



# COMSATS University Islamabad

## Department of Computer Science

### Assignment NO: 4

**COURSE: Introduction to Data Science**

**PROGRAM: BSCS**

**CLASS & SECTION: 7A & 7B**

**DATE: 21<sup>st</sup> Dec, 2022**

**Mapped to CLO-4**

**Question 1) Solve the below given case studies using the concepts of ethics in data science.**

Part a) A company uses a machine-learning algorithm to determine which job advertisement to display for users searching for technology jobs. Based on past results, the algorithm tends to display lower-paying jobs for women than for men (after controlling for other characteristics than gender).

**What ethical considerations might be considered when reviewing this algorithm?**

Part b) A researcher is interested in the relationship of weather to sentiment (positivity or negativity of posts) on Twitter. They want to scrape data from <https://www.wunderground.com> and join that to Tweets in that geographic area at a particular time. One complication is that Weather Underground limits the number of data points that can be downloaded for free using their API (application program interface). The researcher sets up six free accounts to allow them to collect the data they want in a shorter time-frame.

**What ethical guidelines are violated by this approach to data scraping?**

**Part c)** In the United States, most students apply for grants or subsidized loans to finance their college education. Part of this process involves filling in a federal government form called the Free Application for Federal Student Aid (FAFSA). The form asks for information about family income and assets. The form also includes a place for listing the universities to which the information is to be sent. The data collected by FAFSA includes confidential financial information (listing the schools eligible to receive the information is effectively giving permission to share the data with them).

It turns out that the order in which the schools are listed carries important information. Students typically apply to several schools, but can attend only one of them. Until recently, admissions offices at some universities used the information as an important part of their models of whether an admitted student will accept admissions. The earlier in a list a school appears, the more likely the student is to attend that school. Here's the catch from the student's point of view. Some institutions use statistical models to allocate grant aid (a scarce resource) where it is most likely to help ensure that a student enrolls. For these schools, the more likely a student is deemed to accept admissions, the lower the amount of grant aid they are likely to receive.

**Is this ethical? Discuss.**



