Changelog Marc Vogtländer

Abstract, Problem Statement & Research Goal

- **The abstract is hard to read without prior knowledge and should not include references to section numbers. Instead it should be a brief summary of the thesis that can be read without prior knowledge.**
  - Removed reference section numbers .
  - Revised the abstract version and rewrote vague aspects and underpinned findings with evaluation scores.
  - Technical terms explained later in the thesis not mentioned earlier in the thesis. These were replaced with more general words.
- **The societal relevance of the thesis is adequately discussed, the scientific relevance receives much less attention, at least in the Introduction section of the thesis.**
  - Added a subsection with societal and scientific relevance
  - Implemented a research paper proposing to conduct further research on amazon reviews for generalising models on pre-processing techniques
- **RSQ3 should be rephrased given the later binary distinction between "good" and "bad" reviews.**
  - Old RSQ3:
    - To what extent is the impact of sentiment analysis on the predictive power of machine learning models consistent across the review score range?
  - New RSQ3:
    - To what extent is the impact of sentiment analysis on the predictive power of machine learning models consistent across the **binary** positive and negative **classification**?
  - Later in the thesis (experimental setup), 'positive' and 'negative' classification are further substantiated
- **Research subquestions address relevant data science issues, such as class imbalance, model performance, and disparate impact analysis. The formulation of the research question and subquestions is vague: what is predicted from what, review scores from the review text? Or from the description of the product? Which models are compared against each other?**
  - Old RSQ1:
    - Which sentiment analysis techniques are best suited for predicting Amazon product reviews and how do they improve their performance?
  - New RSQ1:
    - Which sentiment analysis techniques are best suited for **classifying** Amazon **text** reviews and how do they improve their performance?
  - Old RSQ2:
    - How does the performance of different machine learning algorithms compare for Amazons reviews?
  - New RSQ2:
    - How does the performance of Naïve Bayes, Support Vector Machines, Gradient Descent, XGBoost, KNN and Decision Tree machine learning algorithms compare for Amazon's reviews?
      - Considered that the RSQ2 is pretty long mentioning all the models but makes it more detailed and therefore implemented
  - RSQ3 (See previous bullet point)

    - RSQ4 Unchanged
- **The contribution of the thesis is not clear, considering the vast literature of sentiment analysis and the fact that the models used in the thesis are classic algorithms and the dataset is popular.**
    - Added literature about SMOTE
    - Removed irrelevant literature
    - Revised the relevance sections
    - Revised section 'Additional value of the thesis'
    - **Literature Review: The structure of the literature review is not very clear. As a reader, I often had to think for a bit how sections related to each other.**
        - Revised literature review, making literature in line with the research sub questions
- **The tables with model performances from previous studies are too much and should be summarized in the text, with the author critically interpreting previous work and connecting it to the thesis research.**
    - Only kept the relevant tables
    - Moved the additional tables to the appendix and put important results in main text.
    - Been more critical about previous studies (e.g. authors not underpinning their theory with performance results)
- **In general, the use of Tables with results from previous studies should be avoided in the literature review. I would highly recommend structuring the literature review section more clearly, with subsections related to the research questions and subsubsections within these subsections.**
    - Implemented subsections & subsubsections throughout the whole thesis.
    - Implemented sections in header text for a more convenient overview
    - Implemented reference numbers
    - Subsections are aligned with research sub questions
- **The overview of the sentiment analysis literature is extensive, but the structure is confusing and too much (unnecessary) detail is included for each study.**
    - Removed unnecessary literature review (e.g. Sentiment analysis pipeline / rule & automated based approach)
    - Unnecessary tables moved to appendix and referenced to in text
    - Added relevant literature review (e.g. SMOTE)
- **Figure 1 seems to be a direct copy of Birjali et al., which is not permitted (unless you ask for the copyright from the authors). This might also apply to Table 1 and 2, unless you have compiled these yourself (in which case you should still specify where these numbers come from in the caption).**
    - Removed figure 1 (was part of unnecessary literature)
    - Tables compiled by myself now have the source specified (except when written by myself)
- **Methodology & Experimental Setup: An EDA is included. The decision to split the ratings into 1-4 and 5 should be motivated, for instance based on the sparsity of data for low ratings.**
    - Motivated in section 5.3
- **Figure 4 is very hard to read. Also, you don't have to refer to the source of the figures if you made them for your thesis. Vogtländer (2022), therefore is unnecessary.**
    - Removed source when compiled by myself
    - Figure 4 (now figure 6) changed to a bar plot showing only the amount of missing values

- **The procedure for resampling and splitting the dataset into train and test data is described clearly.**
- **The hyperparameter tuning process should be discussed in the Experimental Setup section.**
    - Moved to the experimental setup (section 5.9)
- **Methodology and experimental setup should be two separate sections.**
    - Split into two separate sections (section 4 methodology and section 5 experimental setup)
- **The methodology section should provide general descriptions of the methods used, whereas the experimental setup section should provide information about the dataset, the EDA, train/test data, hyperparameter tuning, evaluation metrics, et cetera.**
    - Wrote a new section containing the methodology wherein general descriptions of the methods being used are written (section 4)
    - EDA (section 5) contains the setup of the thesis (e.g. workflow, description, evaluation metrics, pre-processing techniques, software, packages etc.)
    - Added subsection to intrinsic evaluation, providing justification for using Macro vs Weighted F1 score
- **I would recommend starting the experimental setup section with the diagram showing the data science pipeline to help the reader understand the structure of this section.**
    - Implemented in section 5.1 as starting section for the experimental setup
- **In general, I found the structure of the Methodology and Experimental Setup highly confusing.**
    - Split into two sections (methodology and experimental setup)
    - Added subsections
- **Please follow the thesis guidelines and examples in previous theses carefully to help the reader better understand the setup of your thesis research.**
    - Revised DSS master thesis guidelines -vF2022.pdf
    - Looked into S2022 Good theses examples to get a better understanding of the thesis setup and layout
- **I found it strange that any rating <=4 was considered negative, and only 5-star reviews were considered positive. What is the justification for this decision?**
    - Motivated in section 5.3
- **The fine-tuning procedure for each model must be described clearly: what values were considered for each hyperparameter, and how were they set?**
    - Referred to the appendix for visualization of the code for the hyperparameter tuning
    - Added the justification for choosing hyperparameter arguments (and how they were set)
- **It sounds like there is no validation set when dividing the dataset.**
    - Added that with (Random)GridSearchCV validation sets are made based on the amount of folds.
- **Removing stopwords and lemmatizing has been shown not to be effective (or sometimes even harmful) for sentiment analysis when using deep neural networks, so the preprocessing pipeline might not be best for all models used.**
    - Just barely improved the performance (not significantly) added this in the discussion section (section 8.2) with the performance metric
- **Results: The results are presented in a comprehensive manner, but are extremely hard to read due to the abundance of tables and data.**
    - Revised the whole result section. Added comparing tables for every model and pre-processing technique with the corresponding evaluation metric in 1 table (table 3, 5 and 6), instead of multiple tables making it chaotic and unreadable.

    - For more construction the results are now in line with the research question which is proposed in the beginning of each section
- **The results of the hyperparameter tuning process should be discussed in the Experimental Setup section.**
    - Hyperparameter tables moved to appendix VIII
    - Hyperparameter tuning (section 5.9) now contains justification for choosing the hyperparameters with a reference to their results (appendix VIII)
- **Error analyses are included in the form of confusion matrices.**
    - Indeed, but moved the confusion matrices for every single model to appendix VIII for a clearer view and referenced to this appendix for the full results
    - Added tables summarizing the models, pre-processing techniques and evaluation scores
- **The figure about the loss for the train and test data in the RNN should be omitted or moved to the experimental setup section.**
    - Moved to appendix IX
    - Referenced in the experimental setup (section 5.14) to appendix VI and IX for the architecture of the deep learning models and the visualisation for overfitting
- **The results section should be limited to reporting the results themselves.**
    - Now solely reports the results instead of also including justifications for the results (previous thesis)
    - Justifications for the results are moved to the conclusion section
- **The results section is long and unstructured; many experiments on different models used in different configurations are reported but it is almost impossible to see the big picture here. Results must be presented in a compiled (and preferably visualized) way to emphasize relative performance of each model, and allow for easy comparison.**
    - See previous comments e.g. (Error analyses are included in the form of confusion matrices.)
        - Added tables summarizing the models, pre-processing techniques and evaluation scores instead of multiple tables for pre-processing techniques, models, evaluation scores
- **Different research questions must be organized into separate sections with the appropriate analyses reported.**
    - Every section now starts with the research sub question to give a more transparent overview to the results
- **The CNN model is obviously overfitting (most probably due to the simplicity of the architecture and lack of fine-tuning).**
    - Added this in the discussion section (section 7.2.3)
    - Justified why further improvement on this model wasn't made
    - Encouraged further research to improve the architecture of deep learning models for better classificationSome of the tables have major formatting issues (e.g. Tables 8 and 9).
- **Discussion & Conclusion: The discussion section provides some discussion of the results in light of the research questions and subquestions. This discussion, however, should be extended considerably, as should the discussion of the scientific and societal relevance of the thesis research.**
    - Added discussion section about societal relevance (section 7.4)
    - Added discussion section about scientific relevance (section 7.5)
    - Extended section 7.1 about SMOTE
    - Extended section 7.2 about computational power
- **Main findings and limitations of the thesis are sufficiently discussed.**

- **Form & Presentation: There are quite a few grammatical and spelling errors in the thesis. Please proofread the thesis and run it through a spelling and grammar checker.**
    - Proofread the thesis myself
    - Proofread the thesis for spelling and grammar by my girlfriend's aunt (Who's origin is from Australia)
    - Ran through a spelling and grammar checker
- **See the second reader's comment about copyright for figures. In-text references should not use brackets around author names. Please consult the APA style guide to ensure the formatting of references is correct. Overall, the writing style is not what is expected in an academic paper. Please consult previous theses to better understand what is expected in terms of form and presentation. Avoid referring to your thesis as "this report". The thesis needs proofreading and restructuring. Also see my comments on Literature review.**
    - In-text references don't include brackets around author names anymore
    - Checked S2022 theses to improve the form and presentation of the thesis
    - This report is changed into the/this thesis
        - Avoided using the/this thesis in main text (except in sections as; acknowledgements. Additional value of the thesis, and discussion)
    - Restructured the thesis based on S2022 theses
    - Implemented advices from the section 'literature review'
        - See previous sections