

# Bachelor thesis plan

Spring 2024

Department of Advanced Computing Sciences

Board of Examiners

---

## Personal Details

Name: *(Student's family name and first names)*

Buiter Sánchez Arantxa

Student number: *(Student's ID number as provided by university)*

i6294040

Start date of thesis research *(Cannot be prior to receiving BoE permission, hence must be in the future)*

February 5, 2024

Expected graduation date: *(Date of the regular bachelor conference where the thesis will be defended)*

Bachelor conference spring 2024: 21, 24 and 25 June 2024 (resit 23 and 26 August 2024)

Programme:

Data Science and Artificial Intelligence

☐ The student is enrolled in the KE@Work programme *(check if applicable)*

☐ The student is enrolled in the Marble 2.0 programme *(check if applicable)*

---

## Supervision details

*The thesis supervisors are DACS staff members that give process based and content-based supervision. At least one of the thesis supervisor acts as an examiner. External supervisors are listed in the next section.*

Name DACS thesis supervisor 1: Enrique Hortal Quezada

Name DACS thesis supervisor 2: optional

Name DACS thesis supervisor 3: optional

Frequency of contact between DACS thesis supervisor(s) and student:

weekly

## External collaboration

☒ This thesis involves an external collaboration *(check if an external party is actively involved)*

**In case this thesis is not part of an external collaboration proceed to the section "assessment".**

Name external institute: Integrity Software Latin America *(Name the institute of external collaboration. In the case of multiple institutes, name the most important one and add an appendix with an overview of external parties)*

This is: a company

Location external institute: Mexico City, Mexico

Name external supervisor: Rodrigo Castañeda García

Frequency of contact between the external supervisor(s)/advisor(s) and student: weekly

Frequency of contact between the external supervisor(s)/advisor(s) and DACS thesis advisor(s): monthly

☒ The student will be hosted by/have a secondment at the external institute *(Check this if the student will work as an internal at the external institute or be visiting the external institute for a period larger than two weeks)*

The student will be hosted by the external institute for:

- ☒ The entire period  
☐ A period of \_\_\_\_\_ weeks

During this period, the work times of the student are as follows: 8 hours per day, 5 days a week  
And the student will spend 40 hours per week on thesis related work.

☒ There is a confidentiality agreement with the following clauses: *(check when applicable)*

- ☒ The code developed will remain property of the external institute  
☐ Special permission has been obtained for limited confidentiality *(See Rules and Regulations and attach documents)*

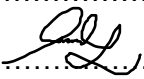
Note that all thesis must be sent to the DACS student affairs office for storing, where they can be labelled “confidential”, but DACS committees and assessment committees (visitation commissions) will always have (confidential) access for auditing purposes.

The student will receive from the external institute:

- ☒ No financial compensation  
☐ An internship fee  
☐ Salary  
☐ Travel costs reimbursement

On behalf of the external institute, *Integrity Software Latin America* consents with this thesis plan

Date of signing: ..... 01/22/2024

Signature: .....  .....

## Assessment

The assessment will be based on:

*(Specify the terms of assessment that at least should contain the thesis, process and the presentation)*

- ☒ The contents and form of the thesis, which must at least be sufficient for passing the thesis  
☒ A (software) product that accompanies the thesis  
☒ The presentation  
☒ The process  
☐ (optional)

## Thesis details

Thesis title: *(Working title of the thesis)*

Utilising Advanced Data Analytics for the Detection and Prevention of Fraudulent Activities

Description of the assignment: *(A single page, describing the assignment and indications of the techniques that will be explored)*

In this research project, the main focus is to employ data analysis techniques to uncover patterns that could be indicative of fraudulent activities. As fraudulent practices become more common in different areas, the methods used to detect them should be updated accordingly. The project specifically aims to contribute to the domain of cybersecurity and risk management by exploring the complexities of fraud detection through data analytics.

The research will adopt a varied approach, combining exploratory data analysis, machine learning algorithms, and statistical modelling. The dataset will be assembled to include relevant variables and features necessary for robust analysis. This is an already existing private dataset stored in an Amazon server and it includes information related to the treasury management system (TMS) implemented by the company for demonstrational purposes. Access to this dataset will be granted through an account provided by the company. The application of various data analysis tools will be measured to ensure the extraction of meaningful insights, allowing for a comprehensive understanding of fraud dynamics.

Awoyemi, O. et al (2017) use machine learning techniques in order to classify credit card transactions. Additionally, Amusan, E. (2021) took the fact that the dataset is imbalanced and resolved this by using the undersampling technique, and trained the model on the now balanced data set. Varmedja, D. et al (2019) do this similarly, by using Synthetic Minority Oversampling Technique (SMOTE) instead of under sampling. In order to avoid information loss, SMOTE will be used for this assignment. By doing so, the machine learning classifier will improve in its performance since it won't be biased towards the majority class. In the three referenced papers, the classifiers achieve a high performance however the best classifier differs. Having said that, the approach taken by Varmedja, D. et al (2019) seems to be the most efficient, as it has the highest accuracy. It is important to note that the accuracy varies because the models are not trained on the same data, however it can also be an indicator of this being the best approach.

The project will continue with the implementation of a baseline agent for fraud detection. This initial agent will serve as a benchmark for assessing the effectiveness of subsequent modifications and the integration of AI models. Following this, the baseline agent will undergo enhancements, incorporating machine learning techniques to improve its ability to detect and prevent fraudulent activities.

In addition, other agents will be introduced, starting with an agent that employs a machine learning approach, in the case of having a low performance, another approach would be considered, or the combination of more than one approach would be implemented. The final agent would be a combination of SMOTE and the best-performing Machine Learning technique(s). SMOTE is employed to ensure a balanced dataset, since the genuine transactions are expected to be of a larger number than the fraudulent ones.

The research will explore the efficacy of three or more distinct agents, one of them being the baseline agent, the machine learning agent as well as the combined agent. These agents will be compared based on their performance metrics and accuracy in identifying fraudulent patterns. This comparative analysis will provide insights into the strengths and weaknesses of the different approaches, potentially contributing to improvements in fraud detection methodologies.

The findings of this research are anticipated to contribute to the field of fraud detection and prevention. By discovering patterns and exploiting advanced analytics. This project signifies a commitment to advancing the limits of fraud detection through data analysis methodologies. By combining theoretical frameworks with practical applications, the research aspires to provide awareness, promoting a more secure landscape for businesses and institutions.

## References

- [1] J. O. Awoyemi, A. O. Adetunmbi and S. A. Oluwadare, "Credit card fraud detection using machine learning techniques: A comparative analysis," 2017 International Conference on Computing Networking and Informatics (ICCNI), Lagos, Nigeria, 2017, pp. 1-9, doi: 10.1109/ICCNI.2017.8123782.
- [2] Amusan, Elizabeth & Oluwaseun, Alade & Emuoyibofarhe, J & Fenwa, Olusayo. (2021). Credit Card Fraud Detection on Skewed Data using Machine Learning Techniques. 2. 49-56.
- [3] D. Varmedja, M. Karanovic, S. Sladojevic, M. Arsenovic and A. Anderla, "Credit Card Fraud Detection - Machine Learning methods," 2019 18th International Symposium INFOTEH-JAHORINA (INFOTEH), East Sarajevo, Bosnia and Herzegovina, 2019, pp. 1-5, doi: 10.1109/INFOTEH.2019.8717766.

**Planning:** *(Provide a complete planning for the thesis period with a reasonable time path for completion, including time for writing the thesis. The preferred format is a Gantt chart that you can attach separately)*

A more extensive gantt chart has been attached.

	week 1	week 2	week 3	week 4	week 5	week 6	week 7	week 8	week 9	week 10	week 11	week 12	week 13	week 14	week 15	week 16	week 17	week 18	week 19
1. Finding relevant sources	05/02/2024	12/02/2024	19/02/2024	26/02/2024	04/03/2024	11/03/2024	18/03/2024	25/03/2024	01/04/2024	08/04/2024	15/04/2024	22/04/2024	29/04/2024	06/05/2024	13/05/2024	20/05/2024	27/05/2024	03/06/2024	10/06/2024
2. Creating a template for the report																			
3. Study collected data and analyse it																			
4. Create the different agents and compare them																			
5. Preparing the first draft																			
6. Thesis checkpoint																			
7. Finish up the sections																			
8. Thesis submission																			

**Problem statement:** *(single sentence describing the overall challenge faced by the student)*

The current state-of-the-art for fraud detection lacks a unified and adaptive framework that integrates advanced data analysis and artificial intelligence models, impeding the development of a comprehensive system capable of efficiently identifying and preventing diverse forms of fraudulent activities.

**Research questions:** *(provide two to five research questions of a single sentence that by answering them, jointly provide a path for addressing the problem statement)*

1. How does the baseline agent for fraud detection compare in terms of performance metrics and accuracy with the enhanced agents incorporating machine learning techniques and the combined agent utilising both the SMOTE technique and the best-performing machine learning approach?
2. What are the key patterns, indicators, or anomalies identified by the exploratory data analysis, and how do these contribute to the development and improvement of fraud detection models in the context of TMS data?
3. How does the combined agent, incorporating the SMOTE technique and the best-performing machine learning approach, outperform or complement the baseline agent and the machine learning agent in terms of early detection and prevention of fraudulent activities, and what insights can be drawn to enhance fraud detection methodologies across industries?

## Signature

The student would like to make the following additional statements to the Board of Examiners regarding this thesis plan:

(optional)

Ensure that you: *(Ensure that you check all)*

- ☒ are familiar with the Education and Examination Regulations (EER) that are published on student portal, as is expected of each student
- ☒ are familiar with the Rules and Regulations that are published on student portal, as is expected of each student
- ☒ are familiar with the concept of plagiarism and will refrain from committing it
- ☒ are aware of academic standards, in particular relating to proper citing
- ☒ have filled out the complete form
- ☒ have included a planning and/or attached a Gantt chart with a planning for your thesis
- ☒ have collected the signatures of both examiners and if applicable, the external institute

To avoid students starting too late with their thesis, the deadline for the thesis plans has been set by the thesis coordinator before some of the grades of the preceding semester have come in. Therefore, the possibility of getting conditional approval has been given.

In separate, motivated requests to the Board of Examiners, dispensation from this rule can be requested for cases of hardship. Such motivated requests can be submitted to [dacs-boe@maastrichtuniversity.nl](mailto:dacs-boe@maastrichtuniversity.nl) and are accompanied by a study plan.

- ☒ I understand that I must have attained 140 ECTS already for this bachelor programme, of which 60 ECTS have been obtained in year 1 and at least 40 ECTS in year 2. I declare that (check one of the following):
  - ☒ I have already obtained the required ECTS
  - ☐ I do not yet have the required number of ECTS, but expect that once the grades of the resits, project/KE@work/MaRBL 2.0 that took place in the semester preceding the thesis semester come in, I will have enough ECTS. Hence, I'm submitting for conditional approval and understand that, if the expected ECTS are not obtained, any conditional approval of the thesis will be automatically withdrawn.
  - ☐ I do not have enough ECTS, but are submitting this plan together with a motivated request for dispensation due to hardship and an individual study plan.

Date: January 18, 2024

Student's signature

.....

## Examiners

*There should be two thesis examiners that approve of this thesis plan. At least one of the examiners acts as a thesis supervisor. The examiners must be DACS staff members (lecturer, assistant, associate, or full professor) with a PhD degree, a uniform teaching qualification (UTQ) and are involved in the programme. See Rules and Regulations for additional conditions. Postdoctoral researchers can only act as a second examiners as part of a training programme.*

The examiners have checked the form for the requirements in the EER and R&R and checked it with the ILOs. The examiners are also available in Maastricht for the bachelor conference indicated.

Name proposed thesis examiner 1:

Signature:

Date of signing:

Choose one from A-M

Enrique Hortal Quesada

OR one from N-Z



18/01/2024

Name proposed thesis examiner 2:

Signature:

Date of signing:

Choose one from A-M

OR one from N-Z

Yusuf Can Semerci



24/01/2024