Personal Development Report

Evaluation 3

Danil Burov

Fontys University of Applied Sciences

January 14, 2025

Contents

1 Introduction

My name is Danil Burov and I am studying Information Technology in Fontys, Venlo. I am specializing in Embedded Software back in Venlo and the reason I chose this minor is because I think that AI is a very ongoing topic and I wanted to get more familiar with it. Especially because I think that AI has a big implication when it comes down to IoT devices.

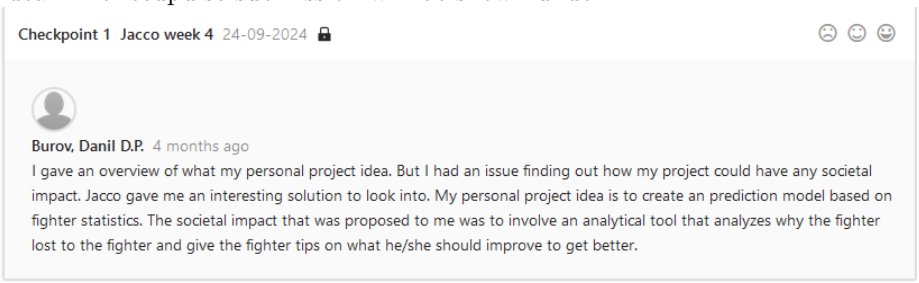
2 Learning outcome 1 - Societal impact

Description
The student is able to approach the context and impact of their own AI project(s) from different perspectives in a sustainable way. In addition, the student is able to reflect on their own choices, taking into account data legislation and the (possible) impact on society.

Explanation:
Societal impact is one of the most important contextual parts of the project. Every time a person starts a project an evaluation on the societal impact should be done in order to understand 'Why?' a project is being done on the given topic.

2.1 First evaluation

During the first weeks of the minor I had the opportunity to think about the societal impact of the projects I will be working on. Since the beginning I had an idea of what my personal project could look like. However, I was not that aware what could be the societal impact of my project. With the help of the technical coach 'Jacco' I established an understanding what could be a potential idea. The feedpulse submission will be shown under:



Self assessment: Orientating

2.2 Second evaluation

From week six to week nine I did not really improve on my societal impact learning outcome. This is mainly due to the reason that I was mainly focusing on modelling my first model and learning how to prepare data for training. In the upcoming weeks I will make sure that I take time as well to work on

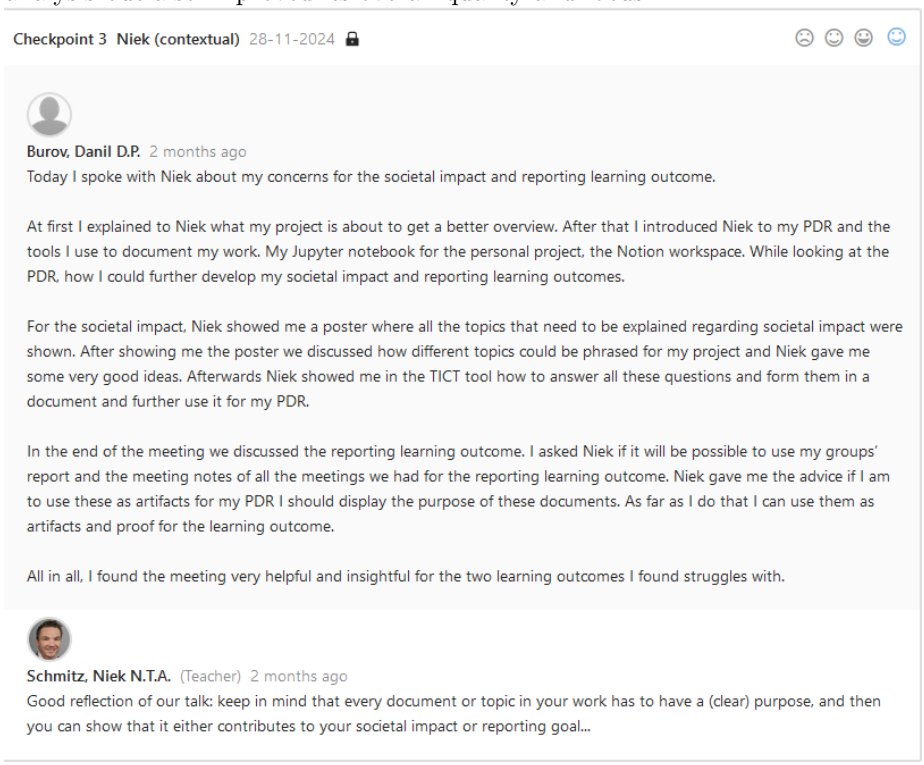
the societal impact goal as well, starting with writing the 'Potential Impact Assessment'.

Self assessment: Orientating

2.3 Third evaluation

During the third evaluation of my project, I delved deeper into analyzing the potential societal impact it might have. With the guidance of my teacher, Niek, I was able to create a solid first draft of the impact analysis. Niek has been a consistent source of support since I began working on this document, helping me navigate its purpose and structure.

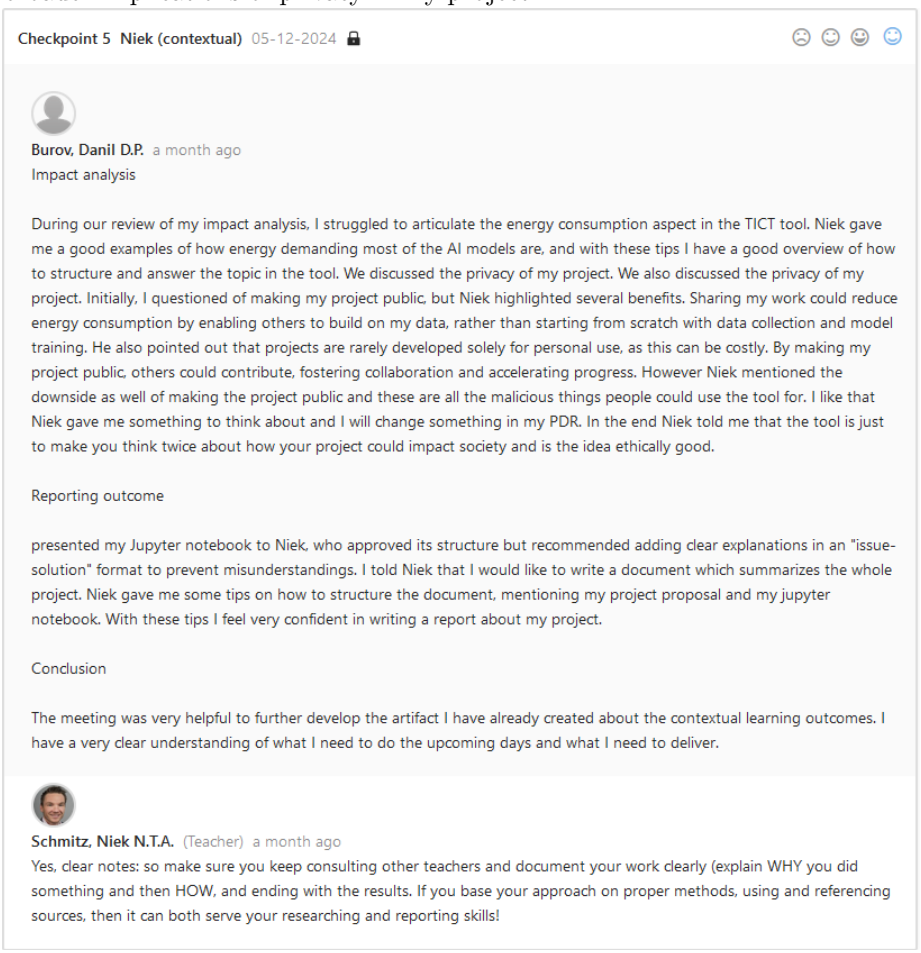
Initially, I struggled to see the relevance of creating an impact analysis. It seemed like an unnecessary artifact. However, as I started writing and consulting with Niek, I began to recognize its significance. The process encouraged me to critically consider the potential impact of my project on various societies, as well as on the individuals whose data I am analyzing and using to train the model. Niek's advice to clearly define the purpose of any document proved invaluable. This guidance not only helped me articulate the intent of the impact analysis but also improved its overall quality and focus.



Despite this progress, I encountered challenges, particularly in aligning the

context of my project with the purpose of the analysis. While I understood the document's goals and expectations, addressing the topic of energy consumption proved difficult. I wasn't entirely sure what was required or how to approach it. Niek's support was crucial here; he helped clarify the energy consumption topic, enabling me to tackle it with a better understanding.

In the first draft of the impact analysis, I was confident that my project should remain strictly private. However, during a feedback session with Niek, he raised a compelling point that challenged my perspective on privacy. His insights made me reconsider my initial stance and reflect more deeply on the broader implications of privacy in my project.



Checkpoint 5 Niek (contextual) 05-12-2024

Burov, Danil D.P. a month ago
Impact analysis

During our review of my impact analysis, I struggled to articulate the energy consumption aspect in the TICT tool. Niek gave me a good examples of how energy demanding most of the AI models are, and with these tips I have a good overview of how to structure and answer the topic in the tool. We discussed the privacy of my project. We also discussed the privacy of my project. Initially, I questioned of making my project public, but Niek highlighted several benefits. Sharing my work could reduce energy consumption by enabling others to build on my data, rather than starting from scratch with data collection and model training. He also pointed out that projects are rarely developed solely for personal use, as this can be costly. By making my project public, others could contribute, fostering collaboration and accelerating progress. However Niek mentioned the downside as well of making the project public and these are all the malicious things people could use the tool for. I like that Niek gave me something to think about and I will change something in my PDR. In the end Niek told me that the tool is just to make you think twice about how your project could impact society and is the idea ethically good.

Reporting outcome

presented my Jupyter notebook to Niek, who approved its structure but recommended adding clear explanations in an "issue-solution" format to prevent misunderstandings. I told Niek that I would like to write a document which summarizes the whole project. Niek gave me some tips on how to structure the document, mentioning my project proposal and my jupyter notebook. With these tips I feel very confident in writing a report about my project.

Conclusion

The meeting was very helpful to further develop the artifact I have already created about the contextual learning outcomes. I have a very clear understanding of what I need to do the upcoming days and what I need to deliver.

Schmitz, Niek N.T.A. (Teacher) a month ago
Yes, clear notes: so make sure you keep consulting other teachers and document your work clearly (explain WHY you did something and then HOW, and ending with the results. If you base your approach on proper methods, using and referencing sources, then it can both serve your researching and reporting skills!

Since then, I have continued to refine the impact analysis. I've been experimenting with different approaches, seeking feedback, and iterating on the document. This process has extended into the fourth quarter, and I'm committed to further improving the analysis to ensure it accurately captures the societal impact of my work.

Self assessment: Beginning

2.4 Final evaluation

Self assesment: Proficient

3 Learning outcome 2 - Investigative problem solving

Description
The student is able to critically look at their own AI project(s) from different perspectives, recognize problems and come up with appropriate solutions.

Explanation:
From this learning outcome I will further develop my investigative and problem solving skills regarding how AI is used in the project, is it doing as intended and if not how to approach the problem and solve it eventually.

3.1 First evaluation

During the first weeks I had to tackle several problems. In the beginning, the course was presented with some stakeholders who presented to us what issues they have with their companies and what they would like to add to their business. I was given the opportunity to think of an idea that solves a business case and so I created a project proposal for one of the companies that visited us. The proposal is based on the problem that the company has.

To see the full proposal go to this link: project proposal

Self assessment: Orientating

3.2 Second evaluation

I have been dealing with a legal issue for the past weeks because, for my personal project, I would love to have as much actual data as possible. In order to do that, I needed to develop a scraper for the websites' statistics. Before doing so I checked the terms and conditions of the website. Unfortunately, the website forbids any scraping and bots developed for the purpose of gathering data. So I investigated a way to still be able to scrape the data and use it for my personal project. In the upcoming weeks I will create a document where I explicitly say what is the intended usage of my project and why the scraper was developed.

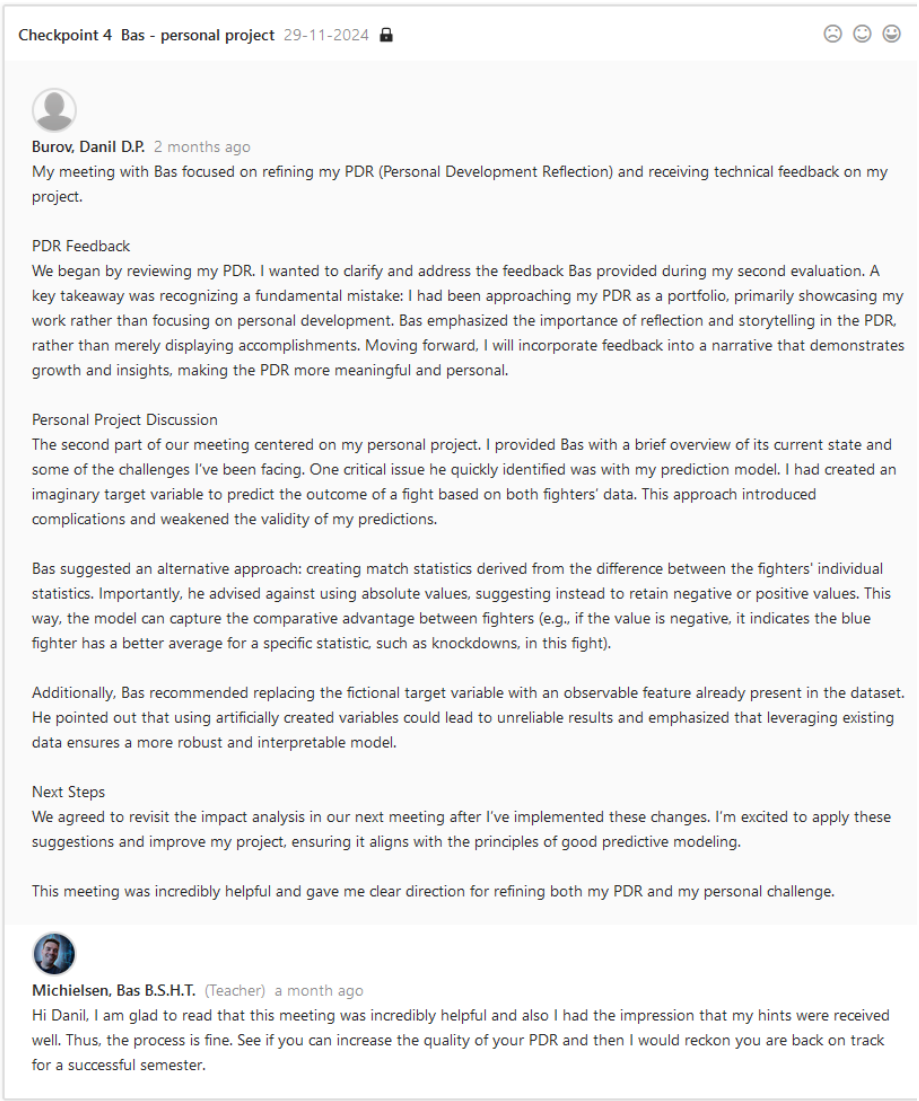
Self assessment: Beginning

3.3 Third evaluation

During this quarter of the semester, I also encountered challenges with my personal project, particularly in data preparation, analyzing societal impacts, and choosing a model for the data. Writing the societal impact analysis required me

to deeply consider who might use my project and how it could potentially be exploited. This exercise helped me brainstorm solutions to mitigate these risks. It also encouraged me to think about the broader future impact of my project, both positive and negative.

Regarding data preparation, I dealt with a significant amount of unstructured data that required cleaning and organizing for modeling. With the guidance of my semester coach, I successfully created a well-structured and cleaned dataset, ready for modeling.



Checkpoint 4 Bas - personal project 29-11-2024

Burov, Danil D.P. 2 months ago

My meeting with Bas focused on refining my PDR (Personal Development Reflection) and receiving technical feedback on my project.

PDR Feedback

We began by reviewing my PDR. I wanted to clarify and address the feedback Bas provided during my second evaluation. A key takeaway was recognizing a fundamental mistake: I had been approaching my PDR as a portfolio, primarily showcasing my work rather than focusing on personal development. Bas emphasized the importance of reflection and storytelling in the PDR, rather than merely displaying accomplishments. Moving forward, I will incorporate feedback into a narrative that demonstrates growth and insights, making the PDR more meaningful and personal.

Personal Project Discussion

The second part of our meeting centered on my personal project. I provided Bas with a brief overview of its current state and some of the challenges I've been facing. One critical issue he quickly identified was with my prediction model. I had created an imaginary target variable to predict the outcome of a fight based on both fighters' data. This approach introduced complications and weakened the validity of my predictions.

Bas suggested an alternative approach: creating match statistics derived from the difference between the fighters' individual statistics. Importantly, he advised against using absolute values, suggesting instead to retain negative or positive values. This way, the model can capture the comparative advantage between fighters (e.g., if the value is negative, it indicates the blue fighter has a better average for a specific statistic, such as knockdowns, in this fight).

Additionally, Bas recommended replacing the fictional target variable with an observable feature already present in the dataset. He pointed out that using artificially created variables could lead to unreliable results and emphasized that leveraging existing data ensures a more robust and interpretable model.

Next Steps

We agreed to revisit the impact analysis in our next meeting after I've implemented these changes. I'm excited to apply these suggestions and improve my project, ensuring it aligns with the principles of good predictive modeling.

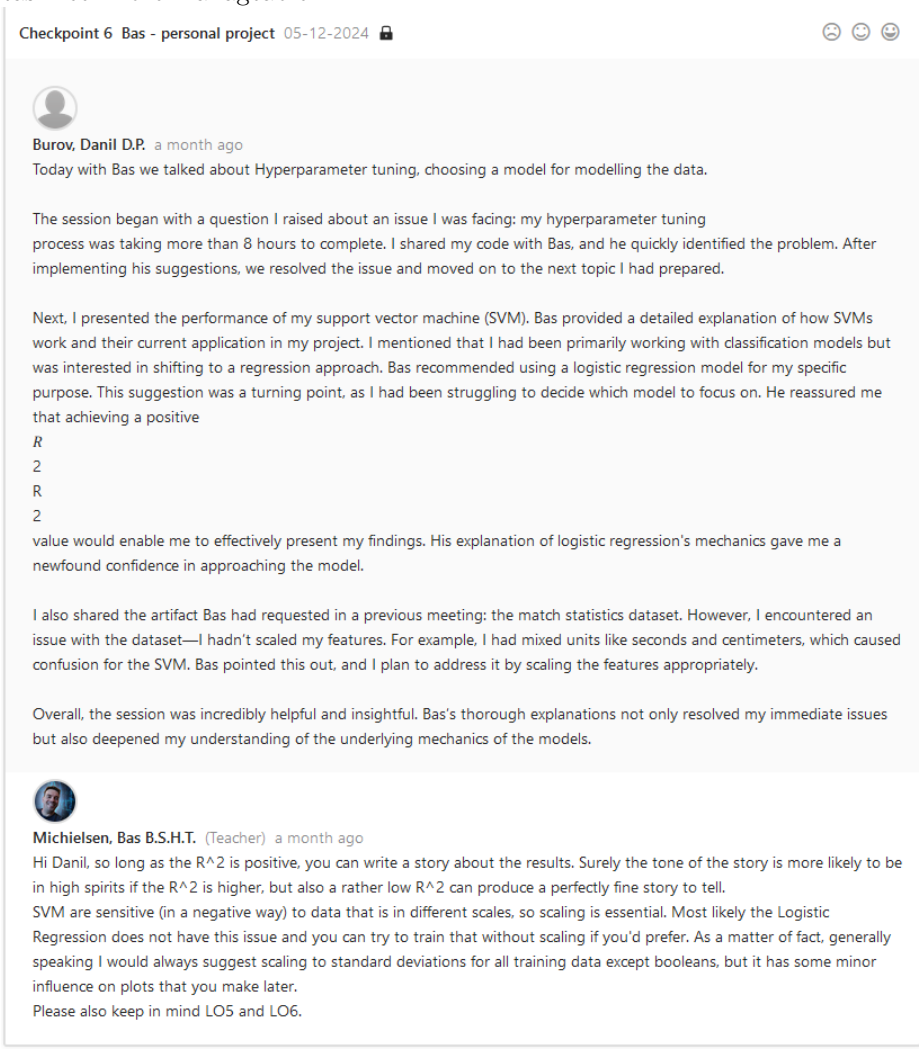
This meeting was incredibly helpful and gave me clear direction for refining both my PDR and my personal challenge.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, I am glad to read that this meeting was incredibly helpful and also I had the impression that my hints were received well. Thus, the process is fine. See if you can increase the quality of your PDR and then I would reckon you are back on track for a successful semester.

Choosing a model proved to be the most challenging aspect for me. I strug-

gled because I needed to understand the underlying technology and how it aligned with my project's goals. Through extensive self-research and discussions with Bas, I realized that selecting a model should not be the immediate priority. Instead, Bas emphasized the importance of building a robust dataset first. As he put it, "Don't rush into modeling; focus on creating a strong dataset, and the modeling will follow." This advice shifted my perspective and made the task feel more manageable.



Checkpoint 6 Bas - personal project 05-12-2024

Burov, Danil D.P. a month ago

Today with Bas we talked about Hyperparameter tuning, choosing a model for modelling the data.

The session began with a question I raised about an issue I was facing: my hyperparameter tuning process was taking more than 8 hours to complete. I shared my code with Bas, and he quickly identified the problem. After implementing his suggestions, we resolved the issue and moved on to the next topic I had prepared.

Next, I presented the performance of my support vector machine (SVM). Bas provided a detailed explanation of how SVMs work and their current application in my project. I mentioned that I had been primarily working with classification models but was interested in shifting to a regression approach. Bas recommended using a logistic regression model for my specific purpose. This suggestion was a turning point, as I had been struggling to decide which model to focus on. He reassured me that achieving a positive

R^2

R
2

value would enable me to effectively present my findings. His explanation of logistic regression's mechanics gave me a newfound confidence in approaching the model.

I also shared the artifact Bas had requested in a previous meeting: the match statistics dataset. However, I encountered an issue with the dataset—I hadn't scaled my features. For example, I had mixed units like seconds and centimeters, which caused confusion for the SVM. Bas pointed this out, and I plan to address it by scaling the features appropriately.

Overall, the session was incredibly helpful and insightful. Bas's thorough explanations not only resolved my immediate issues but also deepened my understanding of the underlying mechanics of the models.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, so long as the R^2 is positive, you can write a story about the results. Surely the tone of the story is more likely to be in high spirits if the R^2 is higher, but also a rather low R^2 can produce a perfectly fine story to tell.
SVM are sensitive (in a negative way) to data that is in different scales, so scaling is essential. Most likely the Logistic Regression does not have this issue and you can try to train that without scaling if you'd prefer. As a matter of fact, generally speaking I would always suggest scaling to standard deviations for all training data except booleans, but it has some minor influence on plots that you make later.
Please also keep in mind LO5 and LO6.

I'm excited about the upcoming challenges in the final quarter of the semester. I feel like I'm on the right track and look forward to tackling new problems and learning from them.

Self assessment: Beginning

3.4 Final evaluation

Self assesment: Proficient

4 Learning outcome 3 - Data preparation

Description
The student is able to collect data and estimate its quality and usability. The student is also able to adjust the data if necessary for proper usage in their project(s).

Explanation:
In order to create any AI model I will need a good, clean dataset that can be used to properly teach a machine. To do so I intend to first of all use a reliable source for the set and inspect it carefully before using.

4.1 First evaluation

The first thing that was done regarding this learning outcome was to select a dataset suitable for the intended usage of the AI model. For the personal project I have selected this dataset. I have been working on a cleaned version of this dataset. Since this dataset does have some empty fields I tried already cleaning some of the data to prepare it for the upcoming modeling. Regarding the data preparation for the project, I have already started analyzing the data for the organization we are doing the project for, however it is still in a very early stage.

Self assessment: Orientating

4.2 Second evaluation

I have been preparing data for the past couple of weeks for both my personal and group projects. In the past weeks in the group project, I have been trying to estimate the quality and the usability of the data that was provided by our stakeholder. To assess the quality and usability I cleaned the data from any 'null' fields and factorized some of the features. I also combined features into a new feature. Especially in my personal project, I had to check each fighters' statistics and based on them I labeled the fighters' fight style.

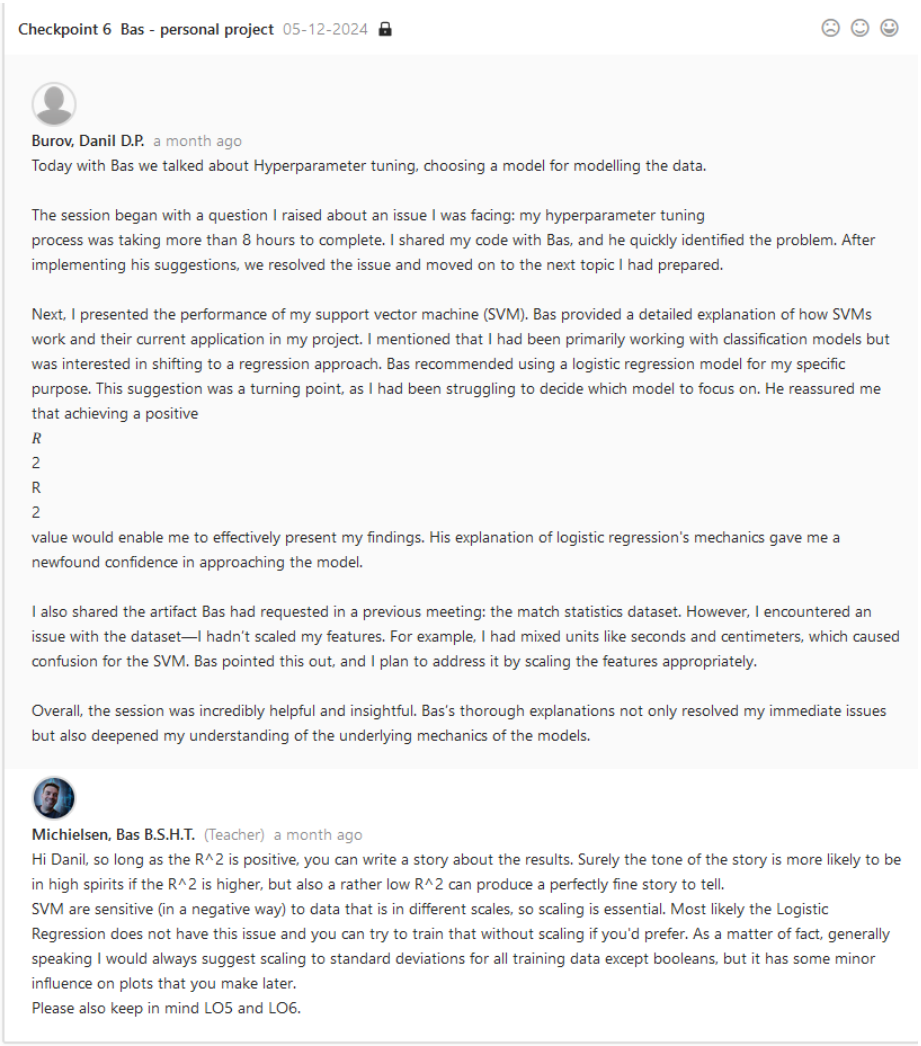
Self assessment: Beginning

4.3 Third evaluation

Since the last evaluation, I feel I have significantly improved in the data preparation outcome. My semester coach and I have worked extensively with a semi-large dataset that contained many unnecessary features. At the beginning of the minor, I wasn't entirely sure what cleaning data entailed. However, I now feel confident in how to approach and start this process. That said, I acknowledge

there's still much to learn.

In my most recent feedback session with my coach, a major issue in my dataset was identified: I hadn't applied standard deviation, which left certain numerical values (e.g., seconds and centimeters) unnormalized. This oversight was particularly problematic for models like Support Vector Machines, which require properly cleaned data to perform well. This feedback session taught me the importance of always double-checking my work, even when I feel confident in its accuracy. I also learned that the quality of the dataset directly determines the model's performance.



Checkpoint 6 Bas - personal project 05-12-2024

Burov, Danil D.P. a month ago

Today with Bas we talked about Hyperparameter tuning, choosing a model for modelling the data.

The session began with a question I raised about an issue I was facing: my hyperparameter tuning process was taking more than 8 hours to complete. I shared my code with Bas, and he quickly identified the problem. After implementing his suggestions, we resolved the issue and moved on to the next topic I had prepared.

Next, I presented the performance of my support vector machine (SVM). Bas provided a detailed explanation of how SVMs work and their current application in my project. I mentioned that I had been primarily working with classification models but was interested in shifting to a regression approach. Bas recommended using a logistic regression model for my specific purpose. This suggestion was a turning point, as I had been struggling to decide which model to focus on. He reassured me that achieving a positive

R^2

R
2

value would enable me to effectively present my findings. His explanation of logistic regression's mechanics gave me a newfound confidence in approaching the model.

I also shared the artifact Bas had requested in a previous meeting: the match statistics dataset. However, I encountered an issue with the dataset—I hadn't scaled my features. For example, I had mixed units like seconds and centimeters, which caused confusion for the SVM. Bas pointed this out, and I plan to address it by scaling the features appropriately.

Overall, the session was incredibly helpful and insightful. Bas's thorough explanations not only resolved my immediate issues but also deepened my understanding of the underlying mechanics of the models.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, so long as the R^2 is positive, you can write a story about the results. Surely the tone of the story is more likely to be in high spirits if the R^2 is higher, but also a rather low R^2 can produce a perfectly fine story to tell.
SVM are sensitive (in a negative way) to data that is in different scales, so scaling is essential. Most likely the Logistic Regression does not have this issue and you can try to train that without scaling if you'd prefer. As a matter of fact, generally speaking I would always suggest scaling to standard deviations for all training data except booleans, but it has some minor influence on plots that you make later.
Please also keep in mind LO5 and LO6.

Bas gave me a valuable analogy that stuck with me: "You can modify a

coffee machine as much as you want, but if the coffee itself is bad, the result will always be bad." This lesson reinforced the idea that a high-quality dataset is the foundation of successful modeling. I've internalized this insight and will prioritize ensuring my datasets are robust before moving forward with modeling.
Self assessment: Beginning

4.4 Final evaluation

Text Self assesment: Proficient

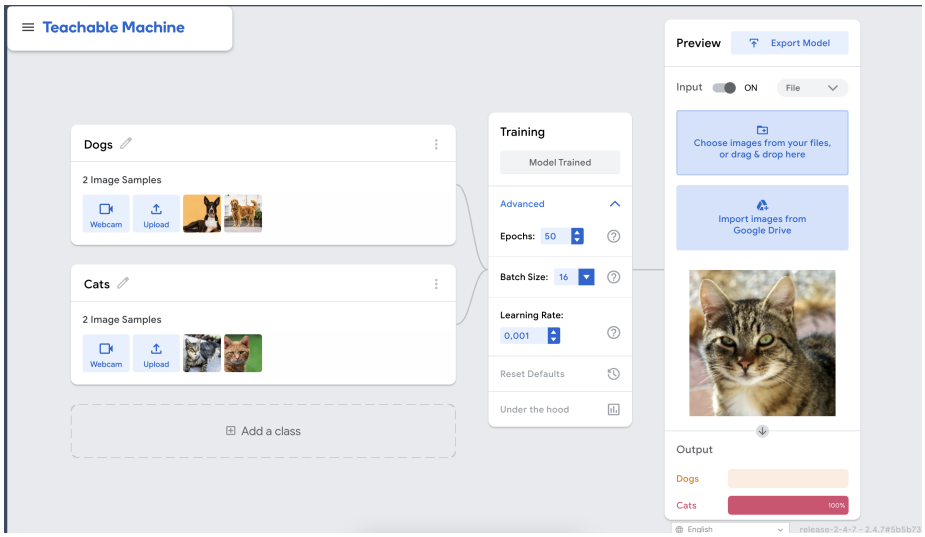
5 Learning outcome 4 - Machine teaching

Description
The student is able to use data to train models in a way that fits the intended purpose. The student is also able to test whether the models have been adequately trained.

Explanation:
After making sure that the dataset that will be used to teach a machine is not corrupted, outdated, etc., I need to ensure that the machine is doing as intended with the dataset. In order to ensure quality, testing the machine teaching is ideal.

5.1 First evaluation

In the first weeks of the minor I explored a website called 'Teachable Machine' in order to get familiar with the steps of creating an AI model. The usage of my model was to recognize cats and dogs.Other than using the website to get familiar with machine learning, I have tried to familiarize myself with different machine learning algorithms, such as Decision Trees, Random Forest, etc. Here you can see the received feedback from one of my lectures:



Self assessment: Orientating

5.2 Second evaluation

I managed to model the data of my personal project, whereas in my group project I mainly analyzed and cleaned the data. The difficulties I found along the way were mainly choosing the model that was actually fitting the purpose of my project. I had to choose between a 'Classifier' and a 'Regressor' first. For my personal project, I chose to use the 'Random Forest Classifier' model, because I was mainly trying to have a prediction of yes or no. However, after my meeting with my semester coach Bas I decided to redirect my focus on actually estimating the likelihood of each fighter winning a match against each other. The upcoming weeks I will be trying to teach a model with the data I have analyzed in my project.

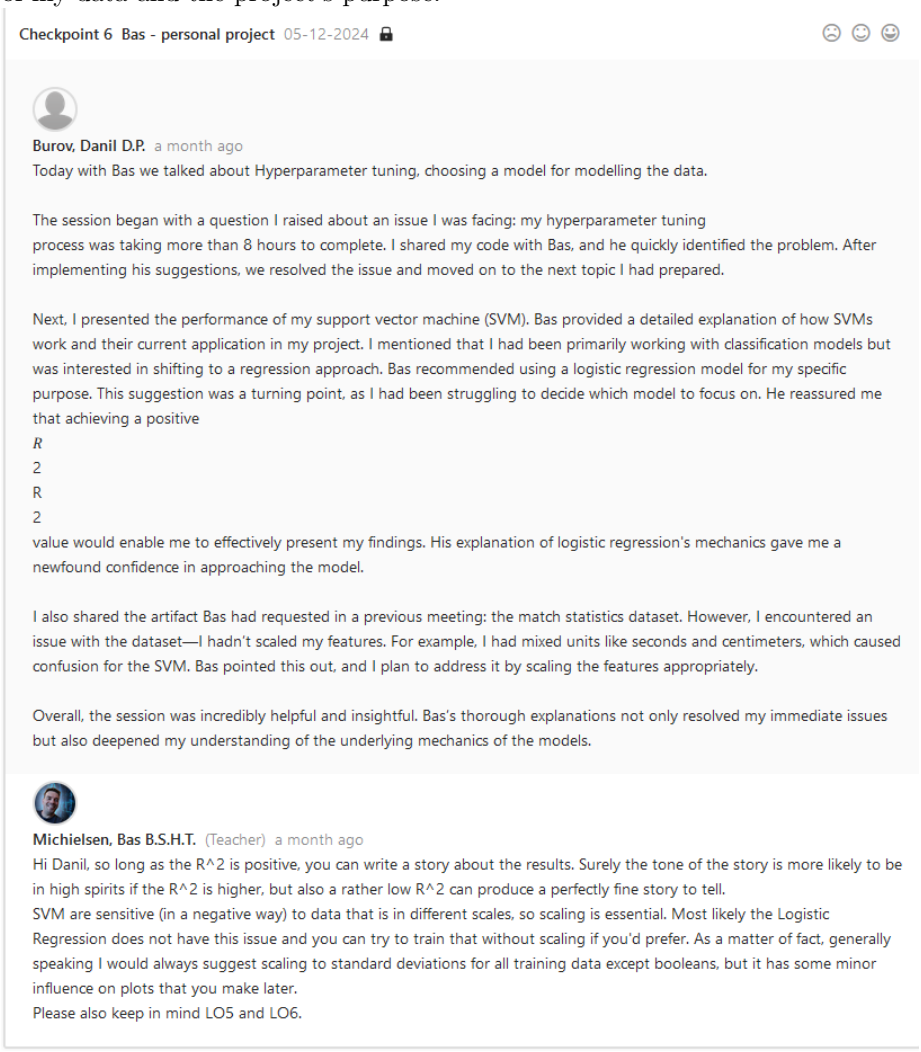
Self assessment: Beginning

5.3 Third evaluation

Machine teaching remains one of the learning outcomes where I feel I still have significant room for improvement. Personally, I find it challenging to teach a machine using a model that I do not fully understand. Fortunately, my semester coach, Bas, has been instrumental in helping me deepen my understanding of why certain models are better suited to specific situations.

In my last feedback session with Bas, I implemented a Support Vector Machine (SVM) for my personal project. However, Bas quickly pointed out that this model would not perform well given the dataset and the objectives of my project. He explained the underlying mechanics of SVMs and why they were not an optimal choice in this scenario. After our discussion, we concluded that

Logistic Regression would be a more suitable option. This decision was based on how Logistic Regression operates, which aligns better with the characteristics of my data and the project's purpose.



Checkpoint 6 Bas - personal project 05-12-2024

Burov, Danil D.P. a month ago

Today with Bas we talked about Hyperparameter tuning, choosing a model for modelling the data.

The session began with a question I raised about an issue I was facing: my hyperparameter tuning process was taking more than 8 hours to complete. I shared my code with Bas, and he quickly identified the problem. After implementing his suggestions, we resolved the issue and moved on to the next topic I had prepared.

Next, I presented the performance of my support vector machine (SVM). Bas provided a detailed explanation of how SVMs work and their current application in my project. I mentioned that I had been primarily working with classification models but was interested in shifting to a regression approach. Bas recommended using a logistic regression model for my specific purpose. This suggestion was a turning point, as I had been struggling to decide which model to focus on. He reassured me that achieving a positive

R^2

R
2

value would enable me to effectively present my findings. His explanation of logistic regression's mechanics gave me a newfound confidence in approaching the model.

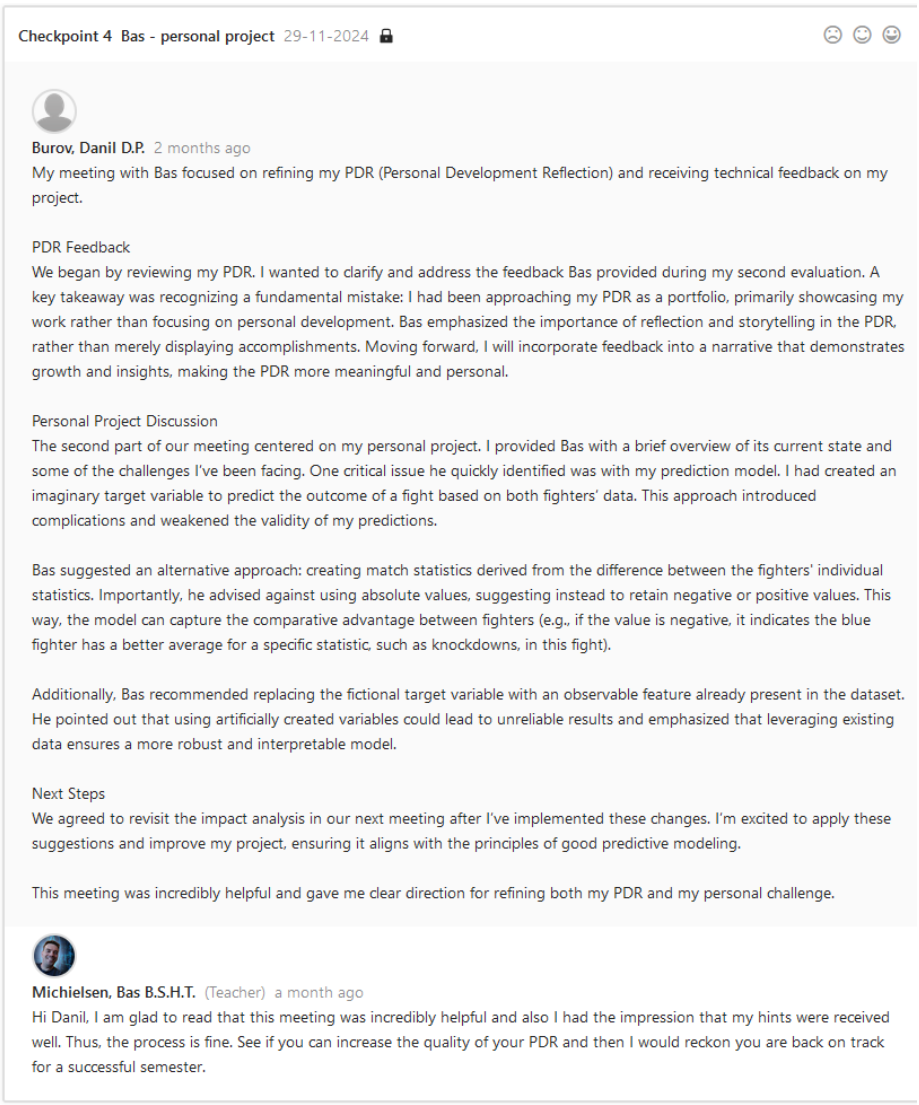
I also shared the artifact Bas had requested in a previous meeting: the match statistics dataset. However, I encountered an issue with the dataset—I hadn't scaled my features. For example, I had mixed units like seconds and centimeters, which caused confusion for the SVM. Bas pointed this out, and I plan to address it by scaling the features appropriately.

Overall, the session was incredibly helpful and insightful. Bas's thorough explanations not only resolved my immediate issues but also deepened my understanding of the underlying mechanics of the models.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, so long as the R^2 is positive, you can write a story about the results. Surely the tone of the story is more likely to be in high spirits if the R^2 is higher, but also a rather low R^2 can produce a perfectly fine story to tell.
SVM are sensitive (in a negative way) to data that is in different scales, so scaling is essential. Most likely the Logistic Regression does not have this issue and you can try to train that without scaling if you'd prefer. As a matter of fact, generally speaking I would always suggest scaling to standard deviations for all training data except booleans, but it has some minor influence on plots that you make later.
Please also keep in mind LO5 and LO6.

At the start of my project, my goal was to use a classification model to output binary results, such as 0 or 1, with no values in between. My initial idea was to predict which fighter would win a match. However, after multiple discussions with teammates and Bas, I realized that regression modeling was a better fit for my case. Regression models differ in that they don't provide an accuracy score; instead, the output itself represents the model's prediction accuracy.



Checkpoint 4 Bas - personal project 29-11-2024

Burov, Danil D.P. 2 months ago

My meeting with Bas focused on refining my PDR (Personal Development Reflection) and receiving technical feedback on my project.

PDR Feedback

We began by reviewing my PDR. I wanted to clarify and address the feedback Bas provided during my second evaluation. A key takeaway was recognizing a fundamental mistake: I had been approaching my PDR as a portfolio, primarily showcasing my work rather than focusing on personal development. Bas emphasized the importance of reflection and storytelling in the PDR, rather than merely displaying accomplishments. Moving forward, I will incorporate feedback into a narrative that demonstrates growth and insights, making the PDR more meaningful and personal.

Personal Project Discussion

The second part of our meeting centered on my personal project. I provided Bas with a brief overview of its current state and some of the challenges I've been facing. One critical issue he quickly identified was with my prediction model. I had created an imaginary target variable to predict the outcome of a fight based on both fighters' data. This approach introduced complications and weakened the validity of my predictions.

Bas suggested an alternative approach: creating match statistics derived from the difference between the fighters' individual statistics. Importantly, he advised against using absolute values, suggesting instead to retain negative or positive values. This way, the model can capture the comparative advantage between fighters (e.g., if the value is negative, it indicates the blue fighter has a better average for a specific statistic, such as knockdowns, in this fight).

Additionally, Bas recommended replacing the fictional target variable with an observable feature already present in the dataset. He pointed out that using artificially created variables could lead to unreliable results and emphasized that leveraging existing data ensures a more robust and interpretable model.

Next Steps

We agreed to revisit the impact analysis in our next meeting after I've implemented these changes. I'm excited to apply these suggestions and improve my project, ensuring it aligns with the principles of good predictive modeling.

This meeting was incredibly helpful and gave me clear direction for refining both my PDR and my personal challenge.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, I am glad to read that this meeting was incredibly helpful and also I had the impression that my hints were received well. Thus, the process is fine. See if you can increase the quality of your PDR and then I would reckon you are back on track for a successful semester.

This shift in approach was particularly enlightening. It is almost impossible to collect data and train a model to accurately predict fight outcomes because human factors are inherently unpredictable. With regression, the goal is not to predict a binary result (win or lose) but rather to produce a continuous range of values that capture all possibilities in between. This approach aligns perfectly with my use case and has given me a deeper appreciation for the nuanced applications of machine teaching.

Self assessment: Beginning

5.4 Final evaluation

Self assesment: Proficient

6 Learning outcome 5 - Data visualization

Description
The student is able to use data to create an interesting, informative and compelling story in an (interactive) data visualization product, tailored to the right target group.
Explanation:
Visualizing data will help further understand the relation between different features. In order to achieve this goal, every correlation found between different features needs to be visualized. These images need to be understandable and self-explanatory.

6.1 First evaluation

I have not yet done anything related to data visualization. I have only educated myself with the provided Python tutorials and self-learning.
Self assessment: Orientating

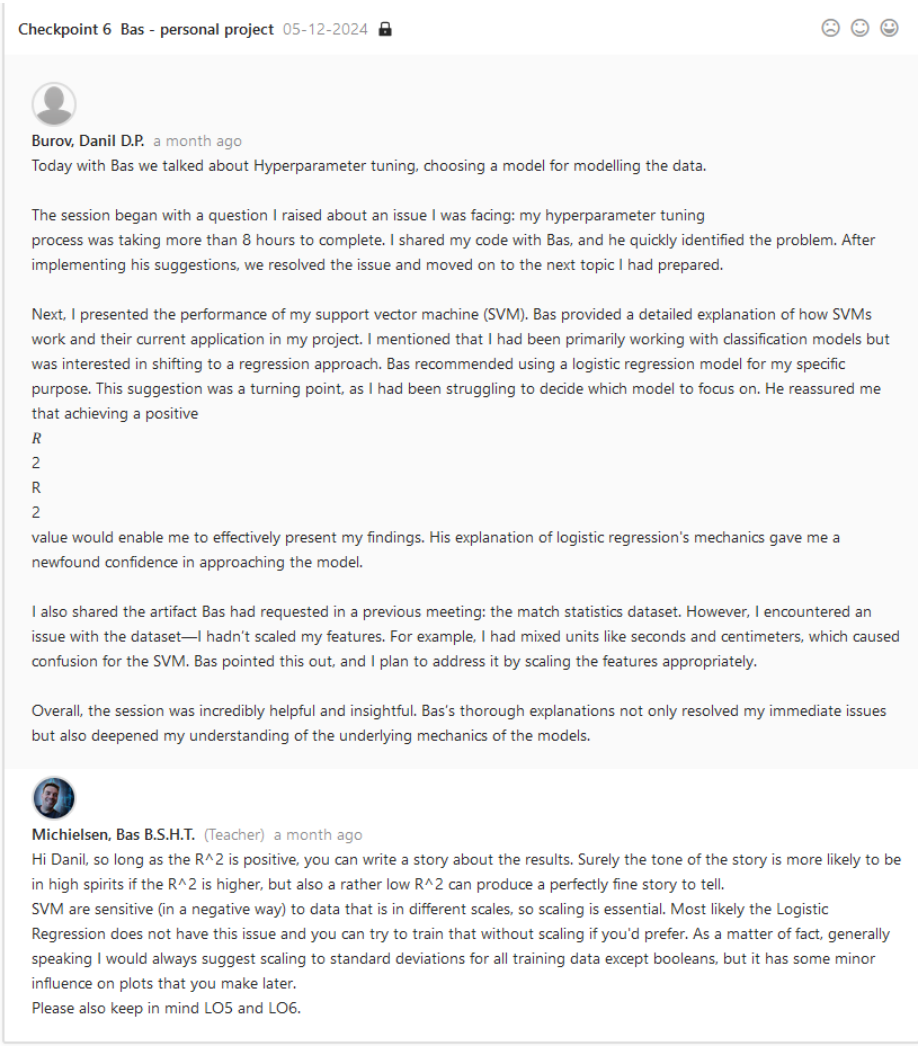
6.2 Second evaluation

I have been able to visualize the correlations with the features for my project. Unfortunately, the correlations are not that strong and I fear that the data is very unaccurate and that is making the data show this information. In the future I would like to find correlations for my personal project, answering the question which fighter statistics make the percentage of a fighters' chance of winning higher.

Self assessment: Beginning

6.3 Third evaluation

Unfortunately, I have not received specific feedback on my data visualizations. However, I did create several plots for both my personal project and my group project. Through this process, I began to appreciate the value of visualizations in gaining deeper insights into data. While Bas and I briefly reviewed my visualizations, we didn't delve deeply into them. Based on the feedback I received for data preparation, I recognize that my visualizations can be significantly improved.



Checkpoint 6 Bas - personal project 05-12-2024

Burov, Danil D.P. a month ago

Today with Bas we talked about Hyperparameter tuning, choosing a model for modelling the data.

The session began with a question I raised about an issue I was facing: my hyperparameter tuning process was taking more than 8 hours to complete. I shared my code with Bas, and he quickly identified the problem. After implementing his suggestions, we resolved the issue and moved on to the next topic I had prepared.

Next, I presented the performance of my support vector machine (SVM). Bas provided a detailed explanation of how SVMs work and their current application in my project. I mentioned that I had been primarily working with classification models but was interested in shifting to a regression approach. Bas recommended using a logistic regression model for my specific purpose. This suggestion was a turning point, as I had been struggling to decide which model to focus on. He reassured me that achieving a positive

R^2

value would enable me to effectively present my findings. His explanation of logistic regression's mechanics gave me a newfound confidence in approaching the model.

I also shared the artifact Bas had requested in a previous meeting: the match statistics dataset. However, I encountered an issue with the dataset—I hadn't scaled my features. For example, I had mixed units like seconds and centimeters, which caused confusion for the SVM. Bas pointed this out, and I plan to address it by scaling the features appropriately.

Overall, the session was incredibly helpful and insightful. Bas's thorough explanations not only resolved my immediate issues but also deepened my understanding of the underlying mechanics of the models.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, so long as the R^2 is positive, you can write a story about the results. Surely the tone of the story is more likely to be in high spirits if the R^2 is higher, but also a rather low R^2 can produce a perfectly fine story to tell.
SVM are sensitive (in a negative way) to data that is in different scales, so scaling is essential. Most likely the Logistic Regression does not have this issue and you can try to train that without scaling if you'd prefer. As a matter of fact, generally speaking I would always suggest scaling to standard deviations for all training data except booleans, but it has some minor influence on plots that you make later.
Please also keep in mind LO5 and LO6.

One of the key issues is that I did not apply standard deviation to my dataset, which led to inaccuracies in my plots. This oversight highlights the importance of ensuring data is properly prepared before creating visualizations, as any errors in the dataset will inevitably affect the clarity and accuracy of the resulting visuals.

Looking ahead, I want to engage in discussions with Bas and other technical consultants to better understand which types of plots would best suit my personal project. I also aim to learn how to interpret these visualizations effectively to maximize their value. By refining this skill, I hope to leverage data visualization as a powerful tool for both analysis and communication

Self assessment: Beginning

6.4 Final evaluation

Self assesment: Proficient

7 Learning outcome 6 - Reporting

Description
The student is able to report in a methodologically sound manner on (the outcome of) their own AI projects (project proposal, process documentation, reporting of final results, etc.).
Explanation:
It is important that documents are written in time and feedback is used to prove the legitimacy of the given goal. The goal would be considered accomplished if all documents are consistent and comprehensible. It is very important that the code documentation is easy to understand and use.

7.1 First evaluation

It is essential for me to document the progress I make in order to keep track of how I have improved during the minor. To do so, I have created my own personal repository where I have already put all the documents I have written up until now, as well as all the exercises I have completed over the past few weeks.

Self assessment: Beginning

7.2 Second evaluation

I have added markdown 'README' file in my repository to make everything more structured and to be easily found. I am trying to work as transparently as possible and document the work I do. In order to keep it transparent I write markdown in the python notebooks,ensuring that everything I do is understandable and comprehensive.

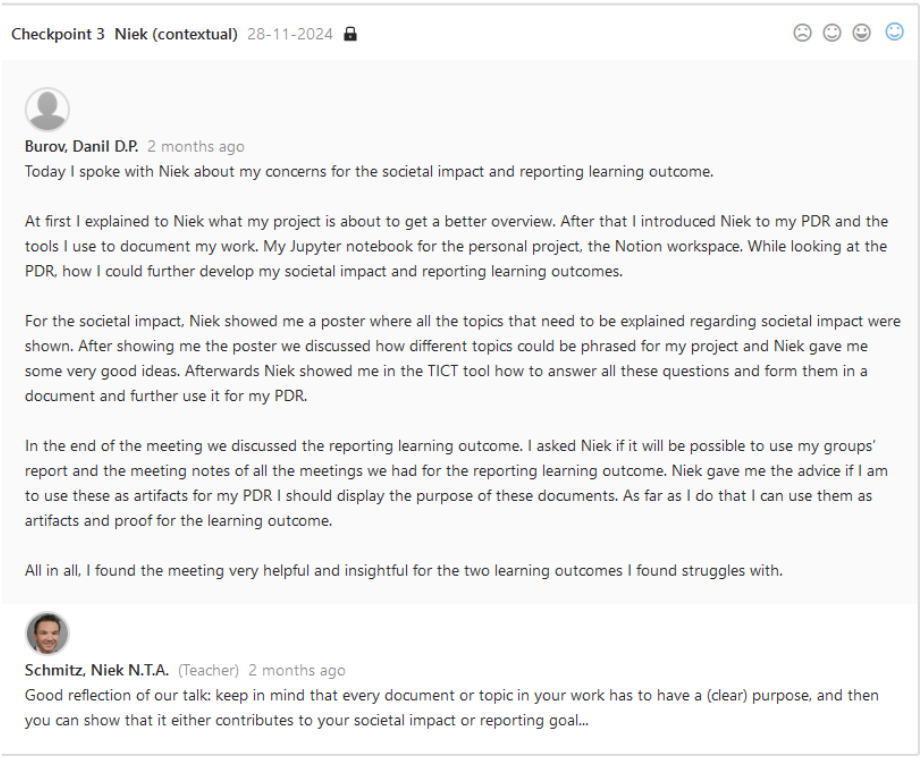
Link to my personal repository: Personal repository for the minor
Link to my personal project notebook: Personal project notebook
Link to my group project notebook: Group project notebook

Self assessment: Proficient

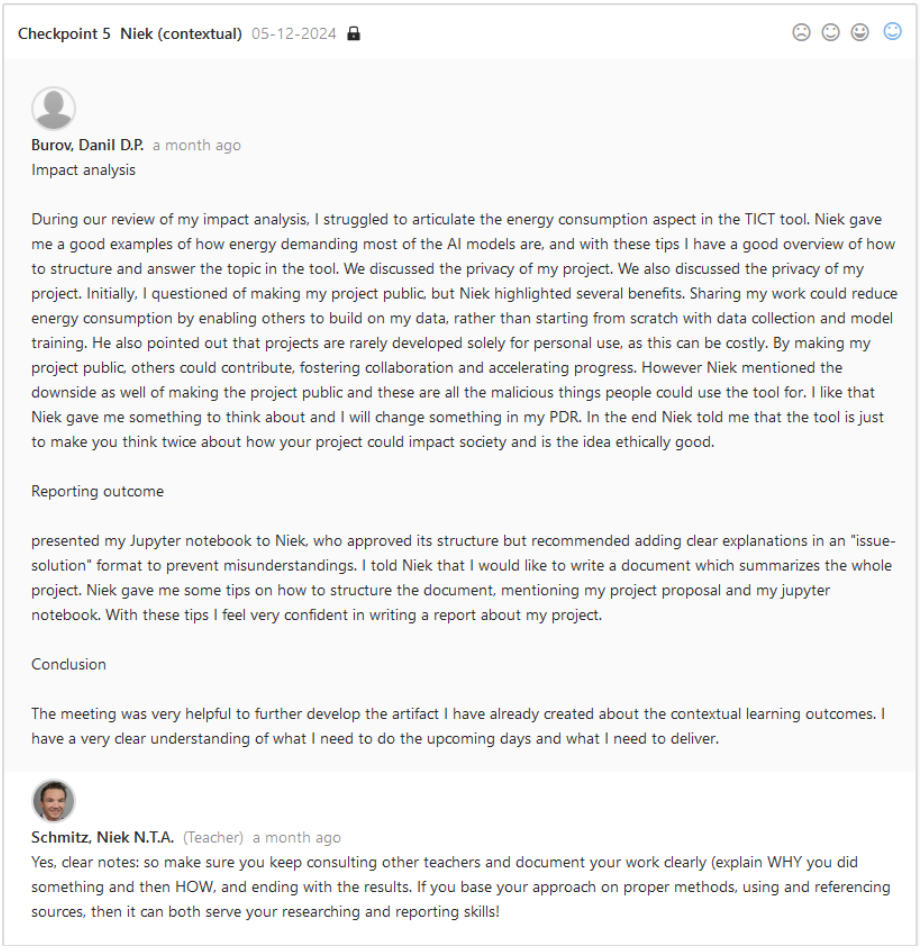
7.3 Third evaluation

Since the beginning of the semester, I have been focusing on documenting my work to improve my reporting skills. To support this effort, I have experimented with different tools to record the progress and results of my work. Seeking a professional perspective on how to align these efforts with my personal goals,

I consulted with Niek. He advised me that as long as a document has a clear purpose, it can effectively showcase my learning process.



This advice made me reconsider my approach. Instead of using all the documentation material I had created separately, I decided to compile a well-rounded document that integrates all my work throughout the minor. When I shared this idea with Niek, he supported it and provided additional guidance. He emphasized that a good report doesn't have to be lengthy; a concise, well-structured document can often be more impactful. This insight surprised me but left a lasting impression.



Checkpoint 5 Niek (contextual) 05-12-2024

Burov, Danil D.P. a month ago

Impact analysis

During our review of my impact analysis, I struggled to articulate the energy consumption aspect in the TICT tool. Niek gave me a good examples of how energy demanding most of the AI models are, and with these tips I have a good overview of how to structure and answer the topic in the tool. We discussed the privacy of my project. We also discussed the privacy of my project. Initially, I questioned of making my project public, but Niek highlighted several benefits. Sharing my work could reduce energy consumption by enabling others to build on my data, rather than starting from scratch with data collection and model training. He also pointed out that projects are rarely developed solely for personal use, as this can be costly. By making my project public, others could contribute, fostering collaboration and accelerating progress. However Niek mentioned the downside as well of making the project public and these are all the malicious things people could use the tool for. I like that Niek gave me something to think about and I will change something in my PDR. In the end Niek told me that the tool is just to make you think twice about how your project could impact society and is the idea ethically good.

Reporting outcome

presented my Jupyter notebook to Niek, who approved its structure but recommended adding clear explanations in an "issue-solution" format to prevent misunderstandings. I told Niek that I would like to write a document which summarizes the whole project. Niek gave me some tips on how to structure the document, mentioning my project proposal and my jupyter notebook. With these tips I feel very confident in writing a report about my project.

Conclusion

The meeting was very helpful to further develop the artifact I have already created about the contextual learning outcomes. I have a very clear understanding of what I need to do the upcoming days and what I need to deliver.

Schmitz, Niek N.T.A. (Teacher) a month ago

Yes, clear notes: so make sure you keep consulting other teachers and document your work clearly (explain WHY you did something and then HOW, and ending with the results. If you base your approach on proper methods, using and referencing sources, then it can both serve your researching and reporting skills!

My plan now is to consolidate all my project documentation into a clear and structured report. Even if the final outcome of the project isn't entirely positive, I want to create a document that effectively communicates the work I have done and is easy for others to read and understand.
Self assessment: Beginning

7.4 Final evaluation

Self assesment: Proficient

8 Learning outcome 7 - Personal Leadership

Description
The student shows an entrepreneurial mindset regarding their own AI project(s)

and personal development, while being aware of their own learning capacity and keeping in mind professional ambitions in their future work field.
Explanation:
The outcome of this learning goal should be that I manage my time correctly without overloading myself. That means following some kind of a schedule, as well as attending lectures to ensure that I stay on track and seeking feedback to make sure I am progressing.

8.1 First evaluation

Ever since the minor started, I have attended all lectures regarding the minor and have taken some insights that will help me further develop my skills in AI modeling. I still need to improve on asking for more feedback from lecturers.
Self assessment: Orientating

8.2 Second evaluation

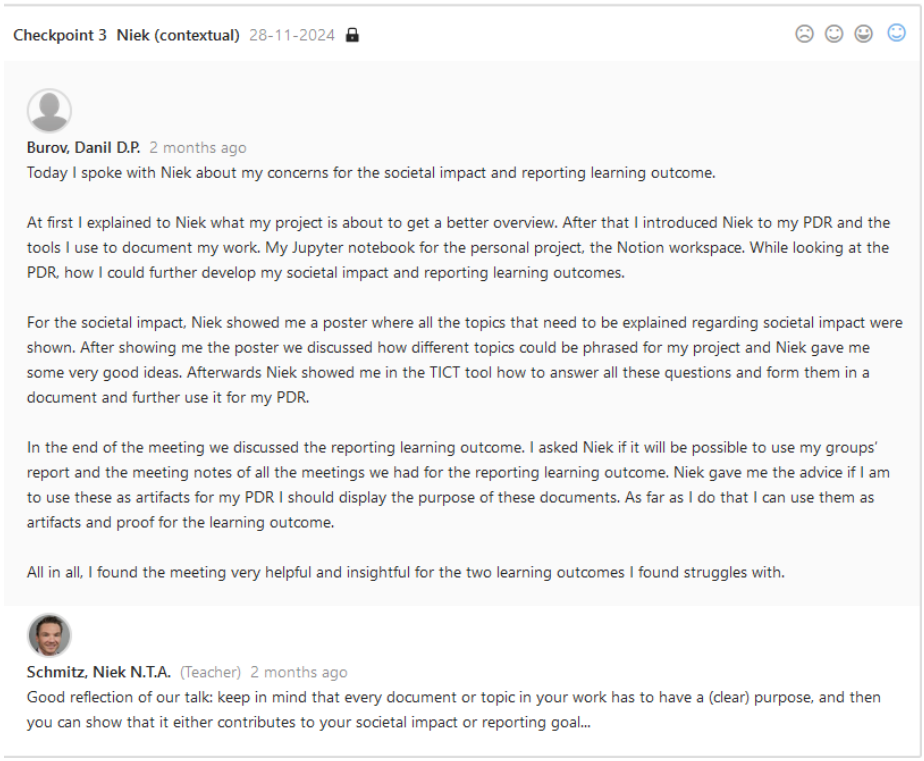
I have been following a sort of strict schedule when it comes down to work. I am working from the morning to the late afternoon every Tuesday and Thursday with my group. That way I ensure that I am always up-to-date with what my teammates are doing in the project. Moreover, I do my personal project work on Monday and Wednesday, leaving Friday to be a day which is free. This was done intentionally, making sure that if I feel that something needs more work I can do it then.

Self assessment: Proficient

8.3 Third evaluation

Since the last evaluation, I have adjusted my approach to personal leadership. I now prioritize incorporating feedback into my workflow, organizing my week around regular feedback sessions. For example, I schedule a feedback session every Thursday, ensuring that I deliver the results from the previous session by that day.

This approach has proven to be highly efficient. It keeps me focused, encourages consistent progress, and helps me accomplish more within a structured timeframe. Additionally, working closely with my tutors allows me to keep them updated on my progress, fostering better collaboration and alignment with my goals.



Checkpoint 3 Niek (contextual) 28-11-2024

Burov, Danil D.P. 2 months ago

Today I spoke with Niek about my concerns for the societal impact and reporting learning outcome.

At first I explained to Niek what my project is about to get a better overview. After that I introduced Niek to my PDR and the tools I use to document my work. My Jupyter notebook for the personal project, the Notion workspace. While looking at the PDR, how I could further develop my societal impact and reporting learning outcomes.

For the societal impact, Niek showed me a poster where all the topics that need to be explained regarding societal impact were shown. After showing me the poster we discussed how different topics could be phrased for my project and Niek gave me some very good ideas. Afterwards Niek showed me in the TICT tool how to answer all these questions and form them in a document and further use it for my PDR.

In the end of the meeting we discussed the reporting learning outcome. I asked Niek if it will be possible to use my groups' report and the meeting notes of all the meetings we had for the reporting learning outcome. Niek gave me the advice if I am to use these as artifacts for my PDR I should display the purpose of these documents. As far as I do that I can use them as artifacts and proof for the learning outcome.

All in all, I found the meeting very helpful and insightful for the two learning outcomes I found struggles with.

Schmitz, Niek N.T.A. (Teacher) 2 months ago

Good reflection of our talk: keep in mind that every document or topic in your work has to have a (clear) purpose, and then you can show that it either contributes to your societal impact or reporting goal...

Checkpoint 5 Niek (contextual) 05-12-2024

Burov, Danil D.P. a month ago
Impact analysis

During our review of my impact analysis, I struggled to articulate the energy consumption aspect in the TICT tool. Niek gave me a good examples of how energy demanding most of the AI models are, and with these tips I have a good overview of how to structure and answer the topic in the tool. We discussed the privacy of my project. We also discussed the privacy of my project. Initially, I questioned of making my project public, but Niek highlighted several benefits. Sharing my work could reduce energy consumption by enabling others to build on my data, rather than starting from scratch with data collection and model training. He also pointed out that projects are rarely developed solely for personal use, as this can be costly. By making my project public, others could contribute, fostering collaboration and accelerating progress. However Niek mentioned the downside as well of making the project public and these are all the malicious things people could use the tool for. I like that Niek gave me something to think about and I will change something in my PDR. In the end Niek told me that the tool is just to make you think twice about how your project could impact society and is the idea ethically good.

Reporting outcome

presented my Jupyter notebook to Niek, who approved its structure but recommended adding clear explanations in an "issue-solution" format to prevent misunderstandings. I told Niek that I would like to write a document which summarizes the whole project. Niek gave me some tips on how to structure the document, mentioning my project proposal and my jupyter notebook. With these tips I feel very confident in writing a report about my project.

Conclusion

The meeting was very helpful to further develop the artifact I have already created about the contextual learning outcomes. I have a very clear understanding of what I need to do the upcoming days and what I need to deliver.

Schmitz, Niek N.T.A. (Teacher) a month ago
Yes, clear notes: so make sure you keep consulting other teachers and document your work clearly (explain WHY you did something and then HOW, and ending with the results. If you base your approach on proper methods, using and referencing sources, then it can both serve your researching and reporting skills!

Checkpoint 4 Bas - personal project 29-11-2024

Burov, Danil D.P. 2 months ago

My meeting with Bas focused on refining my PDR (Personal Development Reflection) and receiving technical feedback on my project.

PDR Feedback

We began by reviewing my PDR. I wanted to clarify and address the feedback Bas provided during my second evaluation. A key takeaway was recognizing a fundamental mistake: I had been approaching my PDR as a portfolio, primarily showcasing my work rather than focusing on personal development. Bas emphasized the importance of reflection and storytelling in the PDR, rather than merely displaying accomplishments. Moving forward, I will incorporate feedback into a narrative that demonstrates growth and insights, making the PDR more meaningful and personal.

Personal Project Discussion

The second part of our meeting centered on my personal project. I provided Bas with a brief overview of its current state and some of the challenges I've been facing. One critical issue he quickly identified was with my prediction model. I had created an imaginary target variable to predict the outcome of a fight based on both fighters' data. This approach introduced complications and weakened the validity of my predictions.

Bas suggested an alternative approach: creating match statistics derived from the difference between the fighters' individual statistics. Importantly, he advised against using absolute values, suggesting instead to retain negative or positive values. This way, the model can capture the comparative advantage between fighters (e.g., if the value is negative, it indicates the blue fighter has a better average for a specific statistic, such as knockdowns, in this fight).

Additionally, Bas recommended replacing the fictional target variable with an observable feature already present in the dataset. He pointed out that using artificially created variables could lead to unreliable results and emphasized that leveraging existing data ensures a more robust and interpretable model.

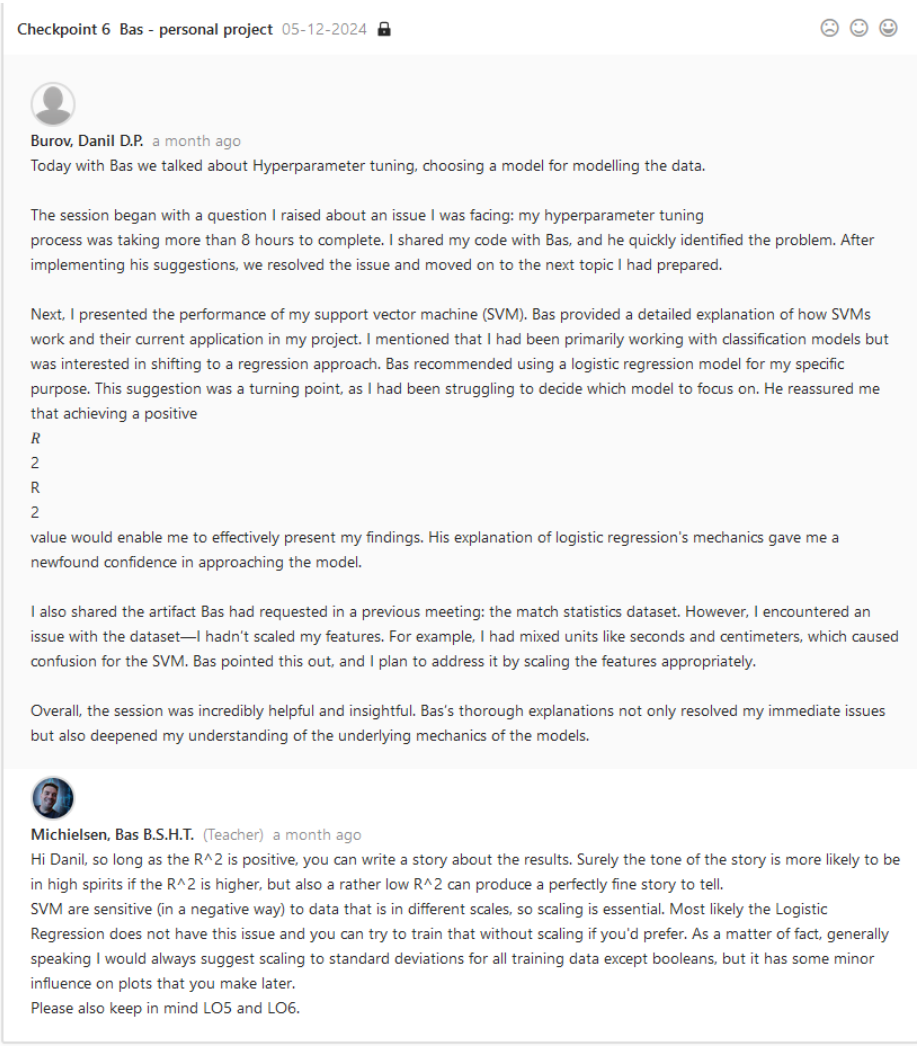
Next Steps

We agreed to revisit the impact analysis in our next meeting after I've implemented these changes. I'm excited to apply these suggestions and improve my project, ensuring it aligns with the principles of good predictive modeling.

This meeting was incredibly helpful and gave me clear direction for refining both my PDR and my personal challenge.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, I am glad to read that this meeting was incredibly helpful and also I had the impression that my hints were received well. Thus, the process is fine. See if you can increase the quality of your PDR and then I would reckon you are back on track for a successful semester.



Self assessment: Beginning

8.4 Final evaluation

Self assesment: Proficient

9 Learning outcome 8 - Personal goal

Description

With this learning outcome, the student can set their own goal in relation to their future field of work. This is always related to your Individual Challenge. Describe this Learning Outcome in your PDR.

Explanation:
My personal goal for this minor is to become very familiar with how to create an AI model, be able to analyze large amounts of data, and prepare this data for further training of a given model. I would like to get more familiar with how machine learning algorithms work and be able to choose the appropriate algorithm for the use case.

9.1 First evaluation

In the first 4 weeks, I managed to choose a dataset for my personal project that fits my needs. I attended all lectures to make sure I progress in my knowledge of AI, both technically and ethically. I managed to import the dataset and clean it based on a condition.

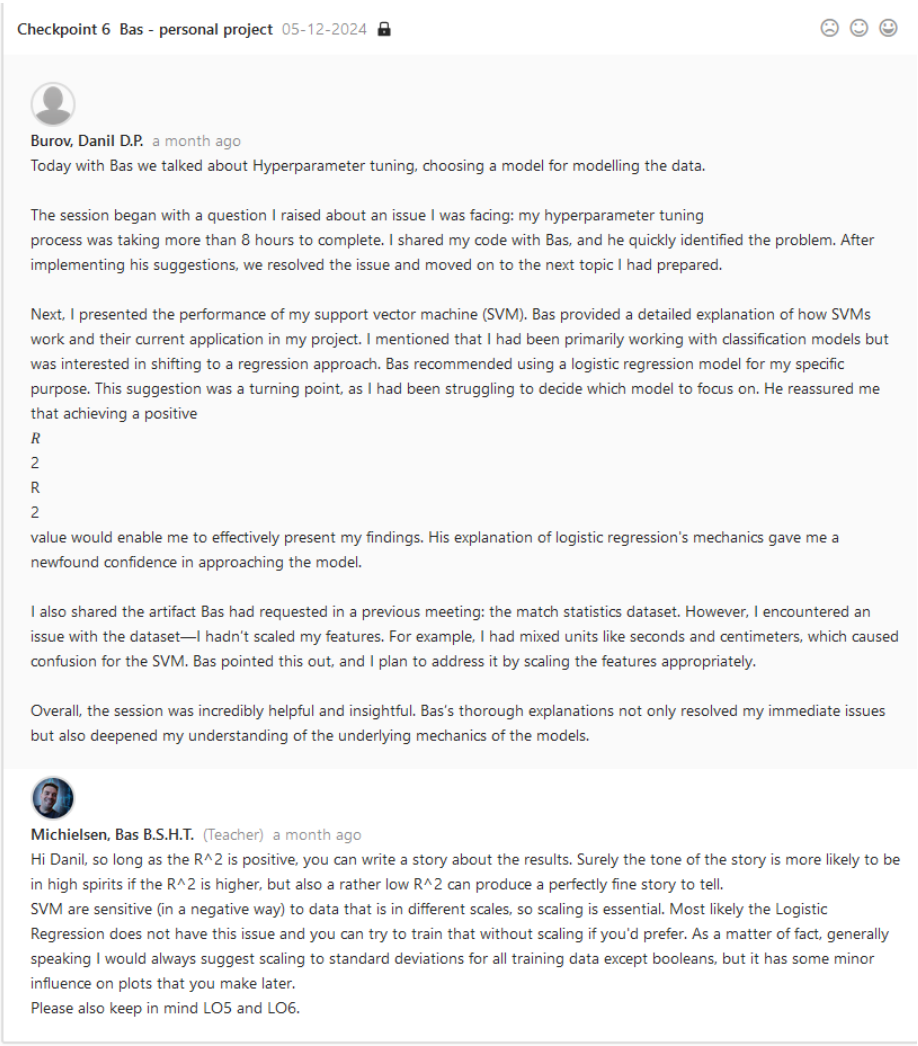
Self assessment: Orientating

9.2 Second evaluation

From week six to nine I believe that I have progressed quite a lot since the previous four weeks. I have done a lot of data preparation for modelling and even managed to train my first model. I am very content with the work I have done regarding my personal project, tackling legal issues as well battling accuracy in the moment. I know that the current dataset may not be enough to train a model sufficiently as well, so I am very excited for the upcoming weeks and how both my projects will progress. I am currently getting familiar with different machine learning algorithms since I found not much success in my personal projects' model I am currently using.

9.3 Third evaluation

Regarding my personal goal, I feel confident in saying that I have made significant progress in approaching the creation of AI models, managing large datasets, and selecting the appropriate model for the data's purpose.



Checkpoint 6 Bas - personal project 05-12-2024

Burov, Danil D.P. a month ago

Today with Bas we talked about Hyperparameter tuning, choosing a model for modelling the data.

The session began with a question I raised about an issue I was facing: my hyperparameter tuning process was taking more than 8 hours to complete. I shared my code with Bas, and he quickly identified the problem. After implementing his suggestions, we resolved the issue and moved on to the next topic I had prepared.

Next, I presented the performance of my support vector machine (SVM). Bas provided a detailed explanation of how SVMs work and their current application in my project. I mentioned that I had been primarily working with classification models but was interested in shifting to a regression approach. Bas recommended using a logistic regression model for my specific purpose. This suggestion was a turning point, as I had been struggling to decide which model to focus on. He reassured me that achieving a positive

R^2

value would enable me to effectively present my findings. His explanation of logistic regression's mechanics gave me a newfound confidence in approaching the model.

I also shared the artifact Bas had requested in a previous meeting: the match statistics dataset. However, I encountered an issue with the dataset—I hadn't scaled my features. For example, I had mixed units like seconds and centimeters, which caused confusion for the SVM. Bas pointed this out, and I plan to address it by scaling the features appropriately.

Overall, the session was incredibly helpful and insightful. Bas's thorough explanations not only resolved my immediate issues but also deepened my understanding of the underlying mechanics of the models.

Michielsen, Bas B.S.H.T. (Teacher) a month ago

Hi Danil, so long as the R^2 is positive, you can write a story about the results. Surely the tone of the story is more likely to be in high spirits if the R^2 is higher, but also a rather low R^2 can produce a perfectly fine story to tell.
SVM are sensitive (in a negative way) to data that is in different scales, so scaling is essential. Most likely the Logistic Regression does not have this issue and you can try to train that without scaling if you'd prefer. As a matter of fact, generally speaking I would always suggest scaling to standard deviations for all training data except booleans, but it has some minor influence on plots that you make later.
Please also keep in mind LO5 and LO6.

Reflecting on the feedback I have received, I realize how much my knowledge has expanded throughout this process. Engaging in discussions about various models with Bas, presenting my cleaned dataset, and receiving constructive feedback have all contributed to my growth. These moments not only highlight my learning but also reinforce the progress I have made toward achieving my goals.

Self assessment: Beginning

9.4 Final evaluation

Self assesment: Proficient

10 Retrospect

11 Conclusion