SUICIDE RELATED TEXT CLASSIFICATION WITH PRISM ALGORITHM

FATIMA CHIROMA[1], HAN LIU[2], MIHAELA COCEA[1]

[1]School of Computing, University of Portsmouth, Portsmouth, United Kingdom
[2]School of Computer Science and Informatics
Cardiff University, Cardiff, United Kingdom
E-MAIL: fatima.chiroma@port.ac.uk, LiuH48@cardiff.ac.uk, mihaela.cocea@port.ac.uk

Abstract:

Raw but valuable user data is continuously being generated on social media platforms. This data is, however, more valuable when they are mined using different approaches such as machine learning techniques. Additionally, this user-generated data can be used to potentially save lives especially of vulnerable social media users, as several studies carried out have shown the correlation between social media and suicide. In this study, we aim at contributing to the research relating to suicide communication on social media. We measured the performance of five machine learning algorithms: Prism, Decision Tree, Naïve Bayes, Random Forest and Support Vector Machine, in classifying suicide-related text from Twitter. The results of the study showed that the Prism algorithm has outperformed the other machine learning algorithms with an F-measure of 0.84 for the target classes (Suicide and Flippant). This result, to the best of our knowledge, is the highest performance that has been achieved in classifying social media suicide-related text.

Keywords:

Text classification; Machine Learning; Prism; Social media; Suicide

1. Introduction

Machine learning is a branch of artificial intelligence that provides techniques which can automatically detect and use patterns for prediction from data [1–3]. According to [3], the use of machine learning has spread rapidly in the last decade especially in computer science, as it has been applied to various and diverse areas such as fraud detection, drug design, web search and recommender systems. Furthermore, one of the most popular tasks in machine learning is classification [3–5], where the category of an unseen instance is judged.

The classification task has been used by employing several machine learning algorithms, one of which is Prism. Prism is a classification rule-learning algorithm that was developed by Cendrowaski in 1987 [6–8]. Although the algorithm is less popular compared with other machine learning algorithmes, such as Decision Tree, Naïve Bayes, Random Forest and Support Vector Machine, it is known to be simple, as well as easy to understand [7, 8]. Additionally, this algorithm uses the separate-and-conquer learning approach [9], which is based on the sequential covering principle [9]: a rule is learned which predicts accurately a value of the target attribute (i.e. class) (the conquer stage); the covered instances are removed (the separate stage) and the process is repeated until all instances are covered.

Therefore, in this study, we will apply the Prism algorithm on short text relating to suicide communications on social media so as to measure its classification performance against four popular machine learning algorithms: Decision Tree (DT), Naïve Bayes (NB), Random Forest (RF) and Support Vector Machine (SVM).

The remainder of the paper is organized as follows: Section 2 discusses the related works on this topic; Section 3 describes the methods used for the data collection and classification; Section 4 discusses the experiment and evaluation; Section 5 and 6 present the results and discussions of the findings; and Section 7 draws the conclusion on the study as well as present some directions for possible future works.

2. Related Work

Numerous research relating to the application of machine learning techniques has been carried out in various areas such as Finance [10], Medicine [11–13], Safety [14] and Economical problems [15] to name a few. However, research relating to Prism rule learning algorithm is limited, therefore research for its application to short informal text is almost non-existent.

The Prism algorithm was originally developed by [6] for addressing the replicated sub-tree problem that usually arises with decision tree learning algorithms. As a rule learning algorithm, the prism algorithm is capable of selecting attributes based on their importance to a specific class. That is, it selects a target class and learns a set of rules which separates the target class from the rest of the classes; this process is repeated by selecting each class as the target class [5]. This process is illustrated in Figure 1.

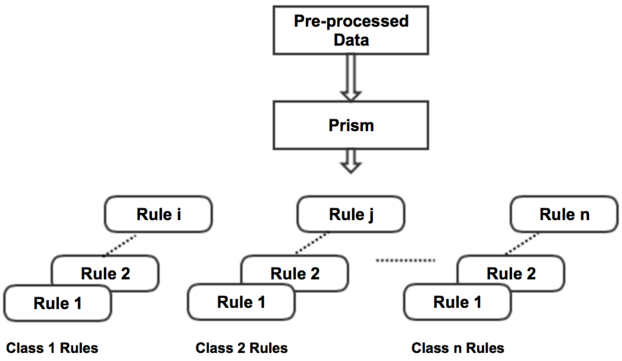

FIGURE 1. Prism workflow overview (adapted from [16]).

Furthermore, for its application, the algorithm in particular, the algorithm was applied for identification of contact lenses types, and the results showed that Prism not only outperformed ID3 (a decision tree learning algorithm) in classification accuracy but also produced a smaller number of simpler rules. Another example of the Prism algorithm application to classification problems is the study carried out by [14]. They used Prism along with other techniques using specifically multimodal features to come up with a multi-media data mining framework that effectively detects soccer goal shots.

Also, the Prism algorithm was used by [5] for multi-task feature selection on an image segmentation data set and those image features were evaluated in terms of their relevance to each specific class, through checking which features have been included in at least one of the rules learned by using Prism, where the rules showed a 92% classification accuracy on the data set. In addition, the characteristics of the Prism algorithm were presented in [16] from a granular computing perspective, which aimed to show theoretically that this kind of algorithms fit text classification.

Therefore, in this study we aim to investigate the classification performance potential of the Prism algorithm when applied to short informal text by comparing it with four popular machine learning algorithms: DT, NB, RF and SVM as well as the result of a related study [17] on suicide communication classification.

3 Data Description and Experiment

In order to carry out this study, short informal text relating to suicide on social media is needed. Therefore, a set of suicide communication tweets that were collected and labelled by [17] using lexicon terms and search keywords from known suicide websites and reported news. These data was classified into seven suicide-related categories by expert human annotators and pre-processed using standard text pre-processing techniques [18, 19] as shown in Figure 2.

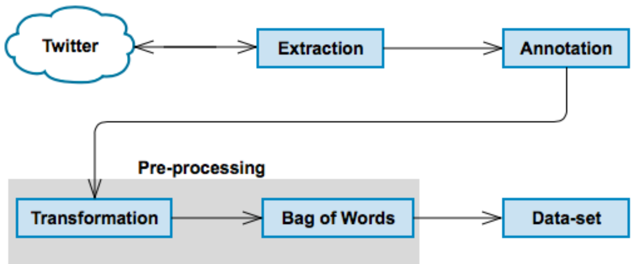

FIGURE 2. Data Preparation Process

A total of 2,000 tweets or instances were collected, however, after rigorous pre-processing only 1060 instances were left for the experiment. Table 1 shows the class categories and the total number of instances per class after pre-processing.

TABLE 1. Instances per class

Description	Class	Instances
Suicide	0	156
Campaign	1	158
Flippant	2	133
Support	3	178
Memorial	4	142
Reports	5	165
Other	6	128
Total	**7**	**1060**

To carry out the experiment, this labelled data was organised into three datasets (one binary and two multi-class) so as to measure the robustness and classification performance of the machine learning algorithms for the classes of interest, i.e. suicide and flippant. Table 2 shows the class distribution for each dataset; the total number is for the processed data.

The experiment for the study was carried out by applying five machine learning algorithms: Prism, DT, NB, RF and SVM on

TABLE 2. Datasets and Instances Distribution

Datasets	Description	*Class*	*Raw*	*Processed*	Total
Binary	Suicide	0	159	156	289
	Flippant	2	133	133	
Three-class	Suicide	0	159	156	1060
	Flippant	2	133	133	
	Non-suicide	1, 3, 4, 5, 6	772	771	
Seven-class	Suicide	0	159	156	1060
	Campaign	1	158	158	
	Flippant	2	133	133	
	Support	3	178	178	
	Memorial	4	142	142	
	Reports	5	165	165	
	Other	5	129	128	

the three data-sets i.e. Binary class, Three-class and Seven-class datasets. Therefore, three different experiments were carried out each for the datasets. 10-fold cross-validation was used to evaluate the performance of the machine learning algorithms. This validation technique is known to limit the influence of the training set's variability on the results [20].

4 Results

In this section we report the results of the experiments, i.e. the performance of the machine learning algorithms when applied to the three datasets, using the standard classification scores, which consist of the Precision (P), Recall (R) and F-measure (F).

4.1 Binary Dataset

The results per class for the binary dataset, which consists of the suicide and flippant classes, showed that the top F-measure value of 0.85 has been achieved by the Prism algorithm. The F-measure achieved by the Prism algorithm belongs to the Suicide, which is the majority class (156 instances) of the two classes. For the other machine learning algorithms, the best F-measure score was achieved by the Random Forest (0.81) also for the suicide class. Table 3 shows the results of this experiment.

TABLE 3. Binary dataset results

Classifier	Measure	*Suicide*	*Flippant*
Prism	Precision	0.85	0.83
	Recall	0.85	0.82
	F-measure	0.85	0.82
DT	Precision	0.80	0.75
	Recall	0.78	0.77
	F-measure	0.79	0.76
NB	Precision	0.54	0.47
	Recall	0.94	0.06
	F-measure	0.69	0.11
RF	Precision	0.72	0.86
	Recall	0.92	0.58
	F-measure	0.81	0.69
SVM	Precision	0.74	0.81
	Recall	0.87	0.63
	F-measure	0.80	0.71

4.2 Three-class Dataset

The Three-class dataset consists of the suicide, flippant and non-suicide (all classes except for suicide and flippant) classes. The results of this experiment showed that the highest performance was achieved by the Prism algorithm, with an F-measure of 0.90 for the Non-suicide class. From the other machine classifiers, the Decision Tree has the best F-measure (0.88) also for the majority class, i.e. non-suicide. Table 4 shows the result of the three-class experiments.

When focusing on the suicide class, we notice that for this dataset, Prism, RF and SVM have the same F-measure of 0.65. For the flippant class, Prism has the highest F-measure of 0.58; NB and RF have very low scores for this class.

TABLE 4. Three-class dataset results

Classifier	Measure	*Suicide*	*Flippant*	*Non-suicide*
Prism	Precision	0.75	0.76	0.86
	Recall	0.58	0.47	0.95
	F-measure	0.65	0.58	0.90
DT	Precision	0.56	0.55	0.86
	Recall	0.65	0.38	0.89
	F-measure	0.60	0.45	0.88
NB	Precision	0.86	0.21	0.79
	Recall	0.48	0.05	0.97
	F-measure	0.62	0.08	0.87
RF	Precision	0.59	0.43	0.83
	Recall	0.73	0.02	0.92
	F-measure	0.65	0.04	0.87
SVM	Precision	0.53	0.38	0.89
	Recall	0.85	0.28	0.82
	F-measure	0.65	0.32	0.86

4.3 Seven-class Dataset

This experiment was carried-out on all the seven classes of the suicide related categories: Suicide, Campaign, Flippant, Support, Memorial, Reports and Other. The result showed that the best F-measure of 0.69 has been achieved by the Random Forest on the campaign class, while the best F-measure obtained by the Prism algorithm was 0.68 on the same class. Table 5 shows the result of this experiment.

TABLE 5. Performance of the Machine Learning Algorithms

	Prism			*DT*			*NB*			*RF*			*SVM*		
Datasets	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*	*P*	*R*	*F*
Suicide	0.69	0.61	0.65	0.48	0.74	0.58	0.88	0.46	0.60	0.42	0.91	0.58	0.48	0.83	0.61
Campaign	0.68	0.68	0.68	0.65	0.68	0.67	0.95	0.23	0.38	0.69	0.70	0.69	0.69	0.60	0.64
Flippant	0.57	0.50	0.53	0.43	0.44	0.44	0.22	0.03	0.05	0.42	0.15	0.22	0.38	0.35	0.36
Support	0.58	0.82	0.68	0.69	0.60	0.64	0.21	0.98	0.34	0.48	0.74	0.58	0.67	0.65	0.66
Memorial	0.66	0.57	0.61	0.32	0.31	0.31	0.48	0.09	0.14	0.43	0.23	0.30	0.39	0.26	0.31
Reports	0.58	0.56	0.57	0.59	0.44	0.50	0.51	0.11	0.18	0.62	0.29	0.40	0.39	0.39	0.38
Other	0.64	0.58	0.61	0.40	0.33	0.36	0.73	0.15	0.25	0.62	0.41	0.49	0.56	0.45	0.50

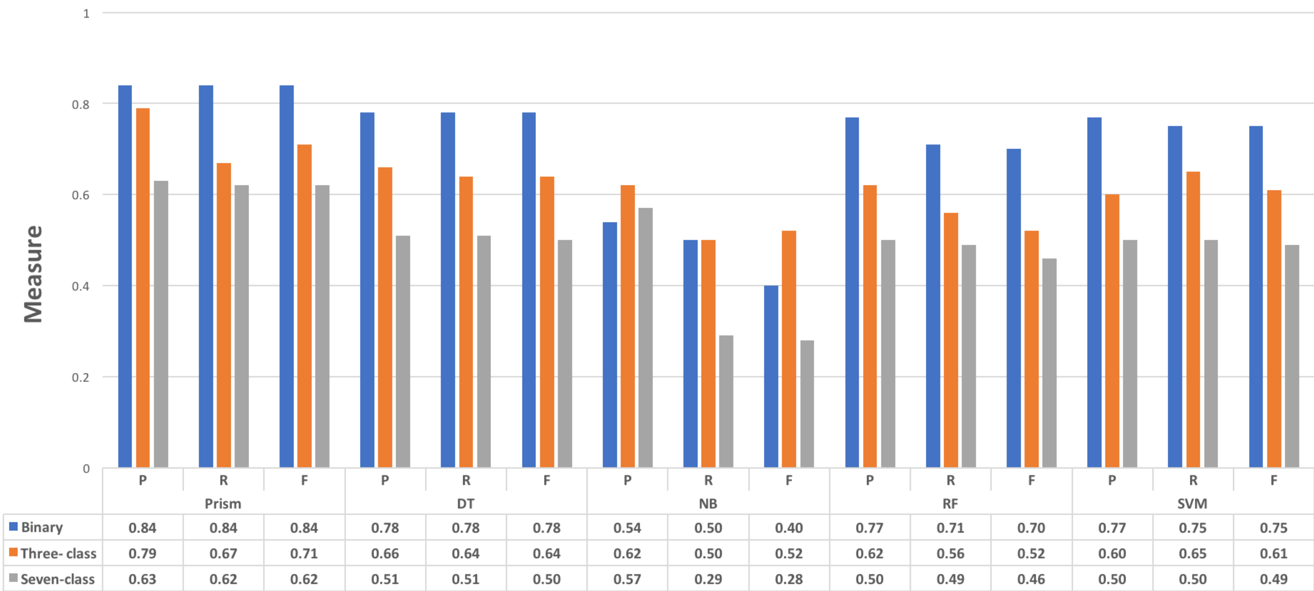

FIGURE 3. Machine Learning Algorithms Comparison

For the suicide and flippant classes, the Prism algorithm performs best, with an F-measure of 0.65 and 0.53, respectively. Similar to the three-class dataset, we notice that NB and RF have particularly low performance for the flippant class.

The overall results for each dataset are illustrated in Figure 3, which shows that the Prism algorithm outperforms the others on all measures, i.e. precision, recall and F-measure.

5 Discussion

Machine learning gained popularity for text classification in the 90s due to its ability to build automatic induction processes such as automatic text classifiers by learning from pre-classified data [21]. Although there are some related work carried out for text classification using machine learning, it is not conclusive on which machine learning algorithm works best with text classification specifically short and informal suicide-related text. Our experiment results indicate that less known algorithms, such as Prism, have the potential to perform well for this task; indeed, our results show that Prism outperformed four popular algorithms for text classification, i.e. Decision Tree, Naive Bayes, Random Forest and Support Vector Machine.

The above phenomenon is likely due to the nature of the Prism algorithm, which aims at learning specifically from instances of the target class to discriminate it from the other classes, unlike other algorithms, which aim at learning generally to discriminate between different classes without considering one class as the target class.

In this study, all the algorithms performed best on the binary dataset, except for NB whose best F-measure was obtained on the three-class dataset.

The Random forest algorithm, which uses an ensemble approach has the highest F-measure (0.69) in the three-class dataset for the campaign class. Interestingly, this was only marginally higher than the F-measure obtained by Prism (0.68).

6 Conclusion & Future Work

The application of machine learning techniques is becoming more popular and some of the techniques, both popular and less popular, have shown their different strengths when applied to short informal text. The aim of the study was to compare the classification performance of the Prism machine learning algorithm with more popular and reliable machine learning techniques.

The result of the study showed that the Prism algorithm has performed better that the other algorithms when applied to different datasets of short texts.

We will further extend this study by exploring the idea of multi-task learning which often gives better results, as it is considered to be more accurate and robust [3, 17, 22, 23] and other feature selection methods which also have benefits such as reduced dimensionality and improved classification performance [18–20, 24].

Acknowledgements

This is independent research commissioned and funded by the Department of Health Policy Research Programme (Understanding the Role of Social Media in the Aftermath of Youth Suicides, Project Number 023/0165). The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health. 9. The authors would also like to extend their gratitude to the Petroleum Technology Development Fund for their support.

References

[1] Y. Baştanlar and M. Özuysal, "Introduction to machine learning," in *miRNomics: MicroRNA Biology and Computational Analysis*. Springer, 2014, pp. 105–128.

[2] C. Robert, "Machine learning, a probabilistic perspective," 2014.

[3] P. Domingos, "A few useful things to know about machine learning," *Communications of the ACM*, vol. 55, no. 10, pp. 78–87, 2012.

[4] H. Liu, M. Cocea, A. Mohasseb, and M. Bader, "Transformation of discriminative single-task classification into generative multi-task classification in machine learning context," *9th International Conference on Advanced Computational Intelligence, ICACI 2017*, pp. 66–73, 2017.

[5] H. Liu, M. Cocea, and W. Ding, "Multi-task learning for intelligent data processing in granular computing context," *Granular Computing*, p. in press, 2017. [Online]. Available: http://link.springer.com/10.1007/s41066-017-0065-2

[6] J. Cendrowska, "PRISM: An algorithm for inducing modular rules," *International Journal of Man-Machine Studies*, vol. 27, no. 4, pp. 349–370, 1987.

[7] I. H. Witten, E. Frank, M. A. Hall, and C. J. Pal, *Data Mining: Practical machine learning tools and techniques*. Morgan Kaufmann, 2016.

[8] M. Aburrous, M. A. Hossain, K. Dahal, and F. Thabtah, "Intelligent phishing detection system for e-banking using fuzzy data mining," *Expert Systems with Applications*, vol. 37, no. 12, pp. 7913–7921, 2010.

[9] J. Fürnkranz, "Separate-and-conquer rule learning," *Artificial Intelligence Review*, vol. 13, no. 1, pp. 3–54, 1999.

[10] T. Trafalis and H. Ince, "Support vector machine for regression and applications to financial forecasting," *Proceedings of the IEEE-INNS-ENNS International Joint Conference on Neural Networks. IJCNN 2000. Neural Computing: New Challenges and Perspectives for the New Millennium*, no. May 2016, pp. 348–353 vol.6, 2000. [Online]. Available: http://ieeexplore.ieee.org/document/859420/

[11] J. A. Cruz and D. S. Wishart, "Applications of machine learning in cancer prediction and prognosis," *Cancer Informatics*, vol. 2, pp. 59–77, 2006.

[12] G. D. Magoulas and A. Prentza, *Machine Learning in Medical Applications*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2001, pp. 300–307.

[13] M. W. Libbrecht and W. S. Noble, "Machine learning applications in genetics and genomics," *Nature Reviews Genetics*, vol. 16, no. 6, pp. 321–332, 2015.

[14] S. C. Chen, M. L. Shyu, C. Zhang, L. Luo, and M. Chen, "Detection of soccer goal shots using joint multimedia features and classification rules." *4th International Workshop on Multimedia Data Mining (MDM/KDD2003)*, no. August 2014, pp. 36 – 44, 2003.

[15] M. Kubat, R. C. Holte, and S. Matwin, "Machine learning for the detection of oil spills in satellite radar images," *Machine Learning*, vol. 30, no. 2-3, pp. 195–215, 1998. [Online]. Available: http://www.scopus.com/inward/record.url?eid=2-s2.0-0031998121&partnerID=tZOtx3y1

[16] X. Zhang, Y. Yin, and H. Yu, "An application on text classification based on granular computing," *Communications of the IIMA*, vol. 7, no. 2, pp. 1–8, 2007.

[17] P. Burnap, W. Colombo, and J. Scourfield, "Machine Classification and Analysis of Suicide-Related Communication on Twitter," *Proceedings of the 26th ACM Conference on Hypertext & Social Media - HT '15*, pp. 75–84, 2015. [Online]. Available: http://dl.acm.org/citation.cfm?doid=2700171.2791023

[18] E. Haddi, X. Liu, and Y. Shi, "The role of text preprocessing in sentiment analysis," *Procedia Computer Science*, vol. 17, pp. 26–32, 2013.

[19] C. D. Manning, P. Ragahvan, and H. Schutze, *Introduction to Information Retrieval*. Cambridge University Press, 2009.

[20] Y. Liu, J. W. Bi, and Z. P. Fan, "Multi-class sentiment classification: The experimental comparisons of feature selection and machine learning algorithms," *Expert Systems with Applications*, vol. 80, pp. 323–339, 2017.

[21] F. Sebastiani, "Machine Learning in Automated Text Categorization," *ACM computing surveys (CSUR)*, vol. 34, no. 1, pp. 1–47, 2002.

[22] M. P. Ponti, "Combining classifiers: From the creation of ensembles to the decision fusion," *Proceedings - 24th SIBGRAPI Conference on Graphics, Patterns, and Images Tutorials, SIBGRAPI-T 2011*, no. Icmc, pp. 1–10, 2011.

[23] C. Poulin, B. Shiner, P. Thompson, L. Vepstas, Y. Young-Xu, B. Goertzel, B. Watts, L. Flashman, and T. McAllister, "Predicting the risk of suicide by analyzing the text of clinical notes," *PLoS ONE*, vol. 9, no. 1, pp. 1–7, 2014.

[24] M. Thelwall, K. Buckley, G. Paltoglou, D. Cai, and A. Kappas, "Sentiment strength detection in short informal text," *Journal of the Association for Information Science and Technology*, vol. 61, no. 12, pp. 2544–2558, 2010.