THE MAIN TITLE OF THE THESIS

THE SUBTITLE OF THE THESIS IF IT HAS ONE

STEFAN WINTER

THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
MASTER OF SCIENCE IN DATA SCIENCE & SOCIETY
AT THE SCHOOL OF HUMANITIES AND DIGITAL SCIENCES
OF TILBURG UNIVERSITY

STUDENT NUMBER

2067606

COMMITTEE

Dr. Peter Hendrix
prof. dr. The Second Reader

LOCATION

Tilburg University
School of Humanities and Digital Sciences
Department of Cognitive Science &
Artificial Intelligence
Tilburg, The Netherlands

DATE

November 17, 2021

ACKNOWLEDGMENTS

Some room for acknowledgements.

THE MAIN TITLE OF THE THESIS

THE SUBTITLE OF THE THESIS IF IT HAS ONE

STEFAN WINTER

Abstract

This is where the abstract goes. Don't forget to change the variables in `main.tex` to change all general placeholders shown in this document. The `frontmatter.tex` file should be left alone.

1 INTRODUCTION

The internet has enabled humankind to access information, exchange ideas and become part of a community. Of course, that also applies to online message boards. Ever since the internet found mainstream adoption, people joined those message boards to discuss trading the stock market. Most recently the Reddit forum wallstreetbets attracted a lot of interest and now counts over 10 million members. In this subreddit, members talk about various investment ideas. However, most of those ideas are of speculative nature with members trying to get rich quick, usually by using risky derivatives like stock options. While the wallstreetbets community undoubtedly minted many millionaires, there are also numerous people who lost their life savings.

Even though the Reddit subforum was already founded in 2012, it got most of its media attention in 2021 due to a short-squeeze of the GameStop (GME) stock that drove the stock price up several hundred percent. However, it was not the rapid price appreciation that amazed market participants. Instead, it was the unprecedented decentralized and coordinated buying of Gamestop shares by members of the wallstreetbets community that attracted attention (**?**). Interestingly, the story repeated itself when forum members sent other stocks, such as AMC Entertainment and BlackBerry to the moon.

Organizing the mass-coordinated buying of stock, however, requires that enough participants share the same sentiment. Some research shows that social media sentiment has a particularly strong impact on uninformed traders

> *Can sentiment analysis of the WallStreetBets Reddit-forum be used to predict daily changes in the stock price of Gamestop?*

The sub-questions can be listed seperately, as such:

RQ1 *Which sentiment analysis approach performs best on predefined key performance indicators?*

RQ2 *How can the domain-specific language of the Reddit forum WallStreetBets best be incorporated into sentiment analysis?*

RQ3 *Which machine learning algorithm delivers the best predictive performance for changes in daily stock prices of selected securities based on the sentiment analysis performed earlier?*

Or, format it as you desire (tip: you can nest `itemize` as well). You can alternate *emph* and **textbf** however you wish. This should cover most of the things required for the introduction.

2 RELATED WORK

Copy paste BibTeX code[1] and put it in `references.bib`. After, you can cite some work (**?**) – using `\citep`. You can refer to the author of e.g. **?** directly like using `\cite` (this does not work when using bracket-citation). If you use bracket-style,[2] you might want use `\citeauthor` when citing, like: see **?** **?**. If you want to add pages you can use brackets in `\citep[][p. 5]{mackay2003information}`, which looks like: (**?**, p. 5). The first brackets can be used for things like *see*, and *e.g.*. If you want to cite multiple authors, simply comma-separate them (`\citep{minsky1961steps,mackay2003information}`) and it will aggregate them automatically (**??**).

3 METHOD

If you define any equations (`\begin{equation}...`), you probably might want to define everything using math operators (e.g., `$D$`) and cite the work (!). So for example, following **?**, representing a document $d \in D$ as $\text{tf}(d)$, we define a probabilistic model $(d|Y = y)$ for all documents in class y, and select y most likely to generate d:

$$\hat{y} = \underset{y}{\text{argmax}}\ P(d|y) \cdot P(y) \tag{1}$$

With this we can detect spam (see Figure 1) or bots. Note that the figures (and tables for that matter) might not always be placed in this section (oh no)! LaTeX determines where to best put your objects, so don't

[1] Using e.g. the quote icon in GScholar, then BibTeX at the bottom.
[2] Find the `natbib` part in the `main.tex` LaTeX script.

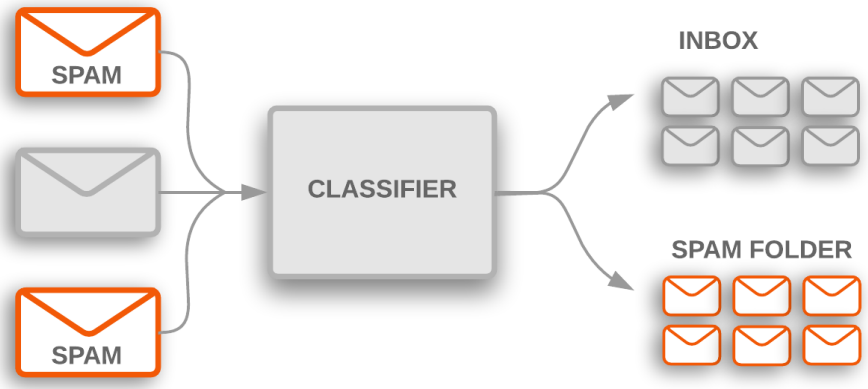

Figure 1: Spam classification example. Source: Google (CC BY 4.0).

Table 1: Best scoring models classifying bots, on Twitter and Facebook respectively. F_1 scores report positive (bot) class. Outline text left (l) and numbers right (r).

		F_1 score	
PCA	Models	Twitter	Facebook
300	Linear SVM ($C = 0.1$)	0.51	**0.91**
	Random Forest ($S = 5, F = 5$)	0.71	0.85
	Naive Bayes	0.61	0.73
500	Linear SVM ($C = 0.1$)	0.55	0.84
	Random Forest ($S = 5, F = 5$)	**0.76**	0.71
	Naive Bayes	0.41	0.64
	Majority	0.50	0.60

worry about that. The reader will find them. **NOTE**: this Figure has a Creative Commons license; you cannot re-use other authors' figures without explicit permission or permissive licensing (as this would mean copyright infringement). You can refer to the equations as well (Equation 1)!

4 RESULTS

You have results and want to show them — probably in a table of some kind as you can see in Table 1. Highlight important scores with `\textbf{}`, use booktabs commands for structure: `\toprule \midrule \bottomrule`. APA does not allow vertical lines.

4.1 *Some Model*

If you have anything specific to talk about, use subsections, and refer to them as Section 4.1. Don't use paragraphs or subsubsections.

5 DISCUSSION

The results were promising!

6 CONCLUSION

Done.

REFERENCES

Anand, A., & Pathak, J. (2021). Wallstreetbets against wall street: The role of reddit in the gamestop short squeeze. *Indian Institute of Management Bangalore Research Paper Series*, *20*.

Lewis, D. D., & Gale, W. A. (1994). A sequential algorithm for training text classifiers. In *Sigir'94* (pp. 3–12).

MacKay, D. J., & Mac Kay, D. J. (2003). *Information theory, inference and learning algorithms*. Cambridge university press.

Minsky, M. (1961). Steps toward artificial intelligence. *Proceedings of the IRE*, *49*(1), 8–30.

APPENDIX A

If you have nothing to append: remove this. You can do a page referral for these, like: Appendix A (page 4).

APPENDIX B

And this!