Homework 1-0

Please answer the following questions based on your reading in Chapter 1 of the “Doing Data Science” text. Brief, but complete answers are best (25 point each, 100 points total).

1. In your own words, briefly explain the meaning of the term “Datafication”. What has made it possible? Provide an example not used in the text. Mention one problem you can associate with the process of “datification”.

Datafication is the process of taking all aspects of life and turning them into data that can then be analyzed. For example, whenever we search the web we are being datafied through cookies. This is now possible due our storage capabilities, sensors, cameras, or Google glasses. All this technology creates new data that data scientists can use for analysis. An example of datafication I have read up about is a company that offers Big Data as a Service; they take twitter feeds and can examine this information for trends such as finding shows that are popular. I think the biggest problem would have to be the issue of security, all of this datafication is leading to many security issues. For example recently Pokemon Go was accessing people’s google accounts, and even had full access to some business accounts.

1. Provide a definition of the term “Data Science” in your own words. Why is it considered a science? Mention three tasks a data scientist typically performs.

Data Science is a field that requires process and systems to understand and incur new information on pre-existing data. This is considered a science since one can ask a question. This means there is a certain hypothesis or question being tested. Thus, we find data to support our claims and prove it, making this a science. The three skills required to be a data science includes: statistics (traditional analysis you’re used to thinking about), data munging (parsing, scraping, and formatting data), and data visualizations (graphs, tools, etc.).

1. “Big Data” is about more than just the size of data sets. In your own words, explain what is meant by the term “Big Data”. Focus your response of what the concept means, not on a quantitative or technical definition. Provide an example not mentioned in the text.

Big data is considered “big data” when the context of the problems you’re trying to solve allow models, algorithms, and include ethical decisions on how the model can affect the world. Big data is data that can be further mined to find new information using preexisting information. For example, a company can use big data to understand efficiency gains or cost saving since it helps provide deeper insights with more data points to analyze.

1. Consider the statement: “You need big data to do data science”. Briefly explain why this statement is false.

This is false since all data can be mined further for further insights, meaning all data can be big data if analyzed in a specific manner using statistics and computer algorithms.