

Data Management Plan for Research Students

1. Overview

Researcher: Ama N, Karthik S, Tanvi M, Miguel P.
Project title: Banking Customer Credit-Score Bracket Classification
Project duration: 15 Weeks
Project context: <i>This project lies in the field of finance, and the main objective is to classify people into credit score brackets through the use of machine learning algorithms based on their bank details and credit-related information.</i>

2. Defining your data/research sources

2.1 Where will your data/research sources come from? <i>The dataset utilized for this project has been obtained from kaggle (linked down below), and it is a public domain dataset with no copyright.</i> https://www.kaggle.com/datasets/parisrohan/credit-score-classification
2.2 How often will you get new data? <i>As stated in the dataset description linked in the box above, the expected update frequency is set at "Never". However, given the size of the dataset (50.000 rows), we might consider samplig techniques to test the Machine Learning Algorithms on different sets of stances.</i>

2.3 How much data/information will you generate?

Try to state this in kB/MB/GB

How much have you got so far?

Try to estimate how this will grow for the rest of the project

2.4 What file formats will you use?

What software is required to access the data? Are free/open alternatives available?

What type of data does each format hold?

3. Organising your data

3.1 How will you structure and name your folders and files?

Are there any set or recommended standards in your discipline?

3.2 What additional information is required to understand each data file?

What would you need to know to reproduce the results from this data?

3.3 What different versions of each data file or source will you create?

How will you differentiate between different versions, for example do you plan to use files names to denote different versions, e.g. V1, V1.1, V2 etc?

4. Looking after your data

4.1 Where will you store your data?

Laptop? USB drive? Network storage? In the cloud?

4.2 How will your data be backed up?

How many copies?

Where are they stored?

How often are copies updated?

4.3 How will you test whether you can restore from your backups?

5. Sharing your data

5.1 Who owns the data you generate?

Is it you? Your supervisor? The University? An external partner?

5.2 Who else has a right to see or use this data?

Your supervisor, collaborators, group members?

5.3 Who else should reasonably have access to this data when you share it?

Readers of your published work? The General Public?

5.4 What should/shouldn't be shared and why?

Consider any ethical, legal or commercial restrictions that may affect what you share, how you share it and who you share it with?

6. Archiving your data

6.1 What should be archived beyond the end of your project?

Everything? Just what you used for your thesis?

What might others need/want?

6.2 For how long should it be stored?

EPSRC guidelines say "10 years from the date of last access"

6.3 When will files be moved into the data archive/repository?

As you complete the analysis of each file? When you submit your thesis?

6.4 Where will the data be stored?

Disciplinary repository (e.g. crystallography databases)? X drive? Opus?

6.5 Who is responsible for moving data to the data archive and maintaining it?

You? Your supervisor? The University?

6.6 Who should have access and under what conditions?

Are there any embargoes necessary?

7. Executing your plan

7.1 Who is responsible for making sure this plan is followed?

You may wish to discuss and agree this with your supervisor

7.2 How often will this plan be reviewed and updated?

You may wish to discuss and agree this with your supervisor

7.3 What actions have you identified from the rest of this plan?

List them here with timescales

7.4 What further information do you need to carry out these actions?

Where can you find this information?

Who might you be able to ask?

Notes on completing this form

- Type as much (or as little) as you feel you need to into each box: it will expand to accommodate what you write;
- You can leave or remove the prompts in grey once you're done;
- For help with completing this DMP, please contact researchdata@brunel.ac.uk