CAS Applied Data Science, University of Bern, S. Haug

**Assignment Work Instructions for Module 1**

The assignment work for Module 1 is a Conceptual Design Report (CDR) for a Data Science project you would like to perform. Maybe you already know what will be your final CAS project, however, that doesn’t have to be? Please use or orient yourself according to the template below. Remove this page for the submission version. Key information for the report is the following.

**Language:** English (exceptionally German)

**Deadline:** To be defined in class

**Expected effort and length:** About 30 hours, minimum 5 pages. You may work in teams (max 3).

**Further formal quality requirements:**

* All references must be listed in corresponding Reference section at the end of the CDR and cited with numbers in text
* All tables and figures should have numbered legends with short explanations (tables above, figures below) and be referenced in text (Figure 1: blabla, Table 1: blablabla).
* Figures to be as self explanatory as possible, e.g. plots with at least axis labeling including units.

**Patricio Becerra**

Grüneckweg 10B, CH-3006, Bern

patriciobecerrav@gmail.com

**Data Science Project**

**Image-based Classification of Sediment Types in Drill Cores**

**Conceptual Design Report**

**10 December 2023**

# Abstract

10-20 lines

Lorem ipsum dolor sit amet, consectetuer adipiscing elit, sed diam nonummy nibh euismod tincidunt ut laoreet dolore magna aliquam erat volutpat. Ut wisi enim ad minim veniam, quis nostrud exerci tation ullamcorper.

# 

# 

# Table of Contents

[**Abstract 1**](#_buc6q0k08dmn)

[**Table of Contents 2**](#_z0ssh6k3nrnr)

[**1 Project Objectives 2**](#_5bx4zkckkkd)

[**2 Methods 2**](#_q672ig652t9r)

[**3 Data 3**](#_wauol8kayjkz)

[**4 Metadata 3**](#_h4s0m7kg1q1p)

[**5 Data Quality 3**](#_vsvk692tupyf)

[**6 Data Flow 3**](#_x5u0l8hx0kbh)

[**7 Data Model 3**](#_mtk27zhjxb66)

[**8 Documentation 4**](#_xyovz8obzdnp)

[**9 Risks 4**](#_akvz1edyy9sm)

[**10 Preliminary Studies 4**](#_r9xh3qn8v2wb)

[**11 Conclusions 4**](#_4hg0yma181gd)

[**Acknowledgements 4**](#_7gj90am4irap)

[**Statement 4**](#_rw3i0hxq4dsj)

[**Appendix X 5**](#_7djimjyxr32l)

[**References and Bibliography 5**](#_g859w1rzj1iv)

# 

# 1 Project Objectives

0.5-1.0 page

The drilling and extraction of subsurface cores is an important process in the geological sciences. In a general sense, core-drilling entails the extraction of an intact (as much as possible) cylindrical core from a specific area of geologic interest. In this core, past surface layers are preserved, from which the geologic, and often, climatologic history of a region, continent and even the whole planet can be

Formulate the goal and purpose of your project here. It is very important to be very specific on your goals. What do you want to find out or predict? What are the numbers and plots you need to do that?

# 2 Methods

0.5-1.0 page

Which infrastructure, tools, software libraries, statistical methods etc do you intend to use. It is clear that you may not know all this at this stage, but try to make yourself some thoughts, even if it is going to change during the CAS.

1. OS (OperationSystem module) [5] - Navigating to file locations and extracting file names

2. Pandas[6], [7]- Importing and exporting csv files, data manipulation.

3. Numpy[8] - Importing and exporting csv files, data manipulation

4. Matplotlib.pyplot[9] - Plotting figures.

5. Scipy[10] - Statistic analyses

6. Statsmodels[11] - Statistic analyses

7. Scikit-learn[12] - Applying machine learning methods

8. TensorFlow[13], [14] – ML, DL and CNN functionalities

9. Mpltern[15] - Plotting ternary plots (if necessary)

10. OpenCV[16] – Image manipulation, processing and analysis

# 3 Data

Which data will be used (with references) 0.5-1.0 page

A couple of plots, maybe some histograms of the columns

A couple table row as an example maybe

Security issues etc (see data management plan, you may attach a SNSF data management plan for your data)

# 4 Metadata

What metadata is required for reproducing your analysis?

Where do you store the metadata, how can people access it?

# 5 Data Quality

What are the quality requirements you need in order to meet your project objectives (data size, data precision missing values, ….)?

Are they met? If not, do you expect a significant impact on your result?

Any measures to improve the data quality?

# 6 Data Flow

Explain with a figure and words how the data flow of your project will be, from the data source to the final plots and numbers.

# 7 Data Model

Draw and explain your data model at the conceptual level, the logical level and the physical level. .

Conceptual

Logical (what columns/features will you use/need)

Physical (infrastructure needs)

# 8 Documentation

How will the project be documented?

# 9 Risks

What can go wrong?

When this and that goes wrong, what counter measures do you have?

What will be the impact on the quality of the aimed output, project time schedule, project cost ?

# 10 Preliminary Studies

Plots and numbers (from Module 2).

# 11 Conclusions

….

# Acknowledgements

Acknowledge persons or institutions that helped you with the CDR here.

# Statement

The following part is mandatory and must be signed by the author or authors.

„Ich erkläre hiermit, dass ich diese Arbeit selbstständig verfasst und keine anderen als die angegebenen Quellen benutzt habe. Alle Stellen, die wörtlich oder sinngemäss aus Quellen entnommen wurden, habe ich als solche gekennzeichnet. Mir ist bekannt, dass andernfalls die Arbeit als nicht erfüllt bewertet wird und dass die Universitätsleitung bzw. der Senat zum Entzug des aufgrund dieser Arbeit verliehenen Abschlusses bzw. Titels berechtigt ist. Für die Zwecke der Begutachtung und der Überprüfung der Einhaltung der Selbstständigkeitserklärung bzw. der Reglemente betreffend Plagiate erteile ich der Universität Bern das Recht, die dazu erforderlichen Personendaten zu bearbeiten und Nutzungshandlungen vorzunehmen, insbesondere die schriftliche Arbeit zu vervielfältigen und dauerhaft in einer Datenbank zu speichern sowie diese zur Überprüfung von Arbeiten Dritter zu verwenden oder hierzu zur Verfügung zu stellen.“

Date: Signature(s):

# Appendix X

If you have something to attach to your report, do it here.

# References and Bibliography

Please number any information source you used in the report with corresponding links here [1]:

[1] S. Haug et al., How to make a CDR, own brain, 2020 (put a weblink or DOI here)

[2]