University of Sheffield

Automatically identifying complaints in social media using transformers



Nitin Sunny Mathew

Supervisor: Nikolaos Aletras

A report submitted in fulfilment of the requirements for the degree of MSc in Data Analytics

in the

Department of Computer Science

September 13, 2023

Declaration

All sentences or passages quoted in this report from other people's work have been specifically acknowledged by clear cross-referencing to author, work and page(s). Any illustrations that are not the work of the author of this report have been used with the explicit permission of the originator and are specifically acknowledged. I understand that failure to do this amounts to plagiarism and will be considered grounds for failure in this project and the degree examination as a whole.

Name: Nitin Sunny Mathew

Signature: Nitin Sunny Mathew

Date: 13-Sep-2023

Abstract

A complaint is a statement made by a person or an entity with the intent to indicate something is unacceptable or unsatisfactory. This is commonly used in various aspects of day-to-day life including when conducting business operations. With the proliferation of social media across our lives and the active enablement of such platforms by organisations for user engagement, it has become a common medium for users to raise complaints. With such complaints being publicly visible, it is imperative for organisations to identify, prioritise and respond to these complaints swiftly. Automatically identifying complaints in social media is an active area of research. In the past few years, the focus has been on using NLP approaches driven by developments in transfer learning and transformer-based models.

In this paper, the use of these approaches are extended by assessing 'lightweight' transformer based models such as DistillBERT and MobileBERT which are meant to reduce the time required for fine-tuning as well as inference. The performance of these 'lightweight' models is compared with the traditional transformer models including BERT, ROBERTA, BERTweet for this particular task. The dataset used will consist of anonymised and annotated Twitter data utilized in previous research and currently available in the public domain. In addition, the nature of complaints will be analysed from a linguistic perspective along with discussions on state-of-the-art approaches for such NLP tasks.

Contents

1	Intr	roduction	1					
	1.1	Background	1					
	1.2	Aims and Objectives	2					
	1.3	Overview of the Report	2					
2	$\operatorname{Lit}\epsilon$	erature Survey	3					
	2.1	A Section that Contains a Figure	3					
	2.2	A Section that Contains a Table	4					
	2.3	Summary	5					
3	Ana	alysis	6					
	3.1	Project Requirements	6					
	3.2	Another Section	7					
	3.3	Ethical, Professional and Legal Issues	7					
4	Pla	nning	8					
	4.1	Risk Analysis	8					
	4.2	Project Plan	8					
	4.3	Another Section if You Need It	9					
5	Cor	nclusions	10					
$\mathbf{A}_{\mathtt{J}}$	ppen	dices	12					
\mathbf{A}	An Appendix of Some Kind							
В	3 Another Appendix 1							

List of Figures

2.1	A simple figure in LATEX. R	Reproduced from	http://tinyurl.com/	nqtrlj5 with
	the permission of the copyris	ght owner		

List of Tables

2.1	A simple table	in LATEX.																							Ę
-----	----------------	-----------	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	--	---

Introduction

1.1 Background

In the act of complaining, dissatisfaction or annoyance is expressed by a person or entity in response to a previous or ongoing event that has negatively impacted them [5]. It provides an avenue to direct dissatisfaction to the appropriate organisation or individual with the hope of rectification or redressal. The event or action could be concerning a product or service procured by the concerned person or entity. The need to recognise, acknowledge and act on complaints is of significant importance to businesses and organisations to retain their customers while maintaining their reputations.

Until the advent of online platforms and specifically social media, the impact of negative word-of-mouth was confined to a relatively limited audience. However, since then complaints posted online have the potential to rapidly go viral, reaching millions of individuals and significantly damaging a company's brand reputation and goodwill in a short period [7]. Customers are able to express their complaints directly, conveniently, and with enhanced effectiveness to organisations through multiple social media channels and platforms [1].

In addition to the timely addressing of customer complaints, automated detection of complaints in natural language has a number of other purposes. Linguists could gain a more detailed understanding of the context, intent, and various types of complaints on a larger scale while psychologists could utilise this information to identify the underlying human traits that drive the behaviour and expression of complaints. Developing downstream natural language processing (NLP) applications, such as dialogue systems is another use case of this task [6].

Attempting to identify complaints manually through the multitude of posts and streams coming through the various social media channels is neither practical nor scalable. Various approaches to automate this task have been explored. The traditional vector-space method utilizing dictionaries has been applied in other text classification tasks [4]. Latent Semantic Indexing based on Singular Value Decomposition along with linguistic style features has

been utilised to classify emails as complaints or not [2]. In recent years, we have seen the use of various Machine learning and Natural Language Processing (NLP) based approaches for similar classification problems. [6] have analysed the performance of logistic regression over various types of feature spaces against neural-network based models like Multi-layer Perceptron (MLP) and Long Short Term Memory (LSTM). The use of more advanced approaches using transformer based models (BERT and its variants) have shown to have better results as explored by [3]. As part of this paper, the use of BERT and its many variants will be explored further including that of lightweight variants that have been created in the recent past.

1.2 Aims and Objectives

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean c

1.3 Overview of the Report

Lorem ipsum dolor sit amet,

Literature Survey

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. [?] Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum.

$$M = \frac{1}{T} \sum_{t=1}^{T} e(t) / \max_{t} [e(t)]$$
 (2.1)

Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

2.1 A Section that Contains a Figure

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim [?]. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Porttitor eu, consequat vitae,

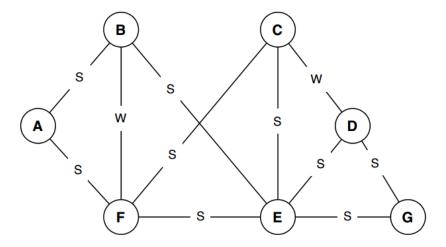


Figure 2.1: A simple figure in LATEX. Reproduced from http://tinyurl.com/nqtrlj5 with the permission of the copyright owner.

eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. See Figure 2.1. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

2.2 A Section that Contains a Table

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, [?] pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula,

porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. See Table 2.1. Nam quam nunc,

A	В	A XOR B
0	0	0
0	1	1
1	0	1
1	1	0

Table 2.1: A simple table in LATEX.

blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

2.3 Summary

Analysis

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

3.1 Project Requirements

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus,

tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

3.2 Another Section

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui.

3.3 Ethical, Professional and Legal Issues

Planning

4.1 Risk Analysis

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

4.2 Project Plan

Lorem ipsum dolor sit amet, consectetuer adipiscing elit. Aenean commodo ligula eget dolor. Aenean massa. Cum sociis natoque penatibus et magnis dis parturient montes, nascetur ridiculus mus. Donec quam felis, ultricies nec, pellentesque eu, pretium quis, sem. Nulla consequat massa quis enim. Donec pede justo, fringilla vel, aliquet nec, vulputate eget, arcu. In enim justo, rhoncus ut, imperdiet a, venenatis vitae, justo. Nullam dictum felis eu pede mollis pretium. Integer tincidunt. Cras dapibus. Vivamus elementum semper nisi. Aenean vulputate eleifend tellus. Aenean leo ligula, porttitor eu, consequat vitae, eleifend ac, enim. Aliquam lorem ante, dapibus in, viverra quis, feugiat a, tellus. Phasellus viverra nulla ut

metus varius laoreet. Quisque rutrum. Aenean imperdiet. Etiam ultricies nisi vel augue. Curabitur ullamcorper ultricies nisi. Nam eget dui. Etiam rhoncus. Maecenas tempus, tellus eget condimentum rhoncus, sem quam semper libero, sit amet adipiscing sem neque sed ipsum. Nam quam nunc, blandit vel, luctus pulvinar, hendrerit id, lorem. Maecenas nec odio et ante tincidunt tempus. Donec vitae sapien ut libero venenatis faucibus. Nullam quis ante. Etiam sit amet orci eget eros faucibus tincidunt. Duis leo. Sed fringilla mauris sit amet nibh. Donec sodales sagittis magna. Sed consequat, leo eget bibendum sodales, augue velit cursus nunc.

4.3 Another Section if You Need It

Conclusions

Bibliography

- [1] Balaji, M. S., Jha, S., and Royne, M. B. Customer e-complaining behaviours using social media. *The Service industries journal* 35, 11-12 (2015), 633–654. Place: London Publisher: Routledge.
- [2] Coussement, K., and Van den Poel, D. Improving customer complaint management by automatic email classification using linguistic style features as predictors. *DECISION SUPPORT SYSTEMS* 44, 4 (2008), 870–882. Place: AMSTERDAM Publisher: Elsevier B V
- [3] JIN, M., AND ALETRAS, N. Complaint Identification in Social Media with Transformer Networks.
- [4] LIANG, C.-Y., Guo, L., XIA, Z.-J., NIE, F.-G., LI, X.-X., Su, L., AND YANG, Z.-Y. Dictionary-based text categorization of chemical web pages. *Information Processing & Management* 42, 4 (2006), 1017–1029.
- [5] OLSHTAIN, E., AND WEINBACH, L. Complaints: A study of speech act behavior among native and non-native speakers of Hebrew. 195–208.
- [6] Preotiuc-Pietro, D., Gaman, M., and Aletras, N. Automatically Identifying Complaints in Social Media.
- [7] TRIPP, T. M., AND GREGOIRE, Y. When unhappy customers strike back on the Internet. MIT Sloan management review 52, 3 (2011), 37. Place: Cambridge Publisher: Sloan Management Review.

Appendices

Appendix A

An Appendix of Some Kind

Appendix B

Another Appendix