

# CSCE 580 Artificial Intelligence

## Quiz 1

Shruti Jadhav

Q1: About data for AI

a) What is open data? Given an example of open data that you produce which others can use?

- Open data is data that is made available to everyone to be reused without any restrictions.
- Example of open data produced by me that can be used by everyone is data collected on daily traffic patterns on my way to and from campus. Since I travel everyday along the same routes, I can produce data such as date-time of travel, travel times, traffic congestion, accidents, as well as time spent on certain intersections. This kind of data, once uploaded publicly, can be used to inform urban/traffic planning on the specific route being traveled.

b) You are analyzing a dataset and some attributes are missing.

b.1) What could be any 2 reasons why they are missing?

2 reasons why attributes can be missing is because:

- the values were not provided or
- the value does not exist or has no practical meaning for the context in which analysis is being performed.

b.2) What are any 2 ways you can still proceed with data analysis despite the missing values. For each, mention what assumption you are making and what are its risks.

Two ways data analysis can be done are:

1. Ignoring the missing value. This technique is called omission, and we assume that the data point for the value does not exist and continue analysis using other values. The risk of this technique is that it will reduce the available data set, and if there is already a limited data set then the resulting analysis may be less comprehensive.

2. Imputing a new value. This technique is called imputation, and we either find a median for the missing value or assign it a default. The risk in this technique is that this can fail to capture the real world behavior of the captured phenomena, and can lead to some bias or generalization if the data set is small.

Q2: Programing activity: resume analysis- **DONE IN GITHUB**

This analysis has to be done as a python notebook or collab. It is saved as “<>/Quiz1/Q2-code.ipynb”.

Q3: Programming activity: data analysis for social impact **DONE IN GITHUB**

Write python code/ demonstrate its working in notebook, and report on the following questions along with your code. . It is saved as “<>/Quiz1/Q3-code.ipynb”.