CE807 – Assignment 2 - Final Practical Text Analytics and Report Student ID: 2202396

Student id: XXX

Abstract

Hate speech recognition is an important problem in natural language processing, given the harmful effects of hate speech on individuals and society. In recent years, machine learning models have been used to automatically identify hate speech in text, with promising results. Two commonly used models for hate speech recognition are the Support Vector Classifier (SVC) and the Random Forest (RF).

1 Materials

- Code.py
- Code.ipynb
- Google Drive Folder containing models and saved outputs
- Presentation Video
- Presentation slides

2 Model Selection (Task 1)

We are using two of the models:

1- Random Forest Classifier.

2- Support Vector Machine Classifier.

2.1 Summary for 2 selected Models

2.1.1 1- Random Forest Classifier

An ensemble learning technique known as the Random Forest algorithm mixes various decision trees to produce classification or regression predictions. The RF algorithm aggregates the predictions from numerous decision trees that have been trained on various random subsets of the training data. The RF model can handle high-dimensional, nonlinear data and can calculate the relative relevance of features.

2.1.2 2- Support Vector Machine Classifier

A well-liked machine learning model for classification issues, particularly in text classification, is the support vector classifier. The SVC technique locates a hyperplane in the feature space that maximally separates the various classes of data points. Depending on the type of data and the issue at hand, the SVM classifier can be employed with many kernel functions, including linear, polynomial, and radial basis function (RBF) kernels.

2.2 Critical discussion and justification of model selection

Several studies have looked into the efficacy of SVC and RF models for recognising hate speech in various datasets. For instance, SVC and RF models were trained on a dataset of tweets annotated for hate speech and assessed on a held-out test set in a work by (Olteanu et al., 2018). The study discovered that the RF model had the maximum precision while the SVC model with an RBF kernel had the highest performance in terms of F1 score.

SVC and RF models were trained on a dataset of tweets labelled for hate speech, offensive language, and neither in a different work by (Davidson et al., 2017). According to the study, the SVC model with a linear kernel outperformed the RF model and produced a high F1 score for recognising hate speech. Despite the positive outcomes of SVC and RF models in the identification of hate speech, there are still issues and restrictions. The definition and annotation of hate speech present a challenge because they might differ among cultures, situations, and people. Another difficulty is the possibility of unfairness and bias problems in the training data and the models, which might exacerbate and reinforce current socioeconomic inequities.

In conclusion, machine learning techniques for recognising hate speech in text include the Support Vector Classifier and Random Forest models. However, the features of the data and the problem will

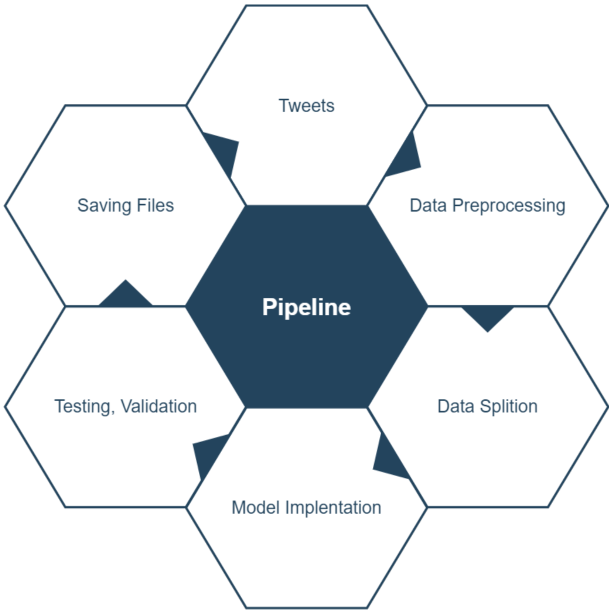

Figure 1: This Diagram explains the overall structure of Program

determine how effective these models are, therefore it is important to carefully analyse their drawbacks.

- It should contain the Figure 1 of both models

3 Design and implementation of Classifiers (Task 2)

The data-set itself is quite impressive combination of a classification problem labels Off and Not and Not almost being the double the amount around 8221 from the total data-set of 12313 entries. For more information look into the Table 1.

The hyper-parameters tuning gave much more precise results but at the price of high computation power and more time complexity which was not worth it and of you would reduce parameters it effected the performance of the model, in case of decision tree it almost predicted True Negatives in classification report as False Positives which is pretty bad, finding the balance between the two was the most hardest part.
In terms of performance, Random Forest classifier was the winner overall as it performs better with long data-sets and the depth of the tree makes the accruacy of the results much more better (Fauzi and Yuniarti, 2018) On the other hand even though the support vector was able to perform well generally and was nit that far behind the Random forest, it is reasonable as SVC tend to be less accurate on data-sets above 1k number of entries (Karamizadeh et al., 2014)

Dataset	Total	% OFF	% NOT
Train	12313	4092	8221
Valid	927	308	619
Test	860	240	620

Table 1: Dataset Details

Model	F1 Score
Model 1	0.6969769769769769
Model 2	0.6461390755726237

Table 2: Model Performance

4 Data Size Effect (Task 3)

As mentioned above the task of the assignment was to divide training data-set into subsequent sets in order to get relatively accurate performance measures to have better understanding and results in terms of the different classifiers. The data-set was divided into 1/4, 1/3, 1/2 ratios respectively.
The hyper tuning unfortunately caused program to give no accurate readings and some times resulting in the time complexity which we as a data scientist try to avoid so had to remove the hyper-parameters.
In terms of performance changing the data-set, it wasn't a huge difference as previously, but definitely there was some difference in accuracy and with more amount of data the accuracy increased automatically

Data %	Total	% OFF	% NOT
25%	3078	1002	2076
50%	6156	2057	4099
75%	9235	3090	6145
100%	12313	4092	8221

Table 3: Train Dataset Statistics of Different Size

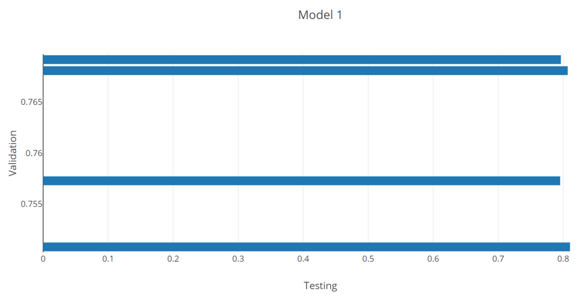

Figure 2: Comparision of Model 1 ccuracy based on different sizes of training datasets

5 Summary (Task 4)

The assignment aims to provide an insight on the performance of classification models and how the change in data-set can impact not only the computing power but also the accuracy of the system. The program works in risk and take where more the data better the accuracy but you risk being much more time consuming in terms of time complexity. The project also had huge insight in using different tolls for natural language processing and different Python libraries and how to work on files being stored on online drive like google drive, saving models that you create and loading them and file management system. All the tasks in the assignment were specifically designed to give an insight on other things

5.1 Discussion of work carried out

I think there is still alot more that can be done in the future given the time, the work so far has a huge insight on how to use different python libraries for different tasks. The most of work was hands on, easy to understand and fun. The project itself was carefully deigned to address all different aspects of typical project giving in more depth overview of different things and possibilities with data-sets.

5.2 Lessons Learned

1- How different methods can be implemented
2- Which models should be selected
3- What ratio of data-set is best suitable

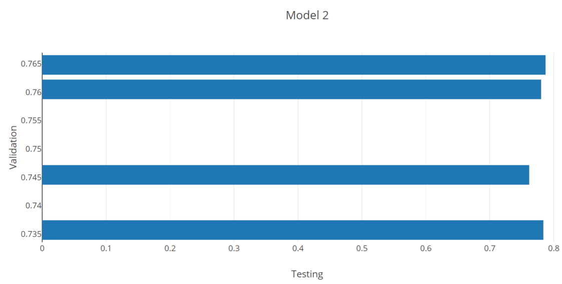

Figure 3: Comparision of Model 2 ccuracy based on different sizes of training datasets

4- Different visualization methods
5- Utilizing Confusion Matrices
Hyper tuning parameters

6 Conlusion

Performance of the model is totally dependent on the type of data-set as well as the structure of method.

References

Thomas Davidson, Dana Warmsley, Michael Macy, and Ingmar Weber. 2017. Automated hate speech detection and the problem of offensive language. In *Proceedings of the international AAAI conference on web and social media*, volume 11, pages 512–515.

M Ali Fauzi and Anny Yuniarti. 2018. Ensemble method for indonesian twitter hate speech detection. *Indonesian Journal of Electrical Engineering and Computer Science*, 11(1):294–299.

Sasan Karamizadeh, Shahidan M Abdullah, Mehran Halimi, Jafar Shayan, and Mohammad javad Rajabi. 2014. Advantage and drawback of support vector machine functionality. In *2014 international conference on computer, communications, and control technology (I4CT)*, pages 63–65. IEEE.

Alexandra Olteanu, Carlos Castillo, Jeremy Boy, and Kush Varshney. 2018. The effect of extremist violence on hateful speech online. In *Proceedings of the international AAAI conference on web and social media*, volume 12.