Case study Guidelines

Course: Case Study: Model Engineering (DLMDSME01)

Table of contents

1. General information on Case Studies

Analyzing and working with case studies plays an important role in university-level coursework because it promotes important skills in the field of action- and solution-oriented thinking and decision-making. Case studies train the practical transfer of theoretical knowledge and especially in IT-related disciplines like data science they help to develop the skillset of how to translate business problems into valuable digital products.

1.1. What Are Case Studies?

Case studies are a teaching method in which you, the student, work on a "case" that demonstrates a situation entailing a realistic, practical (business) problem needing to be solved.

Case studies occur in a wide variety of contexts and are used for various purposes in professional practice, studies, teaching, and research. Classical case studies in the field of economics are often based on real life cases from the past. The data science case studies presented here are based on real life cases, where business stakeholders approach you, the data scientist, to ask if certain parts of the business can be automated via predictive modelling.

Case studies therefore do not contain a structured preparation of knowledge like textbooks do. The tasks within case studies are like real life: complex, incomplete, unstructured, imprecise, and ambiguous. By training them you develop the skillset of how to approach such problems. This is especially important for data scientists, where besides a sound theoretical basis, the experience on many different data sets is a crucial part of education.

1.2 What Do You Learn From Case Studies?

Case studies are used to gain useful insights through analysis and examination. Ideally, these findings can be transferred to other cases and situations. In this way, case studies support the development of analytical skills, sharpen the ability to separate the important from the unimportant, and open up new alternatives for action.

Applied learning with case studies is a central building block in this course. Some of you may have worked with case studies before. However, you have all gained professional experience in different areas, and you are familiar with the contents of this course from the teaching materials provided. With the available case studies, you can now combine your specific experiences and the concepts that you have learned from the course. Further information is provided in this guide and by the course tutor.

1.3 Tips for Working with Case Studies

Case study tasks are intentionally vague. It is part of the task to discover exactly what the case study is about. There is no correct method; instead, what is important here is the process the student engages in to find a valuable solution, such as considering various questions and different competing approaches toward finding a solution.

There are no right or wrong answers in case studies. What is more important is how the proposed solution is approached and justified.

Case study analysis, in general, is not about reproducing learned knowledge, i.e. explaining a method or a concept in detail, but about applying knowledge and experience to a situation that requires action or decisions, i.e. emphasizing the transfer of knowledge and experience.

Case studies in the module "Model Engineering", in particular, consist of conceptual and coding steps. Your observations and visualizations should be embedded into a final document; the format is described in detail in Chapter 4.3. For the coding part, the programming languages **python** and **R**, which are the most popular in the data science community, are recommended. Moreover, the open-source web application **Jupyter Notebook** (https://jupyter.org/) is very helpful for an exploratory study as you can embed code, visualizations and text into one single notebook file.

Do not forget to insert your well-documented code as an attachment to the final submission document since via Turnitin it is only possible to submit one document.

2. Case studies and Tasks

In this section, you can select one of the three listed case studies to work on (see sections 2.1, 2.2, 2.3).

When working on your case study, please take into account the task described in the respective case study itself as well as the notes on submission and formalities in section 4 and the criteria for evaluating the case study analysis in section 5.

In addition, each case study consists of a *.zip folder, which contains a *.csv dataset and *.txt files with further relevant information, like a description of the dataset. The *.zip folders will be made available in the course material.

The course book and in-depth information from the internet or suitable reference books serve as a basis for the case study. For instance, data science community websites, like **kaggle** (https://www.kaggle.com/), are a good source of open databases and public notebooks. It is expected that additional relevant literature and current academic sources will be researched and used to develop the chosen topic. These sources must be correctly referenced in the bibliography; the correct referencing methods are described in section 4.2.

2.1 Case Study 1: Credit Card Routing for Online Purchase via Predictive Modelling

Case description:

This is your first day as a data scientist at one of the world's largest retail companies. Already on your first day, you are invited to a meeting with important business stakeholders from the online payment department, who ask for your help: Over the past year they have encountered a high failure rate of online credit card payments. The company loses a lot of money due to failed transactions and customers become increasingly unsatisfied with the online shop.

Such online credit card payments are performed via so-called payments service providers, referred to as "PSPs" by the business stakeholders. Your company has contracts with four different PSPs and pays transaction fees for every single payment.

The current routing logic is manual and rule-based. Business decision makers, however, hope that with predictive modelling and with your help, a smarter way of routing a PSP to a transaction is possible.

Project Aim:

Help the business to automate the credit card routing via a predictive model. Such a model should **increase the payment success rate** by finding the best possible PSP for each transaction and at the same time **keep the transaction fees low.**

Data Set:

The data set and all relevant information from business side (name of PSPs, transaction fees) are given in a separate *.zip folder, which will be attached to the course material.

Task Description:

The task consists of both coding and conceptual steps. Here is a list of tasks, which should be included in your final document:

- **Structure the project** via the CRISP-DM or Team DS methodologies and give a recommendation of how a git repository for the project could look like. Note that you do not have to structure your final code according to your git-repository proposal.
- **Assess the quality** of the provided data set. **Prepare and visualize** your findings of the initial data analysis in order that business stakeholders can understand them in a clear and easy way.
- **Provide a baseline model** as well as an **accurate predictive model, which fulfils business requirements**, i.e. increase credit card success rate and keep fees low.
- In order that the business places confidence in your model, discuss the importance of the individual features and make the results of the model interpretable. Moreover, a sophisticated error analysis is very important for the business to understand the drawbacks of your approach.
- In the last step of the project, give a proposal of how your model could be used by the business in everyday work, for instance, via a graphical user interface (GUI).

2.2 Case Study 2: Automation of Standby Duty Planning for Rescue Drivers via a Forecasting Model

Case description:

As a data science consultant you are invited to work on a project within Berlin's red-cross rescue service. The HR planning struggles with the current standby-duty plan. They ask for you expertise on predictive models in order to improve the current planning logic.

Every day a certain amount of rescue drivers are on duty. However, due to short-term sickness of rescue drivers or unusual amount of emergency calls often more drivers are needed than initially expected. Hence, a certain

amount of standby-by drivers are kept "on hold" and activated when needed. This is why in the current approach 90 rescue drivers are daily kept on standby.

As colleagues from the planning claim, there are seasonal patterns - for instance, during winter months more rescue drivers call sick – which are not incorporated into the current approach. Moreover, sometimes there are not enough rescue drivers even when all 90 standby-bys are activated, so that drivers are called for work even on their days off.

It is important to know, that the duty plan has to be finished on the 15th of the current month for the upcoming month. This means, for instance, that the duty plan for November has to be finished on 15th of October.

Project Aim:

Help the HR planning to more efficiently estimate the **amount of standby rescue drivers on a daily basis via a prediction model**. Here *efficient* means that the percentage of standbys being activated is higher than in the current approach of keeping 90 drivers on hold. It also means that situations with not enough standbys should occur less often than in the current approach. Note that the plan has to be finished on the 15th of the current month for the upcoming month.

Data Set:

The data set and all relevant information from business side (sickness dataset, number of emergency calls per day) are given in a separate *.zip folder, which will be attached to the course material.

Task Description:

The task consists of both coding and conceptual steps. Here is a list of tasks, which should be included in your final document:

- **Structure the project** via the CRISP-DM or Team DS methodologies and give a recommendation of how a git repository for the project could look like. Note that you do not have to structure your final code according to your git-repository proposal.
- **Assess the quality** of the provided data set. **Prepare and visualize** your findings of the initial data analysis in order that business stakeholders can understand them in a clear and easy way.
- **Provide a baseline model** as well as an accurate **predictive model, which fulfils business requirements**, i.e. the amount of activated standby drivers is maximized, but having not enough standbys is minimized.
- In order that the business places confidence in your model, discuss the importance of the individual features and make the results of the model interpretable. Moreover, a sophisticated error analysis is very important for the business to understand the cases in which your approach could potentially fail.
- In the last step of the project, give a proposal of how your model could be used by the business in everyday work, for instance, via a graphical user interface (GUI).

2.3 Case study 3: Efficient Flight Operations at East Carmen Airlines via a Machine Learning Flight Prediction Model

Case description:

Welcome to East Carmen Airlines! As a strategic data scientist, you are confronted with many use cases to digitalize the airline business. Most important for airlines are accurate predictions of flight arrivals during a day of operations. When it comes to passenger handling, such predictions would admit a proactive steering, in order that the passengers reach the connection flights and therefore save a lot of money.
Your model should be applicable to each aircraft in East Carmen's fleet. Let us take as an example the aircraft with registration name EC-LPD. At scheduled departure of the earliest morning flight from East Carmen's hub Madrid the model should predict all flights of EC-LPD on the given day.

Many business stakeholders do not trust in the robustness of such a model. They claim that for a given aircraft, one could predict the first flight of the day with a decent error, but the errors of second and third flights will be huge. It is now your task to convince the business stakeholders that with a well-prepared data set and meaningful features you can deliver robust flight predictions.

Project Aim:

From different departments business analysts have collected a lot of information on possible features for flight predictions. Now it is your task to use these data to deliver a **prototype forecasting model for all flights of a given aircraft** at scheduled departure of the first flight of the day. For instance, for an aircraft with registration name EC-LPD, first flight of the day (FFOD) scheduled at 5:30 am from MAD to VIE and with the connection chain MAD-VIE-MAD-CDG-MAD-LHR-MAD your model should predict the successive landing times in VIE, MAD, CDG, MAD, LHR, MAD. The colleagues from daily operations are especially interested in how the **confidence level of such a model evolves during a day of operations.**

Data Set:

The data set and all relevant information from business side (operational-, weather data per flight, fleet list) are given in a separate *.zip folder, which will be attached to the course material.

Task Description:

The task consists of both coding and conceptual steps. Here is a list of tasks, which should be included in your final document:

- **Structure the project** via the CRISP-DM or Team DS methodologies and give a recommendation of how a git repository for the project could look like. Note that you do not have to structure your final code according to your git-repository proposal.
- **Assess the quality** of the provided data set. **Prepare and visualize** your findings of the initial data analysis in order that business stakeholders can understand them in a clear and easy way.
- **Provide a baseline model** as well as a **prototype predictive model, which fulfils business requirements** described above in the "project aim".
- In order that the business places confidence in your model, discuss the importance of the individual features and make the results of the model interpretable. Note that it is especially important for the business colleagues to understand the error of predictions for successive flights in the chain.

- In the last step of the project give a proposal of how your model could be used in everyday work, for instance, via a graphical user inferface (GUI). Discuss possible relevant information, which is not yet included into the data set and which could help to improve the forecast.

3. Tutorial support

Students have the option to make use of any one of several opportunities to get support for their case study with the course tutor. Taking advantage of these opportunities is the responsibility of the student and the use of these services is voluntary. The tutor for the course offers regular online tutorials (dates posted on myCampus). It is also possible to contact the tutor via email, especially with formal questions about the case study. Please note: a review of outlines and aspects of the case study is not intended here, since the student's ability to work independently is part of the evaluation and counts as a part of the overall assessment.

4. Formal guidelines and guidelines for submission

A total of 20% of the format, presentation, and language are included in the evaluation of your work on the case study. Therefore, greater attention should be paid to the following requirements in order to avoid points being deducted from the overall grade.

4.1 Quotations and Footnotes

It is absolutely forbidden to use ideas, statements and/or facts that are not the writer's own, without attributing the source. This is plagiarism. Citations should follow the American Psychological Association (APA) style. For direct quotes used in the text, use the following citation format: author's last name, year of publication, and page number, all in parentheses (name, year, page). For example:

The role of Mephisto can be described as follows: "Man loves peace and quiet, and for this reason he needs an opponent of roughly the same strength who constantly hits difficult balls in his life. This is the gift of Mephisto" (Schmidt 2004, p. 102). From this it becomes clear that...

Indirect quotes do not need page numbers, but do still need to include the author's last name and year of publication. While footnotes are not recommended by the APA style guide, if used they should be placed at the end of the corresponding page, and marked within the text by a superscript number[1].

[1] This is an example of how footnotes are used.

4.2 Bibliography

The bibliography is a mandatory part of every academic work. This is a record of the external sources and materials that have been used. A complete bibliography must be included with the submission. The titles are listed in alphabetical order and the bibliography should always be at the end.

The references are formatted according to APA style (the 6th edition has been used here).

- Book

Last name, initials. (year). *Title: Subtitle*. Publisher.

Fisher, R., Ury, W. L., & Patton, B. (2011). *Getting to yes: Negotiating agreement without giving in* (3rd ed.). Penguin Books.

- Article or Chapter in a Book

Last name, initials., & last name, initials. (year of publication). Title of chapter. In initials last name & initials. Editor (Eds.), *Title of book* (pages of chapter). Publisher. DOI if there is one.

Sutherland, F., & Smith, A. C. T. (2013). Leadership for the age of sustainability: A dualities approach to organizational change. In R. T. By & B. Burns (Eds.), *Organizational change, leadership and ethics: Leading organizations towards sustainability*. Routledge.

- Article in a Journal

Last name, initials. (year). Title of article. *Title of Periodical, volume number*(issue number), pages. DOI if there is one.

Galbraith, J. R. (2012). The future of organization design. *Journal of Organization Design, 1*(1), 3–6.

- Database

The APA discourages including database information in bibliographies because of the broad availability of such resources. APA recommends the following if citing a database is necessary:

When citing a work from one of these databases or platforms, do not include the database or platform name in the reference list entry unless the work falls under one of the exceptions described next (databases with original, proprietary content and works of limited circulation).
Likewise, do not include URLs from these academic research databases in reference list entries because these URLs will not resolve for readers.
Instead of a database URL, include a DOI if the work has one.

(https://apastyle.apa.org/style-grammar-guidelines/references/database-information)

- Chapter or Section of a web page

Last name, initials., & last name, initials. (Date of publication). Title of article. In *Title of book or larger document* (chapter or section number). Retrieved from Full URL

Paiz, J.M., Angeli A., Wagner, J., Lawrick L., Moore K., Anderson M., Soderlund L., ... Keck R. (2013). Reference list: Basic rules. In *Perdue Online Writing Lab*. Retrieved from https://owl.english.purdue.edu/owl/resource/560/05/

- Web Article

Last name, initials. (Year, Month Day). Title of article. *Title of Newspaper*. Retrieved from full URL

Fagan, J. (2019, March 25). Nursing clinical brain. *OER Commons.* Retrieved from https://www.oercommons.org/authoring/nursing-clinical-brain/view

Below are some further regulations regarding literature references:

For more than three authors, name only the first author, followed by et al.

If data is missing, abbreviate accordingly:

- No date = n.d.
- No author = (Title, year)
- No title = (Author, year)

Sort several titles from the same author by year of publication (from oldest to youngest); distinguish several titles from the same year with a, b, c ... by year.

The first edition is generally not mentioned.

4.3 Format

Scope	7-10 pages of text
Format of the paper	DIN A4
Margins	Top and bottom 2cm; left 2cm; right 2cm
Type	General text: Arial 11 pt.; Headings Arial 12 pt.
Line height	1.5
Setting	Justified; hyphenation
Footnote	Arial 10 pt., Justified
Paragraphs	6 pt. distance after line break
Headings and Subheadings	Maximum 3 levels (1.Main heading, 1.1 Section, 1.1.1 Subheadings)
	The preceding and following technicalities numbered with Roman numerals (I. List of abbreviations, II. List of illustrations, III. List of tables…)
	Do not underline, use "*italics*" sparingly to emphasize passages
Anti-Plagiarism Pledge	The final case study should be submitted electronically via myCampus. No written draft should be submitted before the due date. Please refer to the "Instructions for submitting a paper to myCampus - Turnitin".

4.4 Submission

Submit the case study through the online platform Turnitin. The instructions for submitting your work can be found in a separate manual on myCampus. There you can also find out how you can view your grade directly on Turnitin after it's published on CARE. Please note that it is not possible to hand in your submission by email or any other means.

For questions about Turnitin please contact pruefungsamt@iubh-fernstudium.de.

5. Evaluation of Your Case Study

The following criteria are used to evaluate the case study:

Evaluation Criteria	Explanation	Weight
Identification	Definition of the problem	10%
Concepts	Application of concepts	25%
Analysis	Quality of analysis	30%
Conclusion	Conclusion and recommendations	15%
Formalities	Compliance with formal requirements	5%
Accuracy	Correctness of spelling and punctuation	5%
Language	Linguistic expression	10%

The analysis of the case study and its processing should take into account the evaluation criteria and explanations listed.

Identification: How is the business problem translated into data science framework, i.e. which data science use case can be established from the given business problem?

- Example: „For the problem at hand we first apply a supervised learning classification algorithm. From the predicted probabilities we find a rule-based logic for the system."

Concepts: Application of relevant concepts learned throughout the course.

- Examples are:
    - how to perform the initial data analysis, like calculating correlations
    - which models are used for predictions (Linear Regression, Random Forest…)
    - which techniques are used to select import features
    - how to prevent overfitting (train hyperparameters, use regularization…)
    - …

Analysis: Deliver a sophisticated, detailed analysis of the data and the modelling approach.

- Example: "In our initial data analysis we observe that for Argentina the number of customers decreased by 50%, whereas in all other countries the decrease is on average 10%. This means that in Argentina we face a huge problem with our business."
- Example: „We observe that choosing a random forest classifier increases the overall accuracy of the model by more than 5%. Moreover, when we apply a trigonometric function on the temporal features the accuracy increases by another 5%."

Conclusion: Describe the important results of your analysis in a clear and quantitative way. This goes hand in hand with the analysis. How can your model be used in everyday work? Are there any improvements necessary until the model goes to production?

- Example: "From the analysis of the model we conclude, that our new data-driven approach outperforms the old, rule-based approach in our test set by 20%. We did not find any situation where the rule-based approach gives better results. This is why we are confident that the prediction model will lead to a significant cost reduction. Our conservative estimate for the next two months is a cost reduction of more than 6% due to the prediction model."

Formalities: Adherence to the guidelines of chapter 4.3

Accuracy: Correctness of spelling and punctuation

Language: Quality of the linguistic expression and adequacy of language style for scientific work

6. Further references and sources

Cengage Learning: Case Studies http://college.cengage.com/business/resources/casestudies/students/overview.htm

Pearson Education: How to Analyze a Case Study http://wps.prenhall.com/bp_laudon_essmis_6/21/5555/1422312.cw/content/index.html

William Ellet: The Case Study Handbook: How to Read, Discuss, and Write Persuasively About Cases.

William Ellet (2018): The Case Study Handbook, Revised Edition: A Student's Guide available at: https://ebookcentral-proquest-com.pxz.iubh.de:8443/lib/badhonnef/detail.action?docID=5180070#

Good luck with your case study!