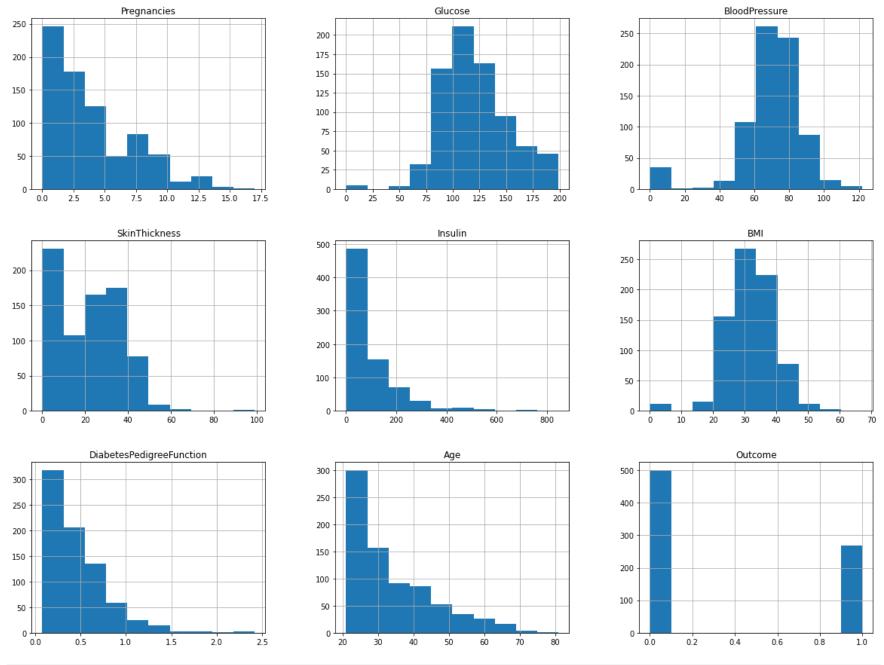
I can already notice that there are null values in the form of zero's.

```
In [4]:
          df.isnull().sum()
Out[4]: Pregnancies
         Glucose
         BloodPressure
         SkinThickness
         Insulin
         BMI
         DiabetesPedigreeFunction
         Age
         Outcome
                                      0
         dtype: int64
        The values show that we do not have any missing data
```

df.info()

```
In [5]:
         df.dtypes
Out[5]: Pregnancies
                                       int64
         Glucose
                                       int64
         BloodPressure
                                       int64
        SkinThickness
                                       int64
        Insulin
                                       int64
         BMI
                                     float64
        DiabetesPedigreeFunction
                                     float64
                                       int64
         Age
        Outcome
                                       int64
        dtype: object
In [6]:
```

```
<class 'pandas.core.frame.DataFrame'>
        RangeIndex: 768 entries, 0 to 767
        Data columns (total 9 columns):
             Column
                                        Non-Null Count Dtype
             Pregnancies
                                        768 non-null
                                                        int64
         1
             Glucose
                                        768 non-null
                                                        int64
             BloodPressure
                                        768 non-null
                                                        int64
             SkinThickness
                                        768 non-null
                                                        int64
         4
             Insulin
                                        768 non-null
                                                        int64
         5
             BMI
                                        768 non-null
                                                        float64
             DiabetesPedigreeFunction 768 non-null
                                                        float64
         7
                                        768 non-null
                                                        int64
             Age
         8
             Outcome
                                        768 non-null
                                                        int64
        dtypes: float64(2), int64(7)
        memory usage: 54.1 KB
In [7]:
         # I'll take a look at the distribution of the different parameters
         df.hist(figsize=(20,15))
Out[7]: array([[<AxesSubplot:title={'center':'Pregnancies'}>,
                 <AxesSubplot:title={'center':'Glucose'}>,
                 <AxesSubplot:title={'center':'BloodPressure'}>],
                [<AxesSubplot:title={'center':'SkinThickness'}>,
                 <AxesSubplot:title={'center':'Insulin'}>,
                 <AxesSubplot:title={'center':'BMI'}>],
                [<AxesSubplot:title={'center':'DiabetesPedigreeFunction'}>,
                 <AxesSubplot:title={'center':'Age'}>,
                 <AxesSubplot:title={'center':'Outcome'}>]], dtype=object)
```



In [8]:
 plt.figure(1 , figsize = (16 , 8))
 cor = sns.heatmap(df.corr(), annot = True)



Insights: There appears to be no strong correlations among the numeric features.

However, there are some slight correlations of Pregnancies, Glucose, BMI, Age with the outcome.

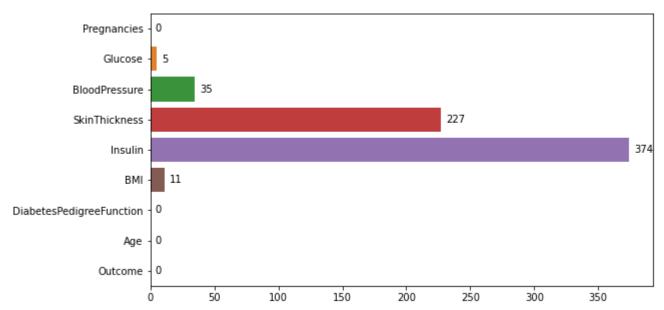
Check for missing values

```
SkinThickness 227
Insulin 374
BMI 11
dtype: int64
```

- Important Observations:
 - As I mentioned above, tt seems like null values are present in the form of zero's.
 - It's impossible to have Glucose, Blood Pressure, SkinThickness, Insulin and BMI to be zero.

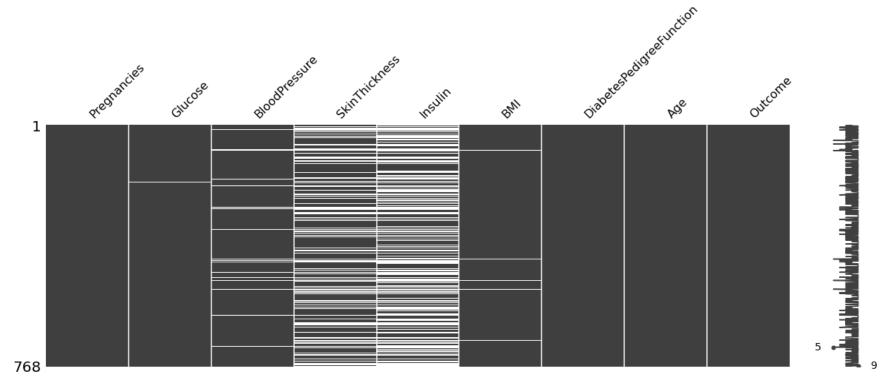
```
In [11]: df.describe().T
```

```
Out[11]:
                                                                                  25%
                                                                                            50%
                                                                                                       75%
                                      count
                                                   mean
                                                                 std
                                                                        min
                                                                                                               max
                         Pregnancies
                                       768.0
                                                3.845052
                                                            3.369578
                                                                       0.000
                                                                               1.00000
                                                                                          3.0000
                                                                                                    6.00000
                                                                                                              17.00
                                       763.0 121.686763
                                                                      44.000
                                                                              99.00000
                                                                                       117.0000
                                                                                                  141.00000 199.00
                             Glucose
                                                           30.535641
                       BloodPressure
                                      733.0
                                               72.405184
                                                           12.382158 24.000 64.00000
                                                                                         72.0000
                                                                                                   80.00000 122.00
                       SkinThickness
                                       541.0
                                               29.153420
                                                           10.476982
                                                                       7.000
                                                                              22.00000
                                                                                         29.0000
                                                                                                   36.00000
                                                                                                              99.00
                              Insulin
                                       394.0
                                             155.548223
                                                          118.775855
                                                                      14.000
                                                                              76.25000
                                                                                        125.0000
                                                                                                  190.00000
                                                                                                             846.00
                                BMI
                                       757.0
                                               32.457464
                                                            6.924988
                                                                      18.200
                                                                             27.50000
                                                                                         32.3000
                                                                                                   36.60000
                                                                                                              67.10
           DiabetesPedigreeFunction
                                       768.0
                                                0.471876
                                                                       0.078
                                                                               0.24375
                                                                                          0.3725
                                                                                                    0.62625
                                                                                                               2.42
                                                            0.331329
                                                                                         29.0000
                                                                                                   41.00000
                                Age
                                       768.0
                                               33.240885
                                                           11.760232 21.000
                                                                              24.00000
                                                                                                              81.00
                            Outcome
                                       768.0
                                                0.348958
                                                            0.476951
                                                                       0.000
                                                                               0.00000
                                                                                          0.0000
                                                                                                    1.00000
                                                                                                               1.00
```



```
import missingno as mno
mno.matrix(df, figsize = (20, 6))
```

Out[13]: <AxesSubplot:>



From the above plot, we can notice that there are missing values in the Glucose, BloodPressure, SkinThickness, Insulin and BMI.

We can see there are lot of null values in SkinThickness and Insulin column. So, after Imputation the mean will change drastically.

```
In [14]:
            #imputing mean instead of null values
            for col in df:
                df[col].replace(np.nan, df[col].mean(), inplace=True)
In [15]:
            df.describe().T
Out[15]:
                                                                             25%
                                                                                         50%
                                                                                                    75%
                                    count
                                                mean
                                                            std
                                                                   min
                                                                                                            max
                       Pregnancies
                                     768.0
                                             3.845052
                                                        3.369578
                                                                  0.000
                                                                           1.00000
                                                                                     3.000000
                                                                                                 6.000000
                                                                                                           17.00
                                                                                                          199.00
                           Glucose
                                     768.0
                                           121.686763
                                                      30.435949
                                                                 44.000
                                                                          99.75000
                                                                                   117.000000
                                                                                              140.250000
                     BloodPressure
                                                                                    72.202592
                                                                                                          122.00
                                     768.0
                                            72.405184
                                                      12.096346
                                                                 24.000
                                                                          64.00000
                                                                                               80.000000
```

25.00000

29.153420

32.000000

99.00

29.153420

8.790942

7.000

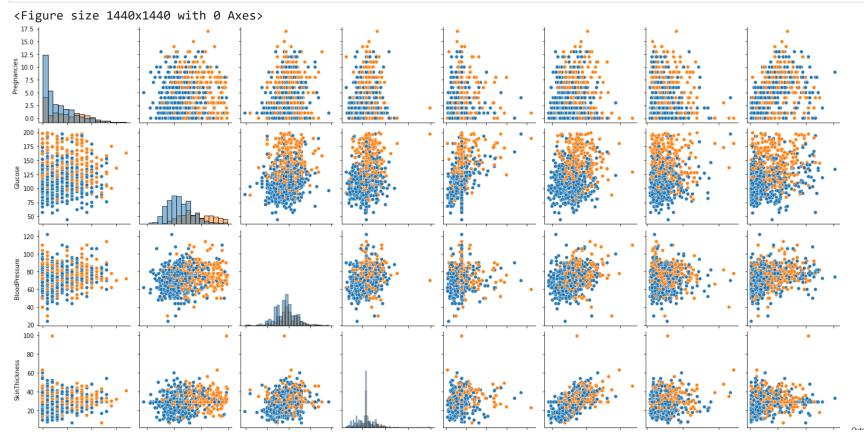
768.0

SkinThickness

	count	mean	std	min	25%	50%	75%	max
Insulin	768.0	155.548223	85.021108	14.000	121.50000	155.548223	155.548223	846.00
ВМІ	768.0	32.457464	6.875151	18.200	27.50000	32.400000	36.600000	67.10
DiabetesPedigreeFunction	768.0	0.471876	0.331329	0.078	0.24375	0.372500	0.626250	2.42
Age	768.0	33.240885	11.760232	21.000	24.00000	29.000000	41.000000	81.00
Outcome	768.0	0.348958	0.476951	0.000	0.00000	0.000000	1.000000	1.00

I can now notice that the minimum value for insulin has increased.

#Plot pairwise relationships in a dataset
plt.figure(figsize=(20,20))
sns.pairplot(data=df, hue="Outcome", diag_kind="hist")
plt.show()





localhost:8889/nbconvert/html/Desktop/ML projects/Diabetes prediction using ANN/Diabetes Prediction ANN.ipynb?download=false

Pregnancies Glucose BloodPressure SkinThickness

Out[18]:

	Pregnancies	Glucose	BloodPressure	SkinThickness	Insulin	ВМІ	DiabetesPedigreeFunction	Age	Outcome
0	6	148.0	72.0	35.00000	155.548223	33.6	0.627	50	1
1	1	85.0	66.0	29.00000	155.548223	26.6	0.351	31	0
2	8	183.0	64.0	29.15342	155.548223	23.3	0.672	32	1
3	1	89.0	66.0	23.00000	94.000000	28.1	0.167	21	0
4	0	137.0	40.0	35.00000	168.000000	43.1	2.288	33	1

Modeling

In [22]:

Splitting the dataset

```
In [19]: X=df.drop(["Outcome"],axis='columns') # dropping the target variable
y=df["Outcome"] # keeping the target variable

In [20]: from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20, random_state=42)
```

Standardize the data

from sklearn.utils import shuffle

```
In [21]: # Import `StandardScaler` from `sklearn.preprocessing`
from sklearn.preprocessing import StandardScaler

# Define the scaler
scaler = StandardScaler().fit(X_train)

# Scale the train set
X_train = scaler.transform(X_train)

# Scale the test set
X_test = scaler.transform(X_test)
```

```
X train, y train = shuffle(X train,y train, random state=14)
In [23]:
       print("Train data length:",len(X train))
       print("Test data length:",len(X test))
       X train.shape[1]
       Train data length: 614
       Test data length: 154
Out[23]: 8
In [24]:
       # Initializing the model
       model = keras.Sequential([
           keras.layers.Dense(64, input shape=(8,), activation='relu'),
           keras.layers.Dense(32, activation='relu'),
           keras.layers.Dense(16, activation='relu'),
           keras.layers.Dense(1, activation='sigmoid')
       1)
In [25]:
       # Compiling the model
       model.compile(optimizer='adam', loss='binary crossentropy', metrics=['accuracy'])
In [26]:
       # Fitting the model to the Training set
       hist = model.fit(X train, y train, epochs=100, validation split=0.2)
       Epoch 1/100
       y: 0.7236
       Epoch 2/100
       0.7236
       Epoch 3/100
       16/16 [============= - 0s 2ms/step - loss: 0.4871 - accuracy: 0.7699 - val loss: 0.5159 - val accuracy:
       0.7154
       Epoch 4/100
       16/16 [============= ] - 0s 2ms/step - loss: 0.4608 - accuracy: 0.7841 - val loss: 0.5035 - val accuracy:
       0.7236
       Epoch 5/100
       16/16 [============== ] - 0s 3ms/step - loss: 0.4417 - accuracy: 0.7943 - val loss: 0.4947 - val accuracy:
       0.7317
       Epoch 6/100
```

```
0.7317
Epoch 7/100
0.7480
Epoch 8/100
16/16 [============] - 0s 3ms/step - loss: 0.4121 - accuracy: 0.7943 - val loss: 0.4836 - val accuracy:
0.7398
Epoch 9/100
16/16 [============== ] - 0s 3ms/step - loss: 0.4050 - accuracy: 0.7963 - val loss: 0.4802 - val accuracy:
0.7561
Epoch 10/100
0.7642
Epoch 11/100
0.7561
Epoch 12/100
16/16 [=============] - 0s 3ms/step - loss: 0.3887 - accuracy: 0.8004 - val loss: 0.4800 - val accuracy:
0.7561
Epoch 13/100
0.8167 - val loss: 0.4869 - val accuracy: 0.7561
Epoch 14/100
16/16 [=============] - 0s 2ms/step - loss: 0.3772 - accuracy: 0.8126 - val loss: 0.4821 - val accuracy:
0.7561
Epoch 15/100
0.7724
Epoch 16/100
0.7724
Epoch 17/100
0.7724
Epoch 18/100
0.7561
Epoch 19/100
16/16 [============] - 0s 3ms/step - loss: 0.3610 - accuracy: 0.8187 - val loss: 0.4896 - val accuracy:
0.7642
Epoch 20/100
0.7642
Epoch 21/100
0.7561
Epoch 22/100
0.7805
```

```
Epoch 23/100
0.7805
Epoch 24/100
16/16 [============] - 0s 3ms/step - loss: 0.3373 - accuracy: 0.8330 - val loss: 0.5084 - val accuracy:
0.7724
Epoch 25/100
0.7642
Epoch 26/100
0.7561
Epoch 27/100
0.7724
Epoch 28/100
0.7724
Epoch 29/100
0.7642
Epoch 30/100
0.7724
Epoch 31/100
0.7724
Epoch 32/100
0.7724
Epoch 33/100
0.7642
Epoch 34/100
0.7805
Epoch 35/100
16/16 [============] - 0s 3ms/step - loss: 0.2887 - accuracy: 0.8534 - val loss: 0.5672 - val accuracy:
0.7642
Epoch 36/100
16/16 [============] - 0s 3ms/step - loss: 0.2844 - accuracy: 0.8737 - val loss: 0.5623 - val accuracy:
0.7724
Epoch 37/100
0.7642
Epoch 38/100
0.7642
Epoch 39/100
```

```
16/16 [============= ] - 0s 3ms/step - loss: 0.2681 - accuracy: 0.8839 - val loss: 0.5911 - val accuracy:
0.7561
Epoch 40/100
16/16 [============== ] - 0s 3ms/step - loss: 0.2588 - accuracy: 0.8880 - val loss: 0.6131 - val accuracy:
0.7561
Epoch 41/100
0.7724
Epoch 42/100
16/16 [============] - 0s 3ms/step - loss: 0.2514 - accuracy: 0.8819 - val loss: 0.6206 - val accuracy:
0.7724
Epoch 43/100
0.7561
Epoch 44/100
0.7642
Epoch 45/100
16/16 [============= ] - 0s 3ms/step - loss: 0.2319 - accuracy: 0.8982 - val loss: 0.6614 - val accuracy:
0.7480
Epoch 46/100
16/16 [============] - 0s 3ms/step - loss: 0.2259 - accuracy: 0.8982 - val loss: 0.6742 - val accuracy:
0.7561
Epoch 47/100
16/16 [============== ] - 0s 3ms/step - loss: 0.2252 - accuracy: 0.9063 - val loss: 0.6739 - val accuracy:
0.7805
Epoch 48/100
0.7805
Epoch 49/100
0.7317
Epoch 50/100
0.7642
Epoch 51/100
16/16 [============] - 0s 2ms/step - loss: 0.2039 - accuracy: 0.9104 - val_loss: 0.7536 - val_accuracy:
0.7398
Epoch 52/100
0.7480
Epoch 53/100
0.7561
Epoch 54/100
0.7724
Epoch 55/100
16/16 [============= ] - 0s 3ms/step - loss: 0.1804 - accuracy: 0.9226 - val loss: 0.7957 - val accuracy:
```

```
0.7398
Epoch 56/100
16/16 [============== ] - 0s 3ms/step - loss: 0.1722 - accuracy: 0.9287 - val loss: 0.7810 - val accuracy:
0.7642
Epoch 57/100
16/16 [============= ] - 0s 3ms/step - loss: 0.1660 - accuracy: 0.9348 - val loss: 0.8168 - val accuracy:
0.7398
Epoch 58/100
16/16 [============== ] - 0s 3ms/step - loss: 0.1640 - accuracy: 0.9246 - val loss: 0.8354 - val accuracy:
0.7480
Epoch 59/100
0.7480
Epoch 60/100
0.7642
Epoch 61/100
0.7480
Epoch 62/100
16/16 [============] - 0s 3ms/step - loss: 0.1434 - accuracy: 0.9450 - val_loss: 0.8734 - val_accuracy:
0.7480
Epoch 63/100
16/16 [============== ] - 0s 3ms/step - loss: 0.1381 - accuracy: 0.9450 - val loss: 0.9113 - val accuracy:
0.7642
Epoch 64/100
16/16 [============= ] - 0s 3ms/step - loss: 0.1302 - accuracy: 0.9532 - val loss: 0.9106 - val accuracy:
0.7398
Epoch 65/100
0.7561
Epoch 66/100
0.7236
Epoch 67/100
0.7154
Epoch 68/100
16/16 [============] - 0s 3ms/step - loss: 0.1178 - accuracy: 0.9633 - val loss: 0.9830 - val accuracy:
0.7480
Epoch 69/100
16/16 [============== ] - 0s 3ms/step - loss: 0.1154 - accuracy: 0.9674 - val loss: 1.0217 - val accuracy:
0.7236
Epoch 70/100
0.7398
Epoch 71/100
0.7398
```

```
Epoch 72/100
16/16 [============== ] - 0s 2ms/step - loss: 0.1017 - accuracy: 0.9756 - val loss: 1.0644 - val accuracy:
0.7317
Epoch 73/100
16/16 [============] - 0s 2ms/step - loss: 0.0990 - accuracy: 0.9796 - val loss: 1.0849 - val accuracy:
0.7480
Epoch 74/100
16/16 [=============] - 0s 2ms/step - loss: 0.0945 - accuracy: 0.9796 - val loss: 1.1072 - val accuracy:
0.7317
Epoch 75/100
0.7154
Epoch 76/100
0.7317
Epoch 77/100
0.7561
Epoch 78/100
0.7236
Epoch 79/100
16/16 [============= ] - 0s 2ms/step - loss: 0.0807 - accuracy: 0.9857 - val loss: 1.2393 - val accuracy:
0.7317
Epoch 80/100
16/16 [=============] - 0s 2ms/step - loss: 0.0803 - accuracy: 0.9776 - val loss: 1.1870 - val accuracy:
0.7561
Epoch 81/100
0.7154
Epoch 82/100
0.7480
Epoch 83/100
0.7317
Epoch 84/100
16/16 [============] - 0s 2ms/step - loss: 0.0680 - accuracy: 0.9898 - val loss: 1.2619 - val accuracy:
0.7561
Epoch 85/100
16/16 [============] - 0s 2ms/step - loss: 0.0655 - accuracy: 0.9898 - val loss: 1.2966 - val accuracy:
0.7154
Epoch 86/100
16/16 [============= ] - 0s 2ms/step - loss: 0.0628 - accuracy: 0.9919 - val loss: 1.3034 - val accuracy:
0.7561
Epoch 87/100
0.7236
Epoch 88/100
```

```
0.7398
Epoch 89/100
0.7398
Epoch 90/100
0.7480
Epoch 91/100
16/16 [============] - 0s 2ms/step - loss: 0.0540 - accuracy: 0.9878 - val loss: 1.4081 - val accuracy:
0.7480
Epoch 92/100
0.7398
Epoch 93/100
0.7480
Epoch 94/100
16/16 [============= ] - 0s 3ms/step - loss: 0.0472 - accuracy: 0.9919 - val loss: 1.4631 - val accuracy:
0.7317
Epoch 95/100
0.7561
Epoch 96/100
16/16 [============] - 0s 2ms/step - loss: 0.0417 - accuracy: 0.9959 - val loss: 1.5240 - val accuracy:
0.7480
Epoch 97/100
0.7398
Epoch 98/100
0.7480
Epoch 99/100
16/16 [============] - 0s 3ms/step - loss: 0.0382 - accuracy: 0.9959 - val_loss: 1.5438 - val_accuracy:
0.7398
Epoch 100/100
16/16 [============= ] - 0s 3ms/step - loss: 0.0358 - accuracy: 0.9980 - val loss: 1.5431 - val accuracy:
0.7480
```

In [27]:

9/15/22, 10:36 AM

```
print(model.summary())
model.evaluate(X_test, y_test)
```

Model: "sequential"

Layer (type)	Output Shape	Param #
dense (Dense)	(None, 64)	576

(None, 32)

dense 1 (Dense)

```
dense 2 (Dense)
                                  (None, 16)
                                                         528
        dense 3 (Dense)
                                  (None, 1)
        ______
        Total params: 3,201
        Trainable params: 3,201
        Non-trainable params: 0
        None
        5/5 [============ ] - 0s 1ms/step - loss: 1.4977 - accuracy: 0.6494
Out[27]: [1.497668743133545, 0.649350643157959]
In [28]:
         from sklearn.metrics import confusion matrix , classification report
         yp = model.predict(X test)
         y_pred = []
         for element in yp:
            if element > 0.5:
                y_pred.append(1)
            else:
                y pred.append(0)
         print(classification report(y test,y pred))
```

2080

	precision	recall	f1-score	support
0	0.72	0.74	0.73	99
1	0.51	0.49	0.50	55
accuracy			0.65	154
macro avg	0.62	0.61	0.61	154
weighted avg	0.65	0.65	0.65	154

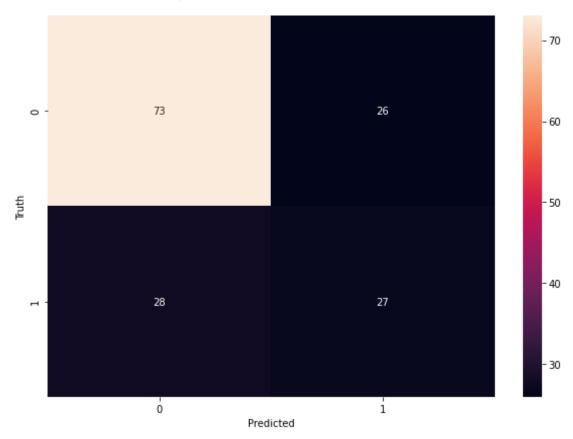
Precision – What percent of your predictions were correct? -> 76%

Recall – What percent of the positive cases did you catch? -> 73%

```
import seaborn as sn
cm = tf.math.confusion_matrix(labels=y_test,predictions=y_pred)

plt.figure(figsize = (10,7))
sn.heatmap(cm, annot=True, fmt='d')
plt.xlabel('Predicted')
plt.ylabel('Truth')
```

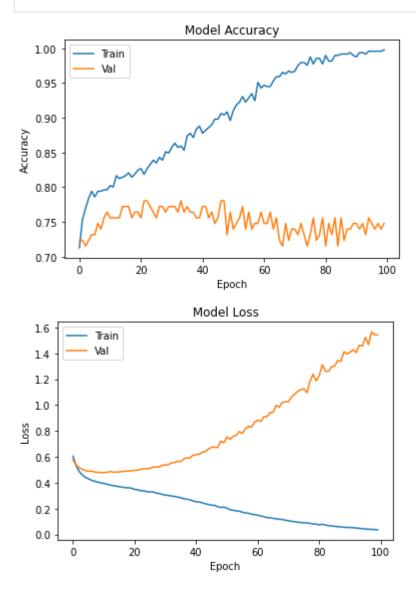
```
Out[29]: Text(69.0, 0.5, 'Truth')
```



```
In [30]:
#accuracy
plt.plot(hist.history['accuracy'])
plt.plot(hist.history['val_accuracy'])
plt.title('Model Accuracy')
plt.xlabel("Epoch")
plt.ylabel("Accuracy")
plt.legend(['Train','Val'],loc='upper left')
plt.show()

#Loss
plt.plot(hist.history['loss'])
plt.plot(hist.history['val_loss'])
plt.title('Model Loss')
plt.xlabel("Epoch")
```

```
plt.ylabel("Loss")
plt.legend(['Train','Val'],loc='upper left')
plt.show()
```



```
# Evaluating the model on the Test set
model.evaluate(X_test, y_test)
```

Model Altering

```
In [33]:
     #create a keras model
     model2 = Sequential()
     #once model object is created from Sequential we need layers to add
     model2.add(Dense(12,input dim=8,activation='relu'))
     model2.add(Dense(12, activation='relu'))
     model2.add(Dense(8,activation='relu'))
     model2.add(Dense(4,activation='relu'))
     model2.add(Dense(1,activation='sigmoid'))
In [34]:
     #compile the model by providing parameters
     model2.compile(loss='binary crossentropy', optimizer='adam',metrics=['accuracy'])
In [35]:
     model2.fit(X train,y train,epochs=250, batch size=15)
     Epoch 1/250
     Epoch 2/250
     Epoch 3/250
     Epoch 4/250
     Epoch 5/250
     Epoch 6/250
     Epoch 7/250
```

```
Epoch 8/250
Epoch 9/250
Epoch 10/250
Epoch 11/250
Epoch 12/250
Epoch 13/250
Epoch 14/250
Epoch 15/250
Epoch 16/250
Epoch 17/250
Epoch 18/250
Epoch 19/250
Epoch 20/250
Epoch 21/250
Epoch 22/250
Epoch 23/250
Epoch 24/250
Epoch 25/250
Epoch 26/250
Epoch 27/250
Epoch 28/250
Epoch 29/250
Epoch 30/250
Epoch 31/250
```

```
Epoch 32/250
Epoch 33/250
Epoch 34/250
Epoch 35/250
Epoch 36/250
Epoch 37/250
Epoch 38/250
Epoch 39/250
Epoch 40/250
Epoch 41/250
Epoch 42/250
Epoch 43/250
Epoch 44/250
Epoch 45/250
Epoch 46/250
Epoch 47/250
Epoch 48/250
Epoch 49/250
Epoch 50/250
Epoch 51/250
Epoch 52/250
Epoch 53/250
Epoch 54/250
Epoch 55/250
Epoch 56/250
```

9/15/22, 10:36 AM

```
Epoch 57/250
Epoch 58/250
Epoch 59/250
Epoch 60/250
41/41 [============== - 0s 1000us/step - loss: 0.3703 - accuracy: 0.8143
Epoch 61/250
Epoch 62/250
Epoch 63/250
Epoch 64/250
Epoch 65/250
Epoch 66/250
Epoch 67/250
Epoch 68/250
Epoch 69/250
Epoch 70/250
Epoch 71/250
Epoch 72/250
Epoch 73/250
Epoch 74/250
Epoch 75/250
Epoch 76/250
Epoch 77/250
Epoch 78/250
Epoch 79/250
Epoch 80/250
```

```
Epoch 81/250
Epoch 82/250
Epoch 83/250
Epoch 84/250
Epoch 85/250
Epoch 86/250
Epoch 87/250
Epoch 88/250
Epoch 89/250
Epoch 90/250
Epoch 91/250
Epoch 92/250
Epoch 93/250
Epoch 94/250
Epoch 95/250
Epoch 96/250
Epoch 97/250
Epoch 98/250
Epoch 99/250
Epoch 100/250
Epoch 101/250
Epoch 102/250
Epoch 103/250
Epoch 104/250
Epoch 105/250
```

```
Epoch 106/250
Epoch 107/250
Epoch 108/250
Epoch 109/250
Epoch 110/250
Epoch 111/250
Epoch 112/250
Epoch 113/250
Epoch 114/250
Epoch 115/250
Epoch 116/250
Epoch 117/250
Epoch 118/250
Epoch 119/250
Epoch 120/250
Epoch 121/250
Epoch 122/250
Epoch 123/250
Epoch 124/250
Epoch 125/250
Epoch 126/250
Epoch 127/250
Epoch 128/250
Epoch 129/250
```

```
Epoch 130/250
Epoch 131/250
Epoch 132/250
Epoch 133/250
Epoch 134/250
Epoch 135/250
Epoch 136/250
Epoch 137/250
Epoch 138/250
Epoch 139/250
Epoch 140/250
Epoch 141/250
Epoch 142/250
Epoch 143/250
Epoch 144/250
Epoch 145/250
Epoch 146/250
Epoch 147/250
Epoch 148/250
Epoch 149/250
Epoch 150/250
Epoch 151/250
Epoch 152/250
Epoch 153/250
Epoch 154/250
```

```
Epoch 155/250
Epoch 156/250
Epoch 157/250
Epoch 158/250
Epoch 159/250
Epoch 160/250
Epoch 161/250
Epoch 162/250
41/41 [===========] - ETA: 0s - loss: 0.3963 - accuracy: 0.80 - 0s 1ms/step - loss: 0.3014 - accuracy:
0.8583
Epoch 163/250
Epoch 164/250
Epoch 165/250
Epoch 166/250
Epoch 167/250
Epoch 168/250
Epoch 169/250
Epoch 170/250
Epoch 171/250
Epoch 172/250
Epoch 173/250
Epoch 174/250
Epoch 175/250
Epoch 176/250
Epoch 177/250
Epoch 178/250
```

```
Epoch 179/250
Epoch 180/250
Epoch 181/250
Epoch 182/250
Epoch 183/250
Epoch 184/250
Epoch 185/250
Epoch 186/250
Epoch 187/250
Epoch 188/250
Epoch 189/250
Epoch 190/250
Epoch 191/250
Epoch 192/250
Epoch 193/250
Epoch 194/250
Epoch 195/250
Epoch 196/250
Epoch 197/250
Epoch 198/250
Epoch 199/250
Epoch 200/250
Epoch 201/250
Epoch 202/250
```

```
Epoch 203/250
Epoch 204/250
Epoch 205/250
Epoch 206/250
Epoch 207/250
Epoch 208/250
Epoch 209/250
Epoch 210/250
Epoch 211/250
Epoch 212/250
Epoch 213/250
Epoch 214/250
Epoch 215/250
Epoch 216/250
Epoch 217/250
Epoch 218/250
Epoch 219/250
Epoch 220/250
Epoch 221/250
6 - ETA: 0s - loss: 0.282
Epoch 222/250
Epoch 223/250
Epoch 224/250
Epoch 225/250
Epoch 226/250
```

```
Epoch 227/250
Epoch 228/250
Epoch 229/250
Epoch 230/250
Epoch 231/250
Epoch 232/250
Epoch 233/250
Epoch 234/250
Epoch 235/250
Epoch 236/250
Epoch 237/250
Epoch 238/250
Epoch 239/250
Epoch 240/250
Epoch 241/250
Epoch 242/250
Epoch 243/250
Epoch 244/250
Epoch 245/250
Epoch 246/250
Epoch 247/250
Epoch 248/250
Epoch 249/250
Epoch 250/250
```

9/15/22, 10:36 AM Diabetes Prediction ANN Out[35]: <tensorflow.python.keras.callbacks.History at 0x1d75d231820>

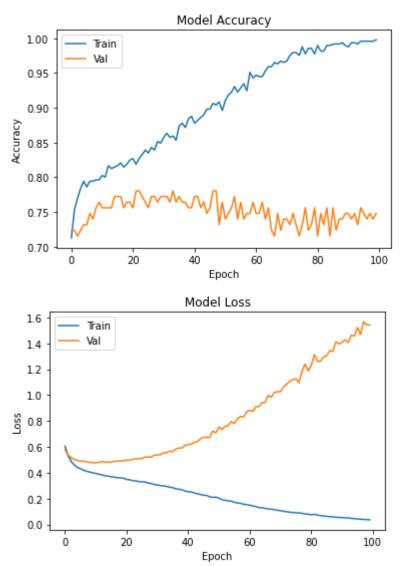
```
Out[35]:
In [36]:
        print(model2.summary())
        model2.evaluate(X test, y test)
        Model: "sequential_1"
        Layer (type)
                               Output Shape
                                                    Param #
        ______
        dense 4 (Dense)
                               (None, 12)
                                                    108
        dense 5 (Dense)
                               (None, 12)
                                                    156
        dense 6 (Dense)
                               (None, 8)
                                                    104
        dense 7 (Dense)
                               (None, 4)
                                                    36
       dense 8 (Dense)
                               (None, 1)
        _____
        Total params: 409
        Trainable params: 409
        Non-trainable params: 0
        None
        Out[36]: [0.9631482362747192, 0.6428571343421936]
In [37]:
        from sklearn.metrics import confusion matrix , classification report
        yp = model2.predict(X test)
        y pred = []
        for element in yp:
            if element > 0.5:
               y pred.append(1)
            else:
               y pred.append(0)
        print(classification report(y test,y pred))
                   precision
                             recall f1-score
                                             support
                0
                       0.73
                               0.70
                                       0.72
                                                 99
                1
                       0.50
                               0.55
                                       0.52
                                                 55
                                       0.64
                                                154
           accuracy
                       0.62
                               0.62
                                       0.62
                                                154
          macro avg
```

weighted avg 0.65 0.64 0.65 154

9/15/22, 10:36 AM

Training the first model resulted in a bit higher accuracy. However, the classification report metrics are better after altering the model. I can notice that the precision and the recall are higher than they were.

```
In [38]:
          print(hist.history.keys())
         dict keys(['loss', 'accuracy', 'val loss', 'val accuracy'])
In [39]:
          #accuracy
          plt.plot(hist.history['accuracy'])
          plt.plot(hist.history['val accuracy'])
          plt.title('Model Accuracy')
          plt.xlabel("Epoch")
          plt.ylabel("Accuracy")
          plt.legend(['Train','Val'],loc='upper left')
          plt.show()
          #Loss
          plt.plot(hist.history['loss'])
          plt.plot(hist.history['val_loss'])
          plt.title('Model Loss')
          plt.xlabel("Epoch")
          plt.ylabel("Loss")
          plt.legend(['Train','Val'],loc='upper left')
          plt.show()
```



```
In [41]: # Predicting
yp = model2.predict(X_test)
```

```
yp[:5]
```

9/15/22, 10:36 AM

Conclusion

Using the Artificial Neural Networks(ANN) model can design and implement the complex medical processes by software. The software systems are more effective and efficient in various medical fields including predicting, diagnosing, treating and helping the surgeons and physicians, and the general population. These systems can be implemented in a parallel way and are distributed in different scales. In general, the Artificial Neural Networks are parallel processing systems that are used to detect complex patterns in the data. The aim of this notebook was to determine effective variables and their impact on diabetes and estimating whether a neural network model can predict diabetes. The results of training the model are not that high for now, as it could be seen on the above figures. The first model consists of 4 layers. The activation function chosen was ReLU for the first layer, ReLU for the two hidden layers and sigmoid for the output layer. The "Altering model" was with 5 layers. The input layer was a ReLU and the other 3 hidden layers were ReLU and the output layer was a segmoid.

References

- [1] Srivastava, S., Sharma, L., Sharma, V., Kumar, A., & Darbari, H. (2019). Prediction of Diabetes Using Artificial Neural Network Approach. SpringerLink. Retrieved September 15, 2022, from https://link.springer.com/chapter/10.1007/978-981-13-1642-5_59? error=cookies_not_supported&code=8c1f8171-1caf-4d1e-a1d0-df1fa6dbc5ff#:%7E:text=Among%20several%20algorithms%20of%20Machine,with%20the%20sample%20test%20data.
- [2] An Improved Artificial Neural Network Model for Effective Diabetes Prediction. (2021, April 22). Retrieved September 15, 2022, from https://www.hindawi.com/journals/complexity/2021/5525271/
- [3] Team, K. (n.d.). Keras: the Python deep learning API. Retrieved September 15, 2022, from https://keras.io/
- [4] Rune, L. P. W. (2021, October 19). Diabetes Classification with Neural Network | Machine Learning with Python (10 h) | Lesson 06 (0-14). YouTube. Retrieved September 15, 2022, from https://www.youtube.com/watch?v=BYY8eSJkJtA&feature=youtu.be

- [5] seaborn: statistical data visualization seaborn 0.12.0 documentation. (n.d.). Retrieved September 15, 2022, from https://seaborn.pydata.org/
- [6] Artificial Intelligence Neural Networks. (n.d.). Retrieved September 15, 2022, from https://www.tutorialspoint.com/artificial_intelligence/artificial_intelligence_neural_networks.htm
- [7] Understanding the Classification report through sklearn Muthukrishnan. (2018, July 7). Retrieved September 15, 2022, from https://muthu.co/understanding-the-classification-report-in-sklearn/