

**Faculty of Engineering and Technology**

**Computer Science Department**

**COMP338**

**Project 2**

**Project Description**

Use R to predict some parameters using (Decision Tree or Linear Regression). Given the input data (see below), each student can choose any column as Target Value (e.g. internet provider, family income, etc.), and use other columns as feature to predict the chosen target value. Each student should use at least 20 columns as features.

**Input data**: Students will be given real data about 8000 Palestinian Families. The data was collected using a survey called “The Palestinian Expenditure and Consumption Survey” in 2011 and 2014. The survey contains 100s of fields. Students can download all files in one package that contains the data itself (spss format), documentation, survey etc., Notice that spss files can be opened using R.

Caution (تحذير): This data is private and students are required to not share this data with anyone, and must delete this data after finishing the project.

**Deliverable:** each student will submit a report (PDF file) via Ritaj. The PDF should include: 1 page introduction about his/her project, 1-2 pages describing the Target Value and selected features, and 2-5 pages screenshots of the results, confusion matrix, accuracy, figures, etc.

**Evaluation**: the discussion of the project (with the instructor and the TA) will include:

Explaining the code, and answering practical and theoretical questions. Students will be also evaluated based on the complexity of the Target Value they selected, how they solve it, the parameters they used, quality of the report, and their answers during the discussion of the project.

The dataset can be download from the shared project folder:

<https://www.dropbox.com/sh/2vcwh4j21433nnv/AAB5FP_Uf5o1DULZssQIkGWka?dl=0>

This folder contains the dataset and **8 short videos**. Each student is required to watch ALL videos in the shared project folder, in order to learn how to use R and how configure the parameters