

# 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 is a powerful linguistic anomaly that when present in text, can alter its meaning entirely. It's detection proves a significant challenge for traditional sentiment analysers. This highlights the scope for innovative machine learning and deep learning solutions to this complex problem.

**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 implemented 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 Do not cite references in the abstract.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

Sarcasm is defined as the use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticize something in a humorous way [6]. Sarcasm transforms the sentiment polarity of a positive or negative utterance into its opposite, thereby posing a significant challenge within sentiment analysis, and the wider field of natural language processing. Sentiment analysis is the task of determining the sentiment polarity of text - typically whether the author is in favour of, or against, a certain topic. It is increasingly common for organisations to use sentiment analysis in order to gauge public opinion on their products or services, however classic sentiment analysers cannot deduce the implicit meaning of sarcastic text and will wrongly classify the author's opinion. This highlights the scope for automatic sarcasm detection, as advancements in sarcasm detection research have the potential to vastly improve the sentiment-analysis task.

A rapidly growing area of natural language processing, there is a demand for solving this problem as classic sentiment analysers incorrectly deduce the sentiment polarity of sarcastic text.

It poses a significant challenge to classic sentiment analysers, which deduce the sentiment polarity of text based-upon surface-level features such as the number of positive or negative influences in a sentence.

Sarcasm transforms the sentiment polarity of a positive or negative utterance into its opposite, thereby posing a significant challenge within the field of natural language processing. For example, classic sentiment analysers typically use surface-level features to determine sentiment polarity, such as the frequency of positive or negative influences in a sentence, hence they are likely to make incorrect deductions when faced with sentiment incongruity. This phenomenon lies at the heart of sarcasm, therefore advancements in sarcasm detection research have the potential to vastly improve the sentiment-analysis task.

Related work in this domain has proven that identifying sarcasm in text is incredibly challenging. This is due, in part, to the lack of an all-encompassing, universal definition of sarcasm. In a 2011 study, González-Ibáñez et al. [4] showed that humans classify statements as either sarcastic or non-sarcastic differently - in their second study, three human judges unanimously agreed on a label less than 72% of the time. This is in spite of the fact that sarcasm is fundamentally a *human* construct, highlighting the differences in sarcasm perception. Oftentimes, behavioural and contextual cues are more powerful indicators of sarcasm than the words themselves, however they disappear when an interaction is transcribed. Consider the scenario where a person is congratulated on their hard work, despite the obviousness of them having not worked hard at all. Without context, the statement becomes insignificant and the sarcastic undertones are lost.

Sarcasm transforms the sentiment polarity of a positive or negative utterance into its opposite, thereby posing a significant challenge within the field of natural language processing. For example, classic sentiment analysers typically use surface-level features to determine sentiment polarity, such as the frequency of positive or negative influences in a sentence, hence they are likely to make incorrect deductions when faced with sentiment incongruity. This phenomenon lies at the heart of sarcasm, therefore advancements in sarcasm detection research have the potential to vastly improve the sentiment-analysis task.

as text can have an implicit meaning which is contradictory to that which is visible at surface-level.

At the heart of sarcasm lies sentiment incongruity, whereby positive or neutral language is used to convey a negative message.

At the heart of sarcasm lies sentiment incongruity, such as the use of positive or neutral language to convey a negative message.

Although it can arise in many forms, it often involves some level of sentiment incongruity, such as the use of positive or neutral language to convey a negative message.

While this accurately describes the composition of *some* sarcastic statements, sarcasm can arise in many forms, hence it cannot be wholly encapsulated under one sweeping generalisation.

as it often involves the use of positive or neutral language to convey negative sentiment. Related work in this domain has proven that identifying sarcasm in text is incredibly challenging. This is due, in part, to the lack of an all-encompassing, universal definition of sarcasm. In a 2011 study, González-Ibáñez et al. [4] showed that humans classify statements as either sarcastic or non-sarcastic differently - in their second study, three human judges unanimously agreed on a label less than 72% of the time. This is in spite of the fact that sarcasm is fundamentally a *human* construct, highlighting the differences in sarcasm perception. Oftentimes, behavioural and contextual cues are more powerful indicators of sarcasm than the words themselves, however they disappear when an interaction is transcribed. Consider the scenario where a person is congratulated on their hard work, despite the obviousness of them having not worked hard at all. Without context, the statement becomes insignificant and the sarcastic undertones are lost.

People perceive sarcasm in different ways

At the heart of sarcasm, lies sentiment incongruity. This describes the use of positive language to insincerely describe a

widespread miscomprehension of sarcastic language

, and what signifies sarcastic intent

Use sentiment features, sentiment incongruity Observing that people are inclined to be more sarcastic towards specific subjects such as weather or work Feed features to trainer such as support vector machine

Represent input words numerically, One-hot encoding has two main disadvantages, firstly, each vector is the size of the vocabulary, with a single 1 in the column marking the word. This creates sparse inefficient representations.

It's usage differs from culture-to-culture and across borders

All of these components make sarcasm detection an extremely complex task for both humans and computers. Despite this, most humans can recognise sarcasm most of the time. If we could replicate this performance, or even perhaps improve upon it, we could move towards a more concrete definition of what *makes* a statement sarcastic.

Oftentimes, sarcasm is made from more than just words - which precisely describes the input to our classification models.

There is no precise formulation of sarcasm

Datasets in this domain are messy, often because they u

Sarcasm has no set rules describing its formulation. So prevalent are these discrepancies, even humans cannot recognise sarcasm in all of its forms.

As more data becomes available, the need to automatically

Misinterpreting sarcasm as a literal statement

Humans struggle with this task.

This highlights the scope for automating this process, and the need for an innovative solution.

The **research questions** guiding this project are as follows –

*Which linguistic cues indicate sarcastic intent in written text? How can a model be used to detect the words that correlate more to sarcastic labels? Do deep learning techniques perform better than machine learning approaches for sarcasm detection? Can the detection of sarcasm improve the sentiment analysis task? Does the proposed solution perform well on other datasets?*

This presents a fatal problem for classic sentiment analyzers, as text that appears to be positive at surface-level can instead convey an alternate negative meaning. Furthermore, failure to recognize sarcasm can be a human issue too, owing to the fact that its presence (or lack thereof) is considered rather subjective. González-Ibáñez et al(2011) [4] showed that different humans do not classify statements as either sarcastic or non-sarcastic in the same way - for example, in their second study all three human judges agreed on a label less than 72% of the time. Truly, the existence of sarcasm can only be conclusively affirmed by the author. Additionally, developmental differences such as autism, as well as cultural and societal nuances cause variations in the way different people define and perceive sarcasm.

In similar studies, the terms irony and sarcasm are used interchangeably. Hence, I have provided an overview of their definitions, accompanied by real examples found among the datasets used in this project.

Table 1: Definition of subjective terms

Term	Definition	Example
Irony	The use of words that are the opposite of what you mean, as a way of being funny. [5]	Example of irony
Sarcasm	The use of remarks that clearly mean the opposite of what they say, made in order to hurt someone's feelings or to criticize something in a humorous way. [6]	Example of sarcasm

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.

## II RELATED WORK

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. This could be due, in part, to the absence of a concrete, all-encompassing definition of sarcasm. N.B. this may also contribute to poor human

accuracy in detecting a fundamentally human construct.

Some studies consider irony and sarcasm identical, [], while others report a discrepancy between their meanings.

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. 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%.

#### **A Main Text**

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

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 [1, 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” [2]. Papers that have been submitted or accepted for publication should be cited as “submitted for publication” as in [3] . You can also cite using just the year when the author’s name appears in the text, as in “but according to Futher (?), 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

- [1] David Budgen. *Software Design*. Addison Wesley, 2nd edition, 2003.
- [2] K. Euther. Title of paper. unpublished, 2006.
- [3] R. Futher. Title of paper 2. submitted for publication, 2006.
- [4] Roberto González-Ibáñez, Smaranda Muresan, and Nina Wacholder. Identifying sarcasm in twitter: a closer look. In *Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: Short Papers-Volume 2*, pages 581–586. Association for Computational Linguistics, 2011.
- [5] Cambridge University Press. Irony: meaning in the cambridge english dictionary. <https://dictionary.cambridge.org/dictionary/english/irony>, 2019. Last accessed: 23 Oct 2019.
- [6] Cambridge University Press. Sarcasm: meaning in the cambridge english dictionary. <https://dictionary.cambridge.org/dictionary/english/sarcasm>, 2019. Last accessed: 23 Oct 2019.