



## COMSATS University Islamabad, Lahore Campus

Course Title:	Machine Learning	Course Code:	CSC668 - PCS 716
Resource Person:	Dr. Muhammad Sharjeel	Marks:	10

### **Assignment 2:**

**Due Date/Time: Sunday, 6<sup>th</sup> October, 23:59**

**Submission: Upload the assignment solution to your GitHub account (private repository).**

### **Instructions:**

- Upload your ARFF file(s) and report (.pdf) on GitHub.
- The name of the report file should be your-roll-number\_assignment2 (e.g., fa24-rs-001\_assignment2)
- Late submissions: 3% penalty per day for 3 days after due date.
- The 'names' dataset is available in your Google Drive shared folder.

**The aim of this assignment is to understand the complete ML pipeline.**

1. Hand crafted features:
  - Extract as many input features as you can by manually observing the text, i.e., the names of people. Create ARFF file(s). You can save the features together as a set of input features or separately one feature per file.  
*Hint: Remember the input feature(s) is the key to getting a good performance from the classifier.*
  - Convert the output feature, i.e., + and – symbols to their numeric equivalent (1 and 0).
2. ML experiments in WEKA:
  - Once you have the ARFF file(s) ready, load it into the WEKA's workbench.
  - View different characteristics of the data (WEKA's main window). If you notice anything interesting about the dataset, record it.
  - Run the j48 classification algorithm and observe/record the results.
3. Write a paragraph about your experience of working with the standard ML pipeline in your own words.