Thesis Proposal Data Science & Society

1. General

The project proposal consists of 4 well-written pages (1.5 line spacing, font size 11, excluding the title page, references, tables, and figures). The proposals will be evaluated on clarity (is it clear what will be done in this project?) and completeness (Is the proposal well-thought out? Does it address all of the important aspects of the proposed project?). The proposals should be written in correct Academic English and adhere to the APA 6 Style. Proposals with spelling, grammar and style mistakes will not be evaluated; instead, you will be asked to resubmit a corrected version.

Submit your proposal as a pdf file in the Canvas Assignment. Your proposal will be reviewed by the intended supervisor of the project and, if necessary, a second reader. You will receive feedback and it is expected that you make use of the feedback for the content of your Presentation. Note that your project plan may need to be adapted as you learn more about the data. This is fine as long as your overall goal (the task you are addressing, the data set you are using) remains generally the same. Should your project change in a major way from what you proposed, you need to get an approval from your supervisor.

Important: On the title page, provide your full name, email address, the name of your internal supervisor and the contact information of your external supervisor, if applicable.

2. Avoiding Plagiarism

As with all assignments, you have to make sure that you do not commit plagiarism. Plagiarism is considered a serious case of fraud that, when suspected, will be reported to the Examination Board. Committing fraud can have serious consequences. At the minimum, when fraud is established by the Examination Board, the assignment is declared invalid and, in the case of a thesis, a new thesis will have to be written. Please see Article 16 of the Rules and Guidelines for TSHD (see below) for the procedure and sections in case of fraud. Note that TiU defines plagiarism as: "Using parts of a text written by someone else, or the reasoning or ideas of others, for a thesis or other assignment, without due acknowledgement." (Source: https://www.tilburguniversity.edu/students/studying/regulations/fraud/whatisplagiarism – this text contains a more elaborate explanation of what is plagiarism.)

2.1. Overlap Detection - Feedback

To prevent accidental plagiarism, we want to ensure that students can gain experience with the overlap detection mechanism implemented by TurnItIn. TurnItIn is a tool that supervisors use to assist them in detecting potential cases of academic fraud. Please note that establishing fraud is

a decision that is always made by the Board of Examiners of the School, not just by the TurnItIn algorithm. To provide you with this experience, the Canvas assignment for the Draft Thesis Proposals will give complete TurnItIn feedback to students as well as supervisors, so that both can learn from the feedback provided by this system. In addition, this Canvas assignment is open for multiple resubmissions, so that students get the opportunity to repair a draft proposal with regard to overlap before submitting their full proposal.

Useful Resources:
- TSHD Education and Examination Regulations (EER), including the Rules and Regulations: https://www.tilburguniversity.edu/students/studying/regulations/eer/humanities
- What is plagiarism?
https://www.tilburguniversity.edu/students/studying/regulations/fraud/whatisplagiarism

3. Outline and Contents

3.1. Project Definition

Provide a clear description of the problem you plan to address.

3.2. Motivation

Briefly explain why this problem is worth addressing, both from a practical/societal and scientific point of view. Make sure that the problem you address has not been solved already.

3.3. Background

Provide a summary of what is known in the scientific literature about this problem. This should be based on at least 5 (preferably, more) relevant sources. These sources need to satisfy the following requirements: (1) recency (published in the last 5 years), (2) quality (published in scientific peer-reviewed journals or conference proceedings), and (3) usefulness (they should help you frame the theoretical background of your project). At this stage of your project, a full literature review is not expected but it will be expected by the time of the final report.

3.4. Dataset Description

Describe the dataset(s) that you will use in your project (size, format, accessibility). Provide rationale as to why you are choosing these data.

3.5. Algorithms and Software

Describe what algorithms and software you plan to use in your project. For data mining/machine learning algorithms, define how you will evaluate your results. For prediction problems

(classification or regression), you will likely make use of standard techniques: Against what baseline methods will you compare your algorithm(s)? How do you plan to obtain ground-truth labeled data so that you can measure accuracy, precision, recall or some other metric? If you are planning to use unsupervised techniques, provide information of how the clustering algorithm will be tested.

3.6. Evaluation Method

Define how you will evaluate your results. For prediction problems (classification or regression), you will likely make use of standard techniques: Against what baseline methods will you compare your algorithm(s)? How do you plan to obtain ground-truth labeled data so that you can measure accuracy, precision, recall or some other metric? If you are planning to use unsupervised techniques, provide information of how the clustering algorithm will be tested.

4. Milestones and Plan

Sketch out what you think will be the major intermediate milestones that you will need to achieve. Give a general idea of your planning.

Version 1.2 - Sept. 30, 2020, M. van Wingerden
Version 1.1 - August 28, 2020. A Hendrickson
Version 1.0 - August 20, 2019 (directly from the 2017-2018 document)