THEORY OF COMPUTATION

BITE306L

DIGITAL ASSIGNMENT - 1

SUBMITTED BY,

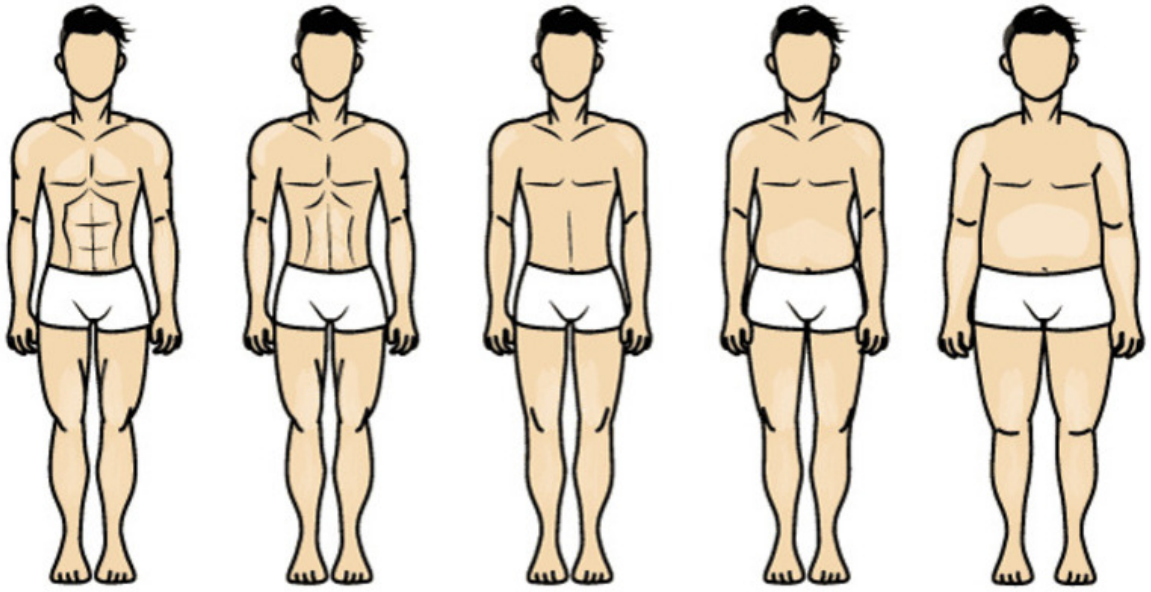
KALIRAJ.A - 21BIT0687

GUIDED BY,

Dr. CHANDRASEGAR.T,
Department of Computer Applications,
School of Information Technology and Engineering,
VIT University, Vellore ,
Tamil Nadu - 632014

DATASET NAME - Body Fat Prediction Dataset

DATASET LINK - https://www.kaggle.com/datasets/fedesoriano/body-fat-prediction-dataset



DATASET DESCRIPTION:

---> Lists estimates of the percentage of body fat determined by underwater weighing and various body circumference measurements for 200 men.

---> This data set can be used to illustrate multiple regression techniques. Accurate measurement of body fat is inconvenient/costly and it is desirable to have easy methods of estimating body fat that are not inconvenient/costly.

---> The variables listed below, from left to right, are:

1)Density determined from underwater weighing
2)Percent body fat from Siri's (1956) equation
3)Age (years)
4)Weight (lbs)
5)Height (inches)
6)Neck circumference (cm)
7)Chest circumference (cm)
8)Abdomen 2 circumference (cm)
9)Hip circumference (cm)
10)Thigh circumference (cm)
11)Knee circumference (cm)
12)Ankle circumference (cm)
13)Biceps (extended) circumference (cm)
14)Forearm circumference (cm)
15)Wrist circumference (cm)

---> Number of columns : 15

---> Number of Rows : 200

---> Data Set Characteristics : Multivarite

---> Associated Tasks : Classification

---> Area : Medical - Health

References:

Bailey, Covert (1994). Smart Exercise: Burning Fat, Getting Fit, Houghton-Mifflin Co., Boston, pp. 179-186.

Behnke, A.R. and Wilmore, J.H. (1974). Evaluation and Regulation of Body Build and Composition, Prentice-Hall, Englewood Cliffs, N.J.

Siri, W.E. (1956), "Gross composition of the body", in Advances in Biological and Medical Physics, vol. IV, edited by J.H. Lawrence and C.A. Tobias, Academic Press, Inc., New York.

Katch, Frank and McArdle, William (1977). Nutrition, Weight Control, and Exercise, Houghton Mifflin Co., Boston.

Wilmore, Jack (1976). Athletic Training and Physical Fitness: Physiological Principles of the Conditioning Process, Allyn and Bacon, Inc., Boston.

In [8]:

```
import pandas as pd
from sklearn.model_selection import train_test_split
from pandas.plotting import scatter_matrix
import matplotlib.pyplot as plt
from sklearn import model_selection
from sklearn import metrics
from sklearn.metrics import classification_report
from sklearn.metrics import confusion_matrix
from sklearn.metrics import accuracy_score
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
import matplotlib.pyplot as plt
import seaborn as sns

print("KALIRAJ.A - 21BIT0687")
data = pd.read_csv('bodyfat3.csv')

# Split-out validation dataset
array = data.values
X = array[:,0:199]
y = array[:,14]


feature_cols = ['Density','BodyFat','Age','Weight','Height','Neck','Chest','Andomen','Hip

# Split dataset into training set and test set
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3) # 70% training a

#Create a Gaussian Classifier
clf=RandomForestClassifier(n_estimators=100)

#Train the model using the training sets y_pred=clf.predict(X_test)
clf.fit(X_train,y_train)
y_pred=clf.predict(X_test)

# Model Accuracy, how often is the classifier correct?
print("Accuracy:",metrics.accuracy_score(y_test, y_pred))

RandomForestClassifier(bootstrap=True, class_weight=None, criterion='gini',
            max_depth=None, max_features='auto', max_leaf_nodes=None,
            min_impurity_decrease=0.0, min_impurity_split=None,
            min_samples_leaf=1, min_samples_split=2,
            min_weight_fraction_leaf=0.0, n_estimators=100, n_jobs=1,
            oob_score=False, random_state=None, verbose=0,
            warm_start=False)
feature_imp = pd.Series(clf.feature_importances_,index=feature_cols).sort_values(ascendin
feature_imp
%matplotlib inline

# Creating a bar plot
plt.figure()
sns.barplot(x=feature_imp, y=feature_imp.index)

# Add labels to your graph
plt.xlabel('Feature Importance Score')
plt.ylabel('Features')
plt.title("Visualizing Important Features in dataset - Bodyfat")
```

```
plt.show()
```

```
KALIRAJ.A - 21BIT0687
Accuracy: 0.9333333333333333
```