Evaluation of different classification algorithms for COVID-19 Pandemic Tweets

Text Mining (732A92)

Stylianos Sidiropoulos (stysi607)

Division of Statistics and Machine Learning
Department of Computer Science
Linköping University

01-2021

Contents

Abstract

Lorem ipsum dolor sit amet, consectetur adipiscing elit. Duis auctor efficitur nisi ac ultricies. Nulla et viverra diam, ut interdum nibh. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nulla facilisi. Aenean consequat aliquet elit eu semper. Fusce posuere quis purus sit amet pretium. Etiam ultrices finibus neque, quis laoreet justo sagittis in.

Etiam iaculis ultricies imperdiet. Cras purus nisl, vehicula nec ipsum hendrerit, tristique bibendum sem. Aenean sed urna euismod, volutpat augue aliquet, congue turpis. Mauris porttitor tempor volutpat. Aliquam ac dui lectus. Maecenas venenatis tempor nibh eu volutpat. Mauris quis posuere est. Nulla purus nisl, viverra facilisis ipsum id, feugiat lacinia lectus. Etiam sodales fermentum euismod. Nulla facilisi. Ut laoreet gravida nisi, at imperdiet magna vestibulum nec. Fusce rutrum ante magna, ut pulvinar sapien mollis posuere.

Keywords: text mining, opinion mining, sentiment analysis, classifing text

GitHub repository: https://github.com/steliossid/covid19-classification

1   Introduction

Introduce the task or research question that you have addressed in your project. What were you trying to do? Why did you choose this project?

The COVID-19 pandemic is having a huge impact on people's lives in terms of not only physical but also mental health. During this pandemic, people changed their daily routine and habits drastically by staying at home, working from home, going outside only to buy food from the supermarket and for physical exercise. People have mixed emotions and feelings about this new routine and they often publish their opinion in social media, such as Twitter. Their opinion includes a small text, limited to 280 characters, where they express how they feel about this pandemic and their new everyday life. Opinions can vary that's why there is a need to classify them into separate categories, which are Positive, Negative and Neutral opinions. Categorizing these opinions means that for each text, a tag/label is assigned to it based on its category, which is a hard and time-consuming task to do for a human, thus it should be done by a computer using Natural language processing and certain text classification algorithms.

The main aim for this project is to analyse some of the most important text classification algorithms and evaluate their performance on classifying correctly tweets related to COVID-19 Pandemic. For this task, the algorithms are trained by getting as input the tweet text and the label it was assigned to it and try to predict the label of unseen tweet texts.

2   Theory

Present relevant theoretical background, and in particular the models that you have used. Where appropriate, use mathematical formulas.

2.1   Multinomial Naive Bayes

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.2   Logistic Regression

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.3 Random Forest Classifier

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.4 Decision Tree Classifier

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.5 Linear Support Vector Classifier (SVC)

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.6 Stochastic Gradient Descent (SGD) Classifier

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.7 Extreme Gradient Boosting (XGB) Classifier

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum

primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

2.8   Recurrent Neural Network (RNN)

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

3   Data

Present your data. What information does it contain? Where did you get it from? What preprocessing did you do, if any?

For this report, an already made dataset from kaggle.com has been used.

Vestibulum non eros tortor. Nullam arcu ante, placerat non rutrum vitae, blandit eget elit. Suspendisse potenti. Interdum et malesuada fames ac ante ipsum primis in faucibus. Sed placerat euismod est nec vulputate. Ut id metus vitae purus pharetra placerat et et sapien. Maecenas lacinia scelerisque leo nec facilisis. Etiam ut mollis dolor. Sed id bibendum elit, non ullamcorper nisl. Morbi sodales mollis commodo. Morbi cursus odio ut mattis accumsan. Aenean ac felis dignissim, luctus est sed, pulvinar velit. Integer turpis elit, posuere nec erat ac, hendrerit tristique ligula.

4   Method

Explain how you carried out your study. Aim to be detailed enough for others to reproduce your results.

Mauris vestibulum venenatis turpis eu blandit. Sed est tellus, auctor id tortor a, pellentesque consequat quam. Phasellus efficitur elit et rutrum egestas. Phasellus sagittis leo non sollicitudin tincidunt. Vivamus rhoncus iaculis aliquam. Proin pretium, risus ut tempor consectetur, nunc augue pretium est, at cursus velit justo vel arcu. Proin dictum ligula a commodo rutrum.

5   Results

Present your results in an objective way. Use tables and charts, but do not forget to also include a summary in text form. Do not interpret your results.
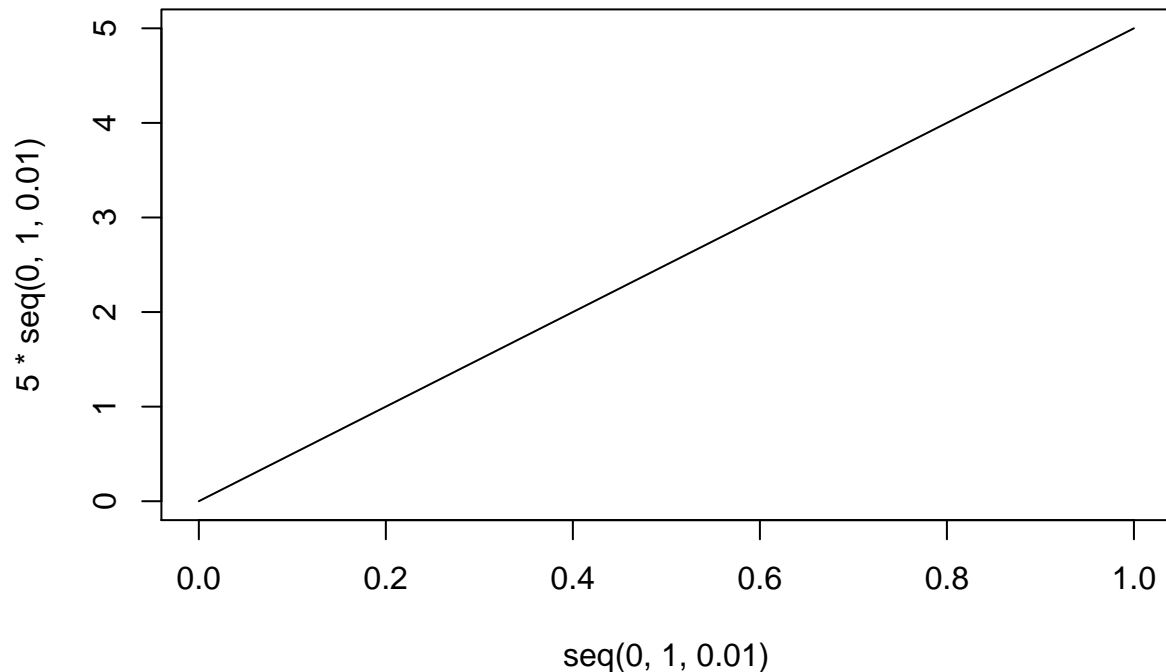
6 Discussion

Analyse your results and discuss the possibilities and limitations of your technical approach. Compare your study to related work.

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

7 Conclusion

Based on your results and their analysis, what new knowledge do you take away from your project?

Phasellus ac consequat nibh, vel condimentum quam. Morbi in est ut odio pharetra feugiat a quis nulla. Donec rhoncus eget nunc vitae ullamcorper. Sed quis bibendum ex. Nullam non augue nec leo facilisis facilisis ullamcorper non turpis. Nullam rhoncus, ipsum non efficitur

viverra, velit ligula hendrerit felis, porta feugiat eros urna ac justo. Vestibulum ante ipsum primis in faucibus orci luctus et ultrices posuere cubilia curae; Nam gravida consectetur rhoncus. Vivamus suscipit ut nulla non vestibulum. Phasellus maximus nisl non odio porta, vel volutpat ligula volutpat.

References