An Analysis of Machine Learning and Deep Learning techniques for the Detection of Sarcasm in Text

Student Name: Molly Hayward

Supervisor Name: Dr Noura Al-Moubayed

Submitted as part of the degree of BSc Computer Science to the

Board of Examiners in the Department of Computer Sciences, Durham University

Abstract —

Context / Background – Sarcasm presents a fatal problem for classic sentiment analysers, as text that appears to be positive at surface-level can instead convey an alternate negative meaning. There exists an extensive body of literature covering machine learning and deep learning approaches to natural language processing problems, however in the application domain of sarcasm detection, they mostly display low accuracy.

Aims – Despite the challenges in this domain, the ultimate aim of this project is to produce a tool that can detect sarcasm with a *high degree* of accuracy.

Method – In my endeavour to realise this aim, I will implement state-of-the-art word embedding and text classification techniques.

Results – Through extensive experimentation, I found that

Conclusions – Following this experimentation, I conclude that

This section should not be longer than half of a page, and having no more than one or two sentences under each heading is advised. Do not cite references in the abstract.

Keywords — Machine learning, Deep learning, Sarcasm detection

I INTRODUCTION

This section briefly introduces the general project background, the research question you are addressing, and the project objectives. It should be between 2 to 3 pages in length. Do not change the font sizes or line spacing in order to put in more text.

Note that the whole report, including the references, should not be longer than 20 pages in length. The system will not accept any report longer than 20 pages. It should be noted that not all the details of the work carried out in the project can be represented in 20 pages. It is therefore vital that the Project Log book be kept up to date as this will be used as supplementary material when the project paper is marked. There should be between 10 and 20 referenced papers—references to Web based pages should be less than 10%.

II RELATED WORK

This section presents a survey of existing work on the problems that this project addresses. it should be between 2 to 4 pages in length. The rest of this section shows the formats of subsections as well as some general formatting information for tables, figures, references and equations.

A Main Text

Title	**Size**	**Mean Length**
News Headlines Dataset For Sarcasm Detection	26709 instances +ve: 43.9% -ve: 56.1%	9.85 words
SARC	1010826 instances +ve: 50% -ve: 50%	10.46 words

The font used for the main text should be Times New Roman (Times) and the font size should be 12. The first line of all paragraphs should be indented by 0.25in, except for the first paragraph of each section, subsection, subsubsection etc. (the paragraph immediately after the header) where no indentation is needed.

B Figures and Tables

In general, figures and tables should not appear before they are cited. Place figure captions below the figures; place table titles above the tables. If your figure has two parts, for example, include the labels "(a)" and "(b)" as part of the artwork. Please verify that figures and tables you mention in the text actually exist. make sure that all tables and figures are numbered as shown in Table 1 and Figure 1.

Table 1: UNITS FOR MAGNETIC PROPERTIES

Symbol	Quantity	Conversion from Gaussian

C References

The list of cited references should appear at the end of the report, ordered alphabetically by the surnames of the first authors. References cited in the main text should use Harvard (author, date) format. When citing a section in a book, please give the relevant page numbers, as in (Budgen 2003, p293). When citing, where there are either one or two authors, use the names, but if there are more than two, give the first one and use "et al." as in , except where this would be ambiguous, in which case use all author names.

You need to give all authors' names in each reference. Do not use "et al." unless there are more than five authors. Papers that have not been published should be cited as "unpublished" (Euther 2006). Papers that have been submitted or accepted for publication should be cited as "submitted for publication" as in (Futher 2006) . You can also cite using just the year when the author's name appears in the text, as in "but according to Futher (2006), we ...". Where an authors has more than one publication in a year, add 'a', 'b' etc. after the year.

III SOLUTION

This section presents the solutions to the problems in detail. The design and implementation details should all be placed in this section. You may create a number of subsections, each focussing on one issue.

This section should be between 4 to 7 pages in length.

IV RESULTS

this section presents the results of the solutions. It should include information on experimental settings. The results should demonstrate the claimed benefits/disadvantages of the proposed solutions.

This section should be between 2 to 3 pages in length.

V EVALUATION

This section should between 1 to 2 pages in length.

VI CONCLUSIONS

This section summarises the main points of this paper. Do not replicate the abstract as the conclusion. A conclusion might elaborate on the importance of the work or suggest applications and extensions. This section should be no more than 1 page in length.

The page lengths given for each section are indicative and will vary from project to project but should not exceed the upper limit. A summary is shown in Table 2.

Table 2: SUMMARY OF PAGE LENGTHS FOR SECTIONS

	Section	Number of Pages
I.	Introduction	2–3
II.	Related Work	2–3
III.	Solution	4–7
IV.	Results	2–3
V.	Evaluation	1-2
VI.	Conclusions	1

References

Budgen, D. (2003), *Software Design*, 2nd edn, Addison Wesley.

Euther, K. (2006), Title of paper. unpublished.

Futher, R. (2006), Title of paper 2. submitted for publication.