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. Dario di Nucci, MSc Thesis Coordinator @ JADS
May, 2020

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 @ JADS is always related to data, involves development of a novel technical solution **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 the existing customer retention model)
2. cost decrease (new solution for existing business)
3. revenue increase (new solution for existing business)
4. new revenue generation/new business development/new product or service.

Low maturity (1) is not enough for a master thesis. You can consider engaging with MKB DataLab @ JADS (https://jadsmkbdatalab.nl) for such projects. All the thesis projects have to be score higher in terms of maturity (level 2, 3 or 4), 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 numbers and experiments/tests, not only assumed, and further delineated in a separate reflection document.

Each master thesis should aim for a scientific contribution in one certain scientific discipline such as human-technology interaction (HTI) and recommender systems, data mining techniques, deep learning, data engineering, entrepreneurship or management.
The research question(s) have to be aligned with the main discipline of the master thesis, steering the particular (aimed) contribution of the master thesis. From the company perspective, the scientific discipline represents the area in which the company would like to get most answers on.
In terms of the contribution, the following can be used as a guideline for each discipline represented at JADS:
1. **Data Science**:
   - **Data modeling and analysis methods**: e.g. new (extreme case) or substantially improved method to analyze data or setup data collection; e.g. tackle an important assumption of a method, which is being often violated in practice, develop the method further and show robustness, or thoroughly exam the behavior of new methods using analytical or empirical (simulation) tools. This work should contribute to research in data mining or statistical methodology and its output should be aligned with standard publications in this field (e.g., a conference contribution). Various methods are covered here including methods to predict, classify, decrease dimensionality, analyze structure and

---

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

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 modelling, system dynamics).

- **Deep learning**: build neural network-based models trying to beat the previous best model in terms of prediction, contributing to the understanding of the success of deep learning, create novel deep learning models to solve new tasks or to deal with novel task requirements.
- **Data engineering**: 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 solution(s); Implement these; Evaluate the results; Draw conclusions

2. **Entrepreneurship/social sciences**: answering novel questions for a certain literature stream by applying neural networks, text mining, etc. as methods, not necessarily extending these methods:
   - **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 and 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 characteristics of the entrepreneurial and financial networks predict different performance outcomes on 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?

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 <u>within 2 months</u> 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!

Dario Di Nucci
Assistant Professor
Jheronimus Academy of Data Science
Sint Janssingel 92
5211 DA 's-Hertogenbosch – The Netherlands
D.DiNucci@tilburguniversity.edu
http://dardin88.github.io