

Introduction to Data Science

Lesson 8: What's next with data science

Activity 1: Data science roles and ethical issues

For this activity you will consider an example of data science in more detail. You can either choose one of the examples you explored in activity 1 of lesson 1 (where you chose two services or organisations that collect data about you) or you can choose an example described by a practising data scientist from one of the *Meet a Data Scientist* videos.

Answer all of the questions below for the example that you have chosen.

1. What activity would a data scientist perform for each of these stages:
 - a. Data sources
 - Where would the data be collected from?*
 - How would the data be collected?*
 - How much data will be collected?*
 - How frequently will the data be updated?*
 - b. Data pre-processing
 - What errors, inconsistencies or omissions might there be in the data?*
 - How will the data be stored and will it need to be pre-processed before it can be analysed?*
 - c. Data exploration/analysis
 - What basic charts or statistical summaries could be produced to check the data?*
 - What groups could the data be split into to help with the analysis?*
 - d. Machine learning
 - What patterns are being searched for in the data?*
 - How is the data being used to make a prediction or a decision?*
 - e. Decision making
 - Who/what will be making decisions based on this data?*
 - What will happen as a result of the analysis/machine learning?*
2. In question 1d you suggested some patterns that machine learning would search for in the data. How accurate would you expect the models to be?

You could comment on the fact that this is using past data to predict future data. You could also comment on whether there are any specific issues related to the data collection that could affect the accuracy of any predictions?
3. What are the ethical issues relating to this example?

Is it appropriate to collect the data?
Is it appropriate to store the data?
Who benefits from the use of this data?