JADS: *The Jheronimus Academy of Data Science*[1]
Master program "Data Science and Entrepreneurship"
Master thesis (Graduation project) information for external companies and organizations
Dr. Geert Monsieur, MSc Thesis Coordinator @ JADS
Version 02.12.2021

Planning

The thesis project is 24 ECTS (672 hours) spread over 2 semesters (September – June, or February-January). Students usually spend 2 days per week working on the thesis in the first semester, and 3 days per week during the second semester. Students will receive the list of all topics suggested by our partners in May and November, and start working on their thesis project in September and February accordingly.

Students have to finish their thesis within 2 semesters, otherwise they (1) get an extension of 4 weeks and a grade of not higher than 6; or (2) if they do not meet this 2nd deadline, they have to start from scratch.

Minimal requirements

The master thesis is always related to data, and has business and/or societal impact.

The targeted impact can be differentiated using the following maturity scale:

1. incremental model improvement (e.g., improve an existing customer retention model)
2. cost decrease (new solution for an existing business)
3. revenue increase (new solution for an existing business)
4. new revenue generation/new business development/new product or service.

Maturity level 1 is assessed as insufficient in the entrepreneurship rubric in the thesis assessment form. The minimum maturity level is 2, with a preference for maturity level 4 (new revenue generation/new business development/new product or service).

Both business and societal impact have to be supported by data and experiments/tests, not only assumed, and further delineated in either a separate reflection document as part of the thesis portfolio or managerial implications section of the thesis report.

Each master thesis should aim for a contribution to a specific scientific discipline. Therefore, scientific contribution requirements for the graduation project differs. The research question(s) must be aligned with the main discipline of the master thesis, steering the particular (aimed) contribution of the master thesis. Below is a list of all the disciplines represented at JADS:

Data Science:

*   Data modeling and analysis methods
*   Deep learning
*   Data engineering and governance
*   HTI and recommender systems
*   Visualization
*   Design/Technology

Entrepreneurship/social sciences:

*   Strategy

---

[1] JADS is a Joint Graduate School of Tilburg University and Eindhoven University of Technology.

- Legal/ethics
- Social networks
- Regulations and institutions
- Cognition and psychology

Datasets and external companies

Since every master thesis project has to feature _both data science and entrepreneurship_ aspects, having at least one dataset as the basis of the master thesis is pretty much unavoidable. The dataset could come from the following sources or a combination thereof:

- External company or organization
- JADS-affiliated research project (i.e. from a researcher)
- Student's own company
- Data scraped or otherwise collected by student

If the external company/organization is not able to deliver the dataset(s) it committed within 2 months after the official master thesis project start, the student has to switch to another dataset/project (could be usually provided by the main supervisor) and redesign the graduation project accordingly.

So the deadline for companies to provide the data they committed for the master thesis project is:

- For students starting in September: **November 1**,
- For students starting in February: **April 1**.

We have to enforce this deadline to make sure that the students get a fair chance to graduate on time and get a good grade (see the Planning section above). Moreover, the more time the students have to work on the data, the better the results will be for your own organization.

Supervision

Each student is assigned to the main supervisor from JADS based on his/her topics of interest that are matched to the expertise of the supervisors. Each student will be also assigned to the 2nd assessor from JADS.

In case the student does the project at an external company or organization, we expect that (s)he will also get a company supervisor who will be responsible for providing input from this company or organization, helping establishing relevant contacts, organizing access to dataset(s) and further facilitating the students' project. While the main burden of the student supervision is with JADS, having a dedicated company supervisor will help your organization make the best out of this master thesis project.

Please let us know if you have any further questions, and looking forward to a productive collaboration!

Geert Monsieur
Assistant Professor
Jheronimus Academy of Data Science
Sint Janssingel 92
5211 DA 's-Hertogenbosch – The Netherlands
g.monsieur@jads.nl

A. Appendix

This section contains suggestions to help students and supervisors to smooth the process of the master thesis trajectory.

A.1. Research directions

Below potential research directions per mentioned research direction within the master are discussed.

Data Science:

- **Data modeling and analysis methods**. Develop new or substantially improved methods to analyze and collect data. Another option is to tackle an assumption of a method that is often violated in practice, further develop the method, show its robustness, or thoroughly examine its behavior using analytical or empirical (simulation) tools. The work should contribute to research in data mining or statistical methodology. Its output should be aligned with the state-of-the-art in this field (e.g., a conference contribution). Various methods are covered here including methods to predict, classify, decrease dimensionality, analyze structure and model/simulate (e.g., Bayesian and frequentist approaches to prediction, time series analysis, random forest classifiers, process mining, topic modeling, clustering, network analysis, agent-based modeling, system dynamics).
- **Deep learning**. Build neural network-based models trying to improve the previous models in terms of prediction, contribute to the understanding of the success of deep learning, create novel deep learning models to solve new tasks or to deal with new task requirements.
- **Data engineering and governance**. Design novel type of data infrastructure that can handle heterogeneous types of data in a novel way with research questions relating to all or parts of the following: (1) Build and maintain a data pipeline to address a specific data-mining problem; (2) Clean and wrangle data into a usable state; (3) machine-learning solutions engineering and deployment from 1 and 2 from above.
- **HTI and recommender systems.** New or improved recommendation algorithms, new implications of human processing of data science models for the design and potentially visualization of such models.
- **Visualization**. A new visualization method (extreme case); experiments to evaluate an existing visualization method(s), ideally with entrepreneurship-oriented DVs (e.g., purchase intent, well-being, engagement, trust, performance).
- **Design/Technology**. Given a non-trivial real-world case: analyze it; derive requirements; find the solution(s); implement these; evaluate the results; draw conclusions.

Entrepreneurship/social sciences:

- **Strategy**. What are the optimal revenue models for data-driven businesses? How should new platforms be optimally designed and scaled up? What are the core failure attributions of bankrupted ventures?
- **Legal/ethics**. How ethical are the permissions requested by apps? Which influence do they have on app demand and performance? What kind of legal implications should these findings have?
- **Social networks**. To what extent do different entrepreneurial and financial network characteristics predict different performance outcomes at the team and company level?
- **Regulations and institutions**. What are the effects of different types of regulations and institutional norms on companies' strategies and performance?
- **Cognition and psychology.** To what extent do different emotions and body language of pitching entrepreneurs influence different kinds of evaluators such as VCs, angel investors, incubator managers, and (potential) clients?

There are different ways such confidentiality/sensitivity can be managed, and often the supervisor can help the student with further suggestions.

The list is indicative of possible contributions, which can be further delineated depending on the discipline. For example, the visualization discipline suggests the following types of papers (with respective contributions) as the accepted norm:

http://ieeevis.org/year/2018/info/call-participation/infovis-paper-types

A.2. NDA

Below, some suggestions to avoid problems with NDA's are indicated. Individual master thesis supervisors or JADS do generally NOT sign any non-disclosure agreements (NDA's) for master thesis projects. We advise students to be extremely careful with signing NDA agreements individually with external companies/organizations and making sure that they can still graduate based on such an NDA and fulfill the aforementioned master thesis requirements. Typically, if the NDA only regards the data used by the student, this should not create any problems for graduation. Please discuss such possible NDA **before signing it** with the main supervisor.

In case the company/organization does not have an own contract/NDA, please check the following TUe page for examples of contracts and NDA agreements:

https://studiegids.tue.nl/organisatie/studentenadministratie/formulieren/#top

Below some other options to avoid problems with NDA's are stated:

- One simple remedy is to use anonymization, making sure that the company name is not listed or recognizable from the thesis.
- If this is not enough or as an alternative solution, the student can agree on a scientific contribution that does not compromise the data and/or code that is sensitive to the company. For example, scientifically, the algorithm/mechanism to identify and segment the prospective customers could be the scientific contribution. At the same time, the company would be interested in the exact prospective names to pursue these leads. Only the prospective names are sensitive, are part of the data that are being generated, and do not necessarily have to be publicly disclosed as part of the thesis.
- The student can also investigate a specific strategic topic relevant to the external company, yet only use publicly available data sources or data that the student collects. For example, the company is interested in the application of distributed ledger for truck platooning: the student researches this topic by collecting and analyzing publicly available secondary data and collecting primary data.