

# Master's thesis Seminar ISM-E5001

Master's program in  
Business Analytics

Liesiö, Vilkkumaa, Malo, Leppänen

23.1.2024



Aalto University  
School of Business



# Introduction

- This presentation includes some additional information to help you in writing a Master's thesis with a Business Analytics (BA) topic
- You should familiarize yourself also with the other three ISM seminar lectures (introduction, quantitative methods, qualitative methods) and Pekka Malo's video
- There will be a separate group for student in the BA master's program if there are enough students



# Examples of BA thesis topics

## **Portfolio Approach to Counter-Party Risk Modelling**

- Applying a simulation model from the literature at a company
- Software: Excel
- Related courses: Business Analytics 2, Simulation

## **Price optimization for consumer loans — Case Finnish fintech company**

- Applying a non-linear optimization model from the literature at a company
- Software: Excel + Solver
- Related courses: Business Analytics 1 & 2

## **Decision support model for cloud adoption strategy - a case study of a Finnish telecom company**

- Developing a multiattribute value model for a company
- Software: Excel
- Related courses: Business Analytics 2, Decision making and choice behavior

## **Solar potential in Helsinki – optimizing the allocation of solar panels on the roofs of Helsinki**

- Developing ILP model, and applying it to open data
- Software: Python + gurobi
- Related courses: Business Analytics 1

## **Bootstrap resampling in second order stochastic dominance based portfolio selection**

- Implementing LP models from existing literature and applying them to open data
- Software: Python + gurobi
- Related courses: Business Analytics 1 & 2



# Examples of BA thesis topics

## **Determinants of central government employee turnover in Finland: A competing risk analysis**

- Applying hazard models to identify determinants the employment duration
- Software: R
- Related courses: Data Science for Business I

## **Comparing external data monetization of two analytics products**

- Semi-structured interviews for identifying challenges and enablers for analytics productization
- Software: -
- Related courses: Business Intelligence

## **Facilitating public procurement processes with multi-criteria decision analysis: Evidence from the HX Fighter Program**

- Qualitative methods for developing a framework to integrated MCDA with public procurement and identifying benefits from and challenges with such integration
- Software: -
- Related courses: Business Analytics 2, Decision making and choice behavior

## **Digital reputation management through a third-party review site**

- Large dataset collection using automated scripts, data analysis including semantic analysis of customer reviews
- Software: Python
- Related courses: Data Science for Business I

## **An Analysis of Attributes Affecting the Price of Plant-Based Meat Alternatives in Finnish Food Markets**

- Empirical data collection and an estimation of a pricing model using regression
- Software: R, Excel
- Related courses: Data Science for Business I



# Examples of BA thesis topics

## **Improving sales support using semantic similarity search**

- Applies neural networks to improve efficiency of sales support process
- Software: Python
- Related courses: Data Science for Business I & II

## **Domain-specific Knowledge Graph Embeddings**

- Development of a new knowledge graph model for the domain of supply chains and packages
- Software: Python
- Related courses: Data Science for Business I & II

## **Investigating the applicability of spatiotemporal transformer networks for air quality forecasting**

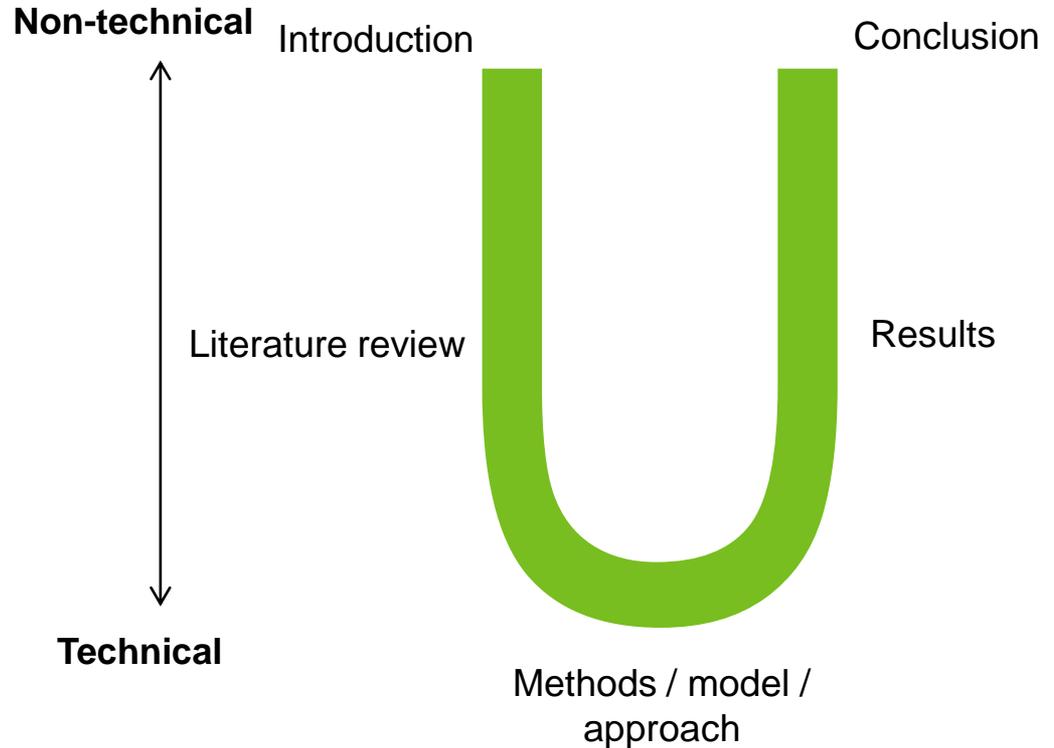
- Application of transformer-based neural networks for spatiotemporal forecasting
- Software: Python
- Related courses: Data Science for Business I & II

## **Empirical asset pricing via machine learning in the context of Nasdaq Helsinki**

- Studied the predictive power of different ML models for portfolio management purposes
- Software: Python/R + Excel
- Related courses: Data Science for Business I & II

# U-model for a BA thesis

- Title page
- **Abstract**
- Table of contents
- **Introduction**
- **Literature review**
- **Methods / model / approach**
- **Results**
- **Conclusions**
- References
- Appendices



# Structure of thesis: Abstract

- A concise description of the objectives, research methods, and key results of the thesis
- Provides a clear, general understanding of the thesis topic, which can also be understood by a non-specialist reader
- Is an independent piece of text with no references
- The abstract should not contain anything which is not presented in the actual thesis

# Structure of thesis: Introduction

1. Motivation and background. Which phenomena or application context is the topic connected to? Why is topic relevant?
  2. What kinds of challenges are related to the topic? How has earlier literature dealt with such challenges? What kinds of gaps there may still be?
  3. In what respects does this work fill the gaps in the literature? What is the contribution of this thesis? What are the research objectives or questions?
  4. How does the thesis address the questions/objectives (e.g. through developing a model/method/approach/framework and applying it to a case/dataset)?
  5. Structure of the thesis, i.e., a brief description of the contents of each section after the introduction.
- The introduction should be easy to read – keep the technical vocabulary at a minimum

# Structure of thesis: Literature review

- Often a descriptive title (e.g., “Earlier approaches to forecasting demand of slow-moving products”) can be better than the generic “Literature review”
- Presents earlier work on the topic
- Format: “Persons X studied subject Y and came to the conclusion that Z”
- You can reflect your thesis contribution to the literature, but keep it in the background
- You may use technical vocabulary here, but introduce formulas or mathematical notation only if they will be referred to later
- Some pieces of the previous literature are introduced already in the introduction to explain “the gap”: Maybe review these in more depth here

# Structure of thesis: Methods / models / approach / framework

- Use a descriptive title (e.g., “Optimization model for sustainable investment portfolio selection”) rather than “The model”
- Methods / models are presented exhaustively and as clearly as possible so that others can reproduce them
- Help the reader by
  - Proceeding in a logical order
  - Defining all symbols immediately when used for the first time
  - Giving examples
- Based on the phenomenon you are studying, justify why you chose to use this model / method / approach / problem formulation / distribution / parameters / etc.
- This is the technical part of your thesis

# Structure of thesis: Results

- A descriptive title (e.g. “Application to historical return data”) often better than the generic title “The results”
- Present all data that you use to obtain your results
  - Simulation: What kinds of parameters (or other assumptions) did you use and why?
  - Empirical data: How did you gather your data?
- You may use tables and figures to illustrate your results
- Reflect on your results: What do the results mean from the point of view of the phenomenon under study? Are the results sensible? Are they statistically significant?
- Sensitivity analyses – how do the model assumptions / choices of parameters affect the results?
- This section should be understood without having thoroughly read the methods section – try not to be overly technical. Yet, it usually makes sense to use math symbols from the method-section to make it clear which data links to which model parameters.



# Structure of thesis: Discussion and conclusions

- Can be two separate sections and some theses benefit from having a separate section “Managerial implications” or “Policy implications”
- What was done in the thesis?
- What are the key results?
- Based on these results, what conclusions can you draw about the phenomenon you are studying?
- How do the assumptions and/or data limit the generality of your conclusions?
- What kinds of questions remain unanswered by your thesis – what should be studied next?
- Keep the conclusions non-technical and easy to read

# Good scientific writing

- Provides evidence, not experience
  - Do not draw unjustified conclusions

Based on a large international data set, Flyvbjerg et al. (2002) show that the realized costs of public infrastructure projects are significantly higher than estimated. 

Based on my experience, a project always lasts ca. 50% longer and costs ca. 33% more than what was initially estimated. 

- Maintain a professional style

The large number of failure combinations presented a computational challenge. To speed up computation, only those combinations whose occurrence probability was at least 0.1% were considered. 

The computation took ages, and boy did that make me go ballistic. 

Does not, was not, has not, is not, large, high 

Doesn't, wasn't, hasn't, isn't, big, big 

# Good scientific writing

- Text is unambiguous
  - Try to minimize the chance of misunderstanding

The optimization model developed in the thesis can be used to determine the production plan with minimal total duration, such that constraints related to schedule and resources are taken into account. 

This thesis developed an optimization model, with which one can form a total duration minimizing production plan, that takes into account the schedule and resource constraints related to it. 

- Use concepts/terms consistently; but avoid unnecessary repetition

The alternatives were evaluated on the aforementioned attributes such that the attribute-specific values had been scaled between  $[0,1]$ . The attribute weights were determined by using the SMARTS-method (Edwards & Barron, 1994). In the SMARTS-method... 

The alternatives were evaluated on the aforementioned attributes such that values on each subobjective had been scaled between  $[0,1]$ . The criterion weights were determined using the SMARTS-method (Edwards & Barron, 1994). 

- Define concepts/terms/variables/symbols immediately, when you use the for the first time

# Good scientific writing

- The author remains in the background
  - Be cautious with "I"

The optimization model developed in this thesis can be used to determine the production plan with minimal total duration, taking into account possible constraints related to schedule and resources. Based on simulation tests, the computation time remains reasonable when...



In this thesis I developed an optimization model that I can use to determine the production plan with minimal total duration, taking into account possible constraints related to schedule and resources. I think that this model works well, and developing it was a very educational experience for me.



# Good scientific writing

- The first sentence of each paragraph is the most important sentence
  - Serves as the “title” of that paragraph
  - Ideally, reading the only the first sentence of each paragraph would make it possible to follow the overall narrative and argumentation of the thesis

...A third complicating factor is that strategy work often requires a balance between preparing for alternative scenarios for the future (i.e., developing a robust strategy), and trying to steer the course of change towards a desirable direction (i.e., developing a market-shaping strategy).

This thesis develops a model-based framework for supporting the development of a robust, market-shaping strategy. This framework combines elements of scenario planning and robust portfolio decision analysis. Scenario planning is utilized to ...

The thesis reports experiences from applying the developed framework to support strategy formation at Nordea Bank Abp, the largest retail bank in the Nordic countries. We also summarize findings...

...A third complicating factor is that strategy work often requires a balance between preparing for alternative scenarios for the future (i.e., developing a robust strategy), and trying to steer the course of change towards a desirable direction (i.e., developing a market-shaping strategy).

Nordea Bank Abp is the largest retail bank in the Nordic countries. The payments unit at Nordea had previous experience in using scenarios in their strategy development processes. This thesis develops a model-based framework for supporting the development of a robust, market-shaping strategy, and applies it to support Nordea in...

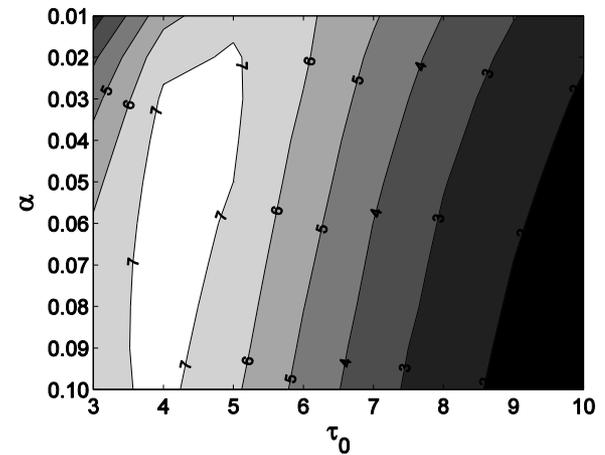
# Figures and tables

- All figures and tables have numbers and a caption
  - Figure captions below the figure
  - Table captions above the table
- All figures and tables are referenced in the text

Figure 1 shows the contours of the option value as a function of the standard deviation  $\tau_o$  and . The projects' utilities and costs are presented in Table 2.



The figure above shows the option value...The projects are in the following table:



**Figure 1.** Option value as a function of standard deviation  $\tau_o$  and threshold value  $\alpha$ .

**Table 1.** Utilities and costs of projects A-G. The selected projects are colored green.

Project ( $j$ )	Utility ( $u_j$ )	Cost ( $c_j$ )
A	9	6
B	7.5	2
C	8.5	5
D	9	2
E	10	6
F	10	3
G	8.5	2

# Mathematical notation

- All notation is introduced when first used
- Use the same notation throughout the thesis
  - The same notation/symbol cannot have multiple meanings
- Use Word equation editor for all mathematical symbols
- Use the “Normal text” style for all text within the equation environment
- Equations are part of the text, i.e., they continue the sentence preceding them
  - If an equation ends a sentence, end it with a full stop
  - An equation might end with a comma, if the sentence continues after the equation
- Number equations if you need to refer to them later in the text
- Sentences should not start with a mathematical symbol or an equation

Let  $x_i^j$  denote the performance of alternative  $j$  with regard to attribute  $i$ . ✓

The value of alternative  $x^j$  is given by

$$V(x^j) = \sum_{i=1}^n w_i v_i(x_i^j), \quad (1)$$

where  $w_i$  is the weight of attribute  $i$  and  $v_i(\cdot)$  is the attribute-specific value function for attribute  $i$ . By utilizing Equation (1) we can formulate... ✓

The value of alternative  $x^j$  is given by

$$V(x^j) = \sum_{i=1}^n w_i v_i(x_i^j) \text{ for all } j \in J. \quad \checkmark$$

The value of alternative  $x^j$  is given by

$$V(x^j) = \sum_{i=1}^n w_i v_i(x_i^j) \text{ for all } j \in J. \quad \times$$

The value of the alternatives is calculated as follows:

$$V(x_j) = \sum w_i \cdot v_i(x_{ji})$$

$W_i$  = weight of attribute  $i$  ✓

# Citations

- Use name & year to cite other work
- If there are more than two authors, use the name of the first author followed by “et al.”
- All items on the list of references must be cited
- All citations must be found in the list of references
- All references must have the same format

This Bayesian variant of the EM algorithm is discussed in more detail by Schafer (1997).



In particular, many empirical studies suggest that the realized values of selected alternatives tend to be systematically lower than estimated, causing the decision-maker to experience post-decision disappointment (Bell, 1985; Harrison & March, 1984; Smith & Winkler, 2006).



Flyvbjerg, B., M. Skamris Holm., S. Buhl. 2002. Underestimating costs in public works projects – error or lie? *Journal of the American Planning Association* **68** 279-295

Smith, J. and Winkler, R. (2006) The Optimizer’s Curse: Skepticism and Postdecision Surprise in Decision Analysis, *Management Science*, Vol. 52:3, pp. 311-322.



# Building a list of references

Google Scholar is a useful tool for searching literature and for getting the references in a consistent format

The screenshot shows a Google Scholar search for "Use of business analytics among the old people". The search results list several articles. A "Cite" dialog box is open over the first article, "A model of health: Using business analytics to identify older Canadian adults with heart disease" by T Ainge (2018). The dialog box displays citation information in MLA, APA, Chicago, Harvard, and Vancouver styles. A green dashed box highlights the "Cite" button in the article's action menu, with a green arrow pointing to the "Cite" button in the dialog box. The URL <https://scholar-google-com.libproxy.aalto.fi/> is highlighted in a green box at the bottom left.

Google Scholar

Use of business analytics among the old people

Articles About 512 000 results (0,38 sec)

Any time

Since 2024

Since 2023

Since 2020

Custom range...

Sort by relevance

Sort by date

Any type

Review articles

include patents

include citations

<https://scholar-google-com.libproxy.aalto.fi/>

Aalto University School of Business

A model of health: Using business analytics to identify older Canadian adults with heart disease [PDF]

T Ainge - 2018 - summit.sfu.ca

... the landscape of chronic disease among older adults in Canada is that by the time people reach old age, it is too late for ...

☆ Save Cite Related articles

[HTML] Developing a business analytics methodology: A case study in the foodbank sector [HTML]

GA Hindle, R Vidgen - European Journal of Operational Research

... on the analytics bandwagon, we have developed a business analytics methodology model and then use this analysis to identify older adults in Canada.

☆ Save Cite Cited by 76

[BOOK] The power of people: Using business analytics to improve business performance

N Guenole, J Ferrar, S Feinzig - European Journal of Operational Research

... along toward the collision between business analytics and other critical business questions that find use in business analytics.

Business analytics based on machine learning: A case study in the foodbank sector [PDF]

KR Varshney, A Mojsilović - IEEE International Conference on Data Mining

... and other critical business questions that find use in business analytics.

Cite

MLA Hindle, Giles A., and Richard Vidgen. "Developing a business analytics methodology: A case study in the foodbank sector." *European Journal of Operational Research* 268.3 (2018): 836-851. [HTML]

APA Hindle, G. A., & Vidgen, R. (2018). Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3), 836-851.

Chicago Hindle, Giles A., and Richard Vidgen. "Developing a business analytics methodology: A case study in the foodbank sector." *European Journal of Operational Research* 268, no. 3 (2018): 836-851.

Harvard Hindle, G.A. and Vidgen, R., 2018. Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*, 268(3), pp.836-851.

Vancouver Hindle GA, Vidgen R. Developing a business analytics methodology: A case study in the foodbank sector. *European Journal of Operational Research*. 2018 Aug 1;268(3):836-51. [PDF]

BibTeX EndNote RefMan RefWorks

# Appendices

- Mathematical proofs, algorithms, source codes for computer implementations, raw data, etc. are often included as appendices
- Appendices are enumerated (Appendix 1, Appendix 2,... or Appendix A, Appendix B,...)

## 4.3. Model estimation

The EM algorithm was used to estimate the unknown model parameters  $\eta$ ,  $\tau^2$ ,  $\bar{\mu}_i$ , and  $\sigma_i^2$  across projects  $i = 1, \dots, 5610$ . A detailed description of this algorithm is given in [Appendix E](#). The estimated values of  $\eta$  and  $\tau$  are  $-0.0347$  and  $0.1264$ , respectively. The estimated values of  $\bar{\mu}_i$  and  $\sigma_i$  across all projects range within  $[11.0, 20.9]$  and  $[0, 3.26]$ , respectively.



# Scientific articles to use as style-examples

## **Analytics development for a particular application:**

- <https://doi.org/10.1080/00207543.2018.1431413>
- <https://doi.org/10.1016/j.omega.2020.102207>
- <https://doi.org/10.1016/j.dss.2015.05.001>
- <https://doi.org/10.1111/dec.12448>
- <https://doi.org/10.1002/asi.24634>

## **Applying a model to open data:**

- <https://doi.org/10.1016/j.ejor.2020.03.042>
- <https://doi.org/10.1016/j.ejor.2021.04.018>

## **Behavioural and empirical studies:**

- <https://doi.org/10.1287/mnsc.2018.3245>
- <https://doi.org/10.1002/joom.1001>

## **Applying a model to company data:**

- <https://doi.org/10.1016/j.indmarman.2021.06.003>
- <https://doi.org/10.1016/j.indmarman.2020.12.012>

# Thesis project management

- **The student is responsible for project management**
  - Supervisor meetings: Scheduling, invitations, agendas, meeting notes...
  - Communications: regular status updates to supervisor; keeping the seminar teachers up-to-date on your supervisor, topic, progress...
  - Feedback: utilizing feedback, keeping track of changes and open questions...

# Questions?