

Data Science Workshop

Lecture 14: Reporting

Marcin Luckner, PhD
mluckner@mini.pw.edu.pl

Version 1.1
November 19, 2020

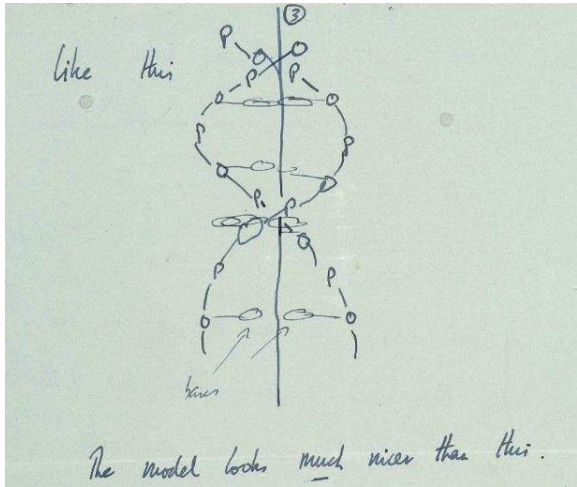


**Warsaw University
of Technology**



**MSc program in Data Science has been developed
as a part of task 10 of the project
„NERW PW. Science - Education - Development - Cooperation”
co-funded by European Union from European Social Fund.**

Scientific results reporting



Francis Crick's letter to his son.

Report preparation order

1. Prewriting (70%)
 - Collect, synthesise, and organise information
 - Prepare take-home messages
 - Work out ideas away from the computer
 - Develop an outline
2. Writing the first draft (10%)
 - Putting your facts and ideas together in organised prose.
3. Revision (20%)
 - Read your work out loud
 - Get rid of clutter
 - Get feedback from others
 - Final check

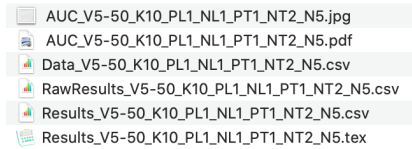
Writing in the Sciences by Kristin Sainani, <https://www.coursera.org/learn/sciwrite/>

Information collection

- Information gathering (e.g. gathering experimental results) is not a part of writing and should be done separately.
- However, it should be done in an organised way to help you with the writing process.
- At least, you should not be stopped during the writing process by searching for some data.
- Create your own organisational system.

Organisational system

- Prepare all your experiments in an automatised and reproducible way.
- Separate data generation, calculations, and reporting.
- Be ready to repeat the whole process in case of any change.



Organisational system - example

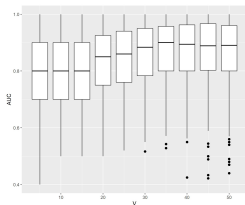


Table 1: Results

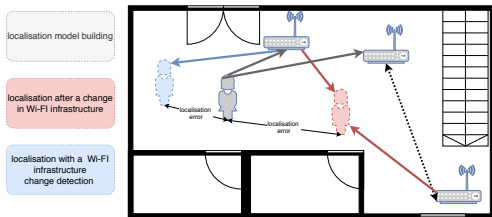
Method	System	V	AUC	ACC	TPR	TNR
RandomForest	Unknow	5.00	0.79	0.79	0.92	0.67
RandomForest	Unknow	10.00	0.79	0.79	0.96	0.62
RandomForest	Unknow	15.00	0.79	0.79	0.98	0.60
RandomForest	Unknow	20.00	0.83	0.83	0.98	0.68
RandomForest	Unknow	25.00	0.84	0.84	0.98	0.70
RandomForest	Unknow	30.00	0.85	0.85	0.98	0.72
RandomForest	Unknow	35.00	0.87	0.87	0.97	0.76
RandomForest	Unknow	40.00	0.87	0.87	0.96	0.77
RandomForest	Unknow	45.00	0.87	0.87	0.95	0.78
RandomForest	Unknow	50.00	0.87	0.87	0.94	0.80

Outline

- Arrange key facts and citations from the literature into a crude outline before writing the first draft.
 - Think in paragraphs and sections
- An issue of changing received signal strength (RSS) from an observed access point (AP).
 - The change can reduce the quality of a Wi-Fi-based Indoor Localisation System.
 - A dynamic system using an RSS estimator recognises AP with changed RSS.
 - The system rebuilds the localisation model excluding the changed AP to keep QoS.
 - The error increase:
 - with the system: 0.25 to 0.61 m
 - without the system 1.21 and 1.98 m
 - The system can be applied to detect various kinds of changes (reconfiguration, malfunction, ageing of the infrastructure).

Take-home message

- Work out take-home messages.
- The main message to remember.
- Memorable lines or graphical abstract.



In the case of an infrastructure change, the system reduces the localisation error by over one meter.

Writing the first draft

- The goal of the first draft is to write the ideas down in ordered complete sentences.
- At this moment, the text does not have to be perfect.
- The draft should be focused on logical organisation more than sentence-level details.

Revision

- Read your writing out loud.
 - The brain processes the spoken word differently than the written word.
- Cut out
 - Empty words and phrases.
 - Unnecessary jargon and acronyms.
 - Repetitive words or phrases.
 - Adverbs.

Revision - example

~~Before details of the platform and design patterns we propose in this study are discussed,~~ in this section we survey both existing design patterns and IoT platforms.

~~First of all, let us observe that~~ due to the development of a variety of IoT software and hardware solutions, a steady increase in the volume of IoT data is being observed. This makes it necessary to propose appropriate platforms and methods for IoT data processing, especially ~~ones performed in near-real time~~ ones.

Hence, the question arises of whether design patterns ~~regarding the way for~~ such platforms should be developed are ~~already~~ sufficiently mature.

Reorganisation

- In each paragraph, underline a phrase or sentence that sums up the main point.
- Move paragraphs around to improve logical flow and group similar ideas together.

Feedback

- Ask someone outside your department to read your manuscript.
- Without any technical background, they should easily grasp:
 - the main findings
 - take-home message
 - significance of your work
- Ask them to point out particularly hard-to-read sentences and paragraphs.

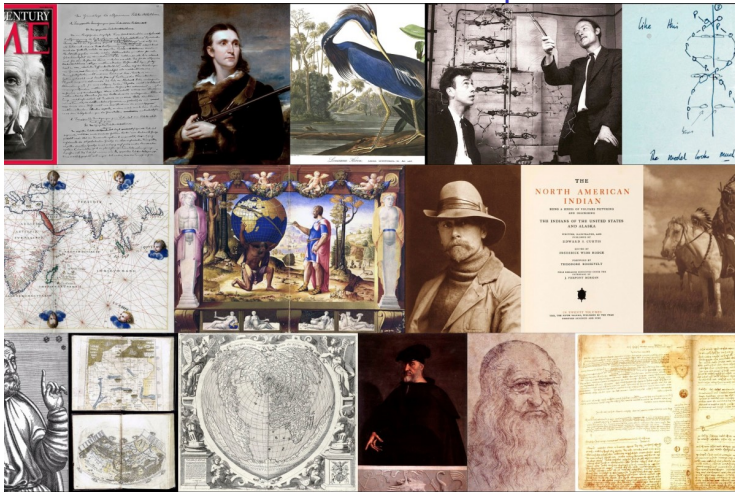
Final check

- Check a coherence of the paper.
 - Are the results and notation the same in all sections?
- Check numbers.
 - Are the numbers the same in figures, tables, and the main text?
- Check references.
 - Are the references correct technically and substantially?

Citations

- Always cite the origin source.
- Assume that other authors make mistakes in the references.
- Richard Dawkins wrongly cited one work in his book. An observation of how the following authors reproduce the mistake brought him to mind an information evolution theory and he created a concept of meme.

Scientific Manuscript



<https://newatlas.com/6-most-valuable-scientific-books-and-manuscripts/45785/>

Scientific Manuscript Structure

1. Abstract
2. Introduction
3. Methods
4. Results
5. Discussion

Introduction

- Structure of the introduction
 1. Background, know information
 2. Knowledge gap, unknown information
 3. Hypothesis, questions, and purpose statement
 4. Approach, plan of attack, and the proposed solution

Methods

- Give a clear overview of what was done.
- Give enough information to replicate the study.
- Give complete information but respect your reader time.
 - Break into smaller sections with subheadings
 - Cite references for commonly used methods
 - Display in a flow diagram or table where possible

Results

- The results section summarises what the data shows
 - Point out simple relationships
 - Describe global trends
 - Cite figures or tables that present supporting data
- Avoid repeating the numbers that are already available in the tables and figures.

Discussion

- Structure of the discussion
 1. Answer the question asked.
 2. Support your conclusion
 - With your data or with other data.
 3. Defend your conclusion
 - Anticipate criticisms
 4. Give the global view take-home message

Abstract

- Overview of the main story.
- Gives highlights from each section of the paper.
- Limited length (100-300 words, typically).
- Stands on its own.
- Used, with title, for electronic search engines.
- Most often, the only part people read.

Manuscript preparation order

1. Tables and Figures
2. Results
3. Methods
4. Introduction
5. Discussion
6. Abstract

References



J. W. Creswell and J. D. Creswell.

Research Design: Qualitative, Quantitative and Mixed Methods.

SAGE, 2014.



B. Gastel and R. A. Day.

How to Write and Publish a Scientific Paper.

Greenwood, 2011.



K. Sainani.

Writing in the sciences.



**Warsaw University
of Technology**



**MSc program in Data Science has been developed
as a part of task 10 of the project
„NERW PW. Science - Education - Development - Cooperation”
co-funded by European Union from European Social Fund.**