Comp1804 Report: Title

ENTER YOUR STUDENT ID ONLY

DATE

Word count: ADD WORD COUNT

Notes (delete these notes from final report!):

* The title should be either “Predicting the severity of road accidents in the UK” or “Predicting the topic of customers’ banking questions” depending on the option you chose. Do not use sub-titles.
* If you use abbreviations, define them.

# Executive summary

Briefly summarize what the report contains. That is: the task you are solving and why it is important; the outline of the ML methods you implemented and any experiments performed; the summary of your results and your conclusions. The executive summary should be between 100 and 200 words.

# 1. Exploratory data analysis

Describe the exploratory data analysis performed and comment on what its implications are for the machine learning task. As part of the exploratory data analysis, you should use dimensionality reduction techniques to show the dataset (including the target labels) in a 2-dimensional plot.

# 2. Data preprocessing

Describe the steps performed for data cleaning, splitting (training/validation/test) and pre-processing (where appropriate: normalization/standardization, imputation of missing values, feature encoding, over/under-sampling, text processing). Provide justifications, based on theory and/or experiments, for your design choices.

# 3. Classification using traditional machine learning

Describe your solution to the classification task (accident severity for option 1 and question topic for option 2) using traditional machine learning techniques. You should describe the final model hyper-parameters in details, ideally in a table, and give a brief explanation of how the algorithm works.

Describe the experiments you did to optimize your model (hyper-parameters optimization and comparison with other models) – these experiment should be rigorous and follow best practice. Provide justifications, based on theory and/or experiments, for your design choices.

Evaluate the model performance using a) a confusion matrix; b) two performance metrics (explain what each metric compute, why it is an appropriate metric to use and what are the implications of the results for the task you are solving); c) a comparison with one “trivial” baseline (for example, random guess or majority class).

Results should be presented in well-formatted figures and tables. Include captions for all your figures and tables. In your captions, please include a brief description of what is in the figure/table and what the reader should focus on when looking at it.

# 4. Classification using neural networks

Describe your solution to the classification task (accident severity for option 1 and question topic for option 2) using neural networks. You should describe the final model hyper-parameters in details, ideally in a table, and give a brief explanation of how the algorithm works.

Describe the experiments you did to optimize your model (hyper-parameters optimization and comparison with other neural networks) – these experiment should be rigorous and follow best practice. Provide justifications, based on theory and/or experiments, for your design choices.

Evaluate the model performance using a) a confusion matrix; b) two performance metrics (explain what are the implications of the results for the task you are solving); c) a comparison with one “trivial” baseline (for example, random guess or majority class). Unless you are using performance metrics that are different from the previous section, you do not need to explain again what each metric computes. However, you should compare the results with those from the previous section.

Results should be presented in well-formatted figures and tables.

# 5. Ethical discussion

Identify and discuss some of the social and ethical implications of your chosen task, from data collection and processing to the ML prediction. It is highly recommended that you structure the discussion using either Data Hazard Labels or the Ethical OS Toolkit. The discussion should take into account communities and people that may be affected by the ML system.

# 6. Recommendations

You should provide three bullet points detailing the following:

* Which of your machine learning model is the best candidate for the task and why.
* Whether the final model is good enough to be used in practice and why (or why not).
* Your top suggestion for future improvements and why.

# 7. Retrospective

The last section in the report is a reflection on the work you have done for this coursework. You should write a maximum of 50 words answering the following question: if you were to start the coursework all over again, what aspect of it would you want to investigate more in depth and why?

# 8. References

If needed, cite references and sources used at the end. These can be academic papers, blogs, code repositories, and more. Remember to give credit if someone else’s work has helped you complete your coursework.