**B. Tech. III (CSE) Semester – V**

**MACHINE LEARNING (CORE - 11)**

**CS305**

**Practical 1. Implement classification and regression techniques.**

1. **Regression**
   1. **Linear Regression**

*Dataset*: iris.csv

The iris(the flower) dataset attached contains five variables namely,

1. SepalLength(Cm)
2. SepalWidth(Cm)
3. PetalLength(Cm)
4. PetalWidth(Cm)
5. Species

All you need to do is performing Linear Regression in python on this dataset taking Sepal Length as Response or dependent variable and rest of the variables as independent ones.

Before performing Linear Regression, please check

1. whether there exists any multicollinearity in the independent variables with correlation matrix and suitable scatter plots.
2. Find the correlation between dependent variable and each independent variable.
3. Find if there is any outlier in the variables given with suitable boxplots.
   1. **Logistic Regression**

1. *Title of Database:* Abalone data (abalone.csv)

2. *Sources*:

(a) Original owners of database:

Marine Resources Division

Marine Research Laboratories - Taroona

Department of Primary Industry and Fisheries, Tasmania

GPO Box 619F, Hobart, Tasmania 7001, Australia

(contact: Warwick Nash +61 02 277277, wnash@dpi.tas.gov.au)

(b) Donor of database:

Sam Waugh (Sam.Waugh@cs.utas.edu.au)

Department of Computer Science, University of Tasmania

GPO Box 252C, Hobart, Tasmania 7001, Australia

(c) Date received: December 1995

3. *Number of Instances*: 4177

4. *Number of Attributes*: 8

5. *Attribute information*:

Given is the attribute name, attribute type, the measurement unit and a

brief description

Name Data Type Meas. Description

---- --------- ----- -----------

Sex nominal M, F, and I (infant)

Length continuous mm Longest shell measurement

Diameter continuous mm perpendicular to length

Height continuous mm with meat in shell

Whole weight continuous grams whole abalone

Shucked weight continuous grams weight of meat

Viscera weight continuous grams gut weight (after bleeding)

Shell weight continuous grams after being dried

Rings integer +1.5 gives the age in years

6. From the above description of the dataset predict the Sex of abalone using Logistic Regression Classifier and make notebook.

1. **Classification**

*Dataset*: Diabetes data.csv

Please perform Classification Analysis using Decision Tree & Random Forest Classifier on the diabetes dataset attached.

* Split the data into train & test set.
* The data set may contain missing values so check before diving into applying the algos.
* Please visualize the classification report using ROC & AUC plot