|  |
| --- |
| **Intent Classification for Bank Customer Queries** |
|  |
| **Yumi Heo**  230003122  MSc Data Science  [yumi.heo@city.ac.uk](mailto:yumi.heo@city.ac.uk) |

Problem statement and Motivation

Please summarise the primary goal of your project. What problem are you intending to address. What is the motivation of the problem.

General guidelines: Please follow the instructions as specified in the “Final Coursework” document provided on moodle ([link](https://moodle.city.ac.uk/pluginfile.php/2824561/course/section/588887/coursework%20details%20and%20marking%20criteria.pdf)). Carefully follow the marking guidelines. A minimum of 8-pages is required for PG programmes and 4-pages (at least) for UG-programmes. This coursework report must be more elaborate than your proposal. Please attach the code as an “.ipynb” notebook (download your code from colab as “.ipynb” notebook). Additionally provide the colab link to the notebook.

Please use this template or the LaTeX version of this template (provided on overleaf) for the coursework. Also, either attach or provide a link to a short video presentation (about 5 min) describing your work and major findings.

Research hypothesis

In this section, you should clearly state your research hypothesis, which should be a concise and testable statement that answers your research question. Explain why you believe that the method you are proposing is a good approach for the problem. Keep in mind that the research question should flow logically from the topic you are investigating. When stated as a single sentence, the research question becomes your research hypothesis or thesis statement. Make sure to clearly articulate your hypothesis and provide a convincing argument for why it is worth testing.

1. Related work and background
2. Provide a survey of prior work related to your project.
3. Make sure to cite at least 10 relevant papers in this section.
4. You don't have to describe all the papers in detail, but for a couple of the most related papers, briefly describe what they did and how/why your approach differs.
5. Your related work section should not read like a list of paper descriptions but rather form a cohesive section.
6. Look for relevant papers by checking out the top NLP conferences (on the website of ACL anthology).
7. Make sure to properly cite the papers. If you're not familiar with LaTeX, add entries to yourbib.bib to get them to show up when you cite them.
8. Your related work section should be no longer than 2 columns in this format.
9. If you need more information, have a look at the overleaf documentation on citation: <https://www.overleaf.com/learn/latex/Biblatex_citation_styles>.
10. Accomplishments
11. Provide a short list of the proposed project tasks from your project proposal.
12. State whether you completed each task or not.
13. If you failed to complete a task, briefly explain why.
14. This section is a summary and should be no longer than a column. You can provide more detail in the full report.
15. For example:
    * Task 1: Preprocess dataset – Completed
    * Tokenised the dataset -- completed
    * Task 2: Build and train (specific baseline model) on collected dataset and examine its performance - Completed
    * Task 3: Build and train fancy model and examine its performance - Completed
    * Task 4: Make fancy model perform better than baseline model - Failed due to time constraints
    * Task 5: Perform in-depth error analysis to figure out what kinds of examples our approach struggles with - Completed.
16. Approach and Methodology
17. What is your approach and how does it work? Explain the core idea behind your approach and how it solves the problem at hand.
18. Do you expect it to fail in similar ways to your baselines? Discuss the limitations of your approach and compare them to the limitations of your baselines.
19. Did you manage to complete a working implementation? If so, explain the key components of your implementation and how they fit together to form the final product.
20. What libraries did you use to accomplish this? List the libraries you used to implement your approach and explain how they contributed to the final product.
21. Did you rely on help from any existing implementations? If so, provide links to the existing implementations and explain how you used them in your work.
22. What models did you implement yourself, and what files in your uploaded code are associated with these models? Provide a list of the models you implemented and specify which files in your uploaded code correspond to each model.
23. Are there any issues that you could not solve? Discuss any challenges or roadblocks you encountered during the implementation and explain how you attempted to address them.
24. Dataset
25. Emphasize the importance of examining the dataset.
26. Provide specific examples from your dataset that illustrate the task you are working on.
27. Explain the properties of the data that make your task challenging.
28. Report the source of the dataset, including its basic statistics such as size, number of words/sentences/documents.
29. Additionally, provide other statistics that are relevant to your task.

For instance, you could structure this section as follows:

1. Introduction to the dataset
2. Examples of the dataset
3. Properties of the data that make the task challenging.
4. Source of the dataset and basic statistics
5. Other statistics relevant to the task

## Dataset preprocessing

Please state the types of preprocessing that you have applied to the data. What were the difficulty associated with this. Why were these the suitable set of preprocessing techniques?

1. Baselines

What baselines are you using? Why are these useful baselines for the task and the dataset?

1. Results, error analysis

Report the performance of your model and compare it to the baselines. What results did your model achieve, and how do these results compare to your baselines? Provide detailed results from your experiments and compare them to the results of your baselines. Explain what these results mean in the context of your project and how they contribute to the overall goals of your work.

Analyze the kinds of inputs that your baselines fail to handle and examine whether your approach addresses these failures. Identify any syntactic or pattern based similarities among the difficult examples. I strongly recommend conducting manual error analysis by annotating 20-50 examples that were not handled correctly by the model and analyze various properties of these examples.

1. Lessons learned and conclusions

It is important to reflect on the lessons learned, observations made, and outcomes achieved. In the conclusion section, you should summarize the most salient observations and draw relevant conclusions.

Consider what you have learned from completing this project. Were there any techniques that you found particularly useful? Did you encounter any unexpected challenges or difficulties? What did you do to overcome these challenges? What new knowledge or insights have you gained from this experience?

Additionally, reflect on the outcomes of your project. Did you achieve your original goals and objectives? If not, what factors contributed to these outcomes? What could you have done differently to improve the results?

References

It would be ideal if the references are in the same format as shown below. This is the ACL format.   
Or you could follow the APA format. You can use google scholar to get references in this format.

Galen Andrew and Jianfeng Gao. 2007. Scalable training of L1-regularized log-linear models. In *Proceedings of the 24th International Conference on Machine Learning*, pages 33–40.

Isabelle Augenstein, Tim Rocktäschel, Andreas Vlachos, and Kalina Bontcheva. 2016. Stance detection with bidirectional conditional encoding. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*, pages 876–885, Austin, Texas. Association for Computational Linguistics.

James W. Cooley and John W. Tukey. 1965. An algorithm for the machine calculation of complex Fourier series. *Mathematics of Computation*, 19(90):297–301.

1. Appendices

Any additional data that you consider useful and essential to support your analyses that you think is interesting to include can be included here.