SMART Skills Python: Course Handbook

Dr. Kirils Makarovs

February 2022

The demand for high-level numeracy skills and knowledge of statistical and computational methods is ever increasing. As organisations are having to adapt to a world that is becoming ever more data-driven, particularly in the age of 'big data', they look to recruit those with data skills. This is an issue that is even shared by the academic research community as fields that were traditionally non-quantitative, such as archaeology and the humanities, are becoming increasingly compelled by data. The pace at which this demand has developed has resulted in a noticeable skills gap within the UK, where individuals who wish to pursue any one of a large number of careers find themselves unable to do so due to a lack of data analysis experience being offered to them.

The SMART Skills programme aims to address this gap. It does so by offering the opportunity for individuals to gain a theoretical understanding of statistical and computational methods, as well as develop practical skills of data preprocessing, analysis, and visualisation through an intense – but thoroughly supportive – training course.

This SMART Skills 'Introduction to text analysis in Python' course is designed to acquaint students with approaches to analyzing text as data using Python programming language. Over five days, students will learn how to preprocess text data, summarize its contents, visualize textual information, and run more complex types of analysis using Python via Jupyter Notebooks.

Why take this course?

Text data surrounds us in our lives and comes in different shapes and sizes, e.g. newspaper articles, tweets, product reviews, song lyrics, etc. While it might seem at first glance that this information can hardly be summarized and compared, certain computational techniques allow extracting meaningful information from text data. The topic of the document, its sentiment, the relationship between various named entities – this is to name a few examples of what sort of information one can get from text by applying relevant data analysis techniques.

This course provides the foundations for you to understand, execute and communicate text data analysis in a widely recognised software platform that was built for data analysis. Over five days, you will gain valuable skills that you can market to employers, gain confidence in your ability to work with data, and create a knowledge base that you can build on for years to come.

Learning Outcomes

This course is aimed at individuals with some prior experience or knowledge of statistical methods and data analysis techniques. However, it is expected that you will not have used Python to implement these, or at the least, you won't have received formal text analysis training in Python. Through a mix of formal teaching and practical exercises, at the end of this course you will:

- Have an understanding of a reproducible research workflow in Python via Jupyter Notebooks
- Be able to prepare raw text data for analysis
- Be able to perform descriptive text analysis
- Be able to run more complex types of text data analysis e.g. sentiment analysis
- Be confident in communicating the results of your analysis
- Have a greater understanding of the research process

In addition, you will gain a number of transferable skills that will enhance your employability. During the course, you will develop the following key skills:

- Coding in Python
- Data preprocessing
- Data analysis and visualisation
- Collaborative project work

The delivery of the course takes place over five sessions of two and a half hours each, and they will be a mix of lectures and guided practical exercises; the final session will be a group project where you produce a data-focused report.

Course Content

The software used throughout this course is Python, which is a very flexible and user-friendly open-source programming language that can handle various data analysis and statistical tasks. We will work in Python via Jupyter Notebooks provided by Google Colab, which is a free environment for interactive computing. To access Google Colab, you need to have a Google (Gmail and Drive) account.

Day	Topic
Session 1	Introduction to Jupyter Notebooks workflow, descriptive text analysis
Session 2	Text data preprocessing – stemming, lemmatisation

Session 3	Approaches to quantifying text data
Session 4	Computational models on text data e.g. sentiment analysis
Session 5	Group report project

Contact

For further information on the SMART Skills courses, please contact Dr. Kirils Makarovs at k.makarovs@exeter.ac.uk or Chloe Harvey at c.harvey3@exeter.ac.uk