**Research Questionnaire**

**Note:**

* **Fill up *Table 2* for each paper. (COPY TABLE 2 AND PASTE AT THE END OF THIS FILE FOR NEXT PAPER)**
* ***Green* – Write few or more lines of required in your own words.**
* ***Red* – write down the list of what is required and description about each in the list**

|  |  |
| --- | --- |
| **TABLE 1** | |
| ***Reg. No. & Name*** | 17BCE2208- Varshitha Chennamsetti |
| ***Team No.*** | **18** |
| ***Paper Title*** | A Novel Stream Clustering Framework for Spam Detection in Twitter |
| ***Citation*** *(APA style)* | Tajalizadeh, H., & Boostani, R. (2019). A Novel Stream Clustering Framework for Spam Detection in Twitter. IEEE Transactions on Computational Social Systems, 6(3), 525-534. |

|  |  |
| --- | --- |
| **TABLE 2** | |
| **Problem answered in this paper.**  **(1-2 lines)** | We can use clustering as a method for spam filtering but the big microclusters do not always have a symmetric distribution. Therefore, to enhance the precision of the previous methods, this paper suggests replacing the Euclidean distance. |
| **Detailed description about the problem**  **(5-8 lines)** | Data Stream clustering is a technique used for spam filtering. But the big micro clusters sometimes have an asymmetric distribution. Even though the supervised techniques tend to give better results, the conventional techniques lack the flexibility and applicability. So in order to enhance the precision, there needs to be a replacement in the Euclidean distance with Incremental Naïve Bayes (INB) classifiers. |
| **Why that problem is chosen in this paper? Scope of the problem and solution** (*Refer Introduction*)  **(5-8 lines)** | There had been an increase in spam messages with the increase in the usage of social networks, especially twitter. Normally the spam messages can be detected by checking up on the members. If the members send messages to more people than the number of friends, they are generally considered spam. But there had been an increase of spammers who send limited amount of messages. Therefore, there needs to be a more flexible method than the conventional data stream clustering method to enhance the spam filtering process. This paper suggests to replace the Euclidean distance with the Incremental Naïve Bayes (INB) classifiers. |
| **History of the problem.** (*Refer Introduction*)  **(8-10 lines)** | Social Networks like twitter, facebook, instagram have gotten a lot of spam with the increase in the amount of usage of them. Most of the spam, especially on Twitter, is due to shortened URLs. Shortened URLs have become a problem because of the character limit on twitter and spammers take advantage of this. But not all accounts who have shortened URLs are spamming accounts. This is where the challenge is, to recognise the spammers. Recently, there have been many machine learning techniques which are both supervised and unsupervised. Even though better results can come through the supervised techniques, they lack the flexibility and applicability. This is the reason we need to make some changes to the conventional type of data stream clustering techniques. |
| **List of the related/similar problems** (*Refer Related work*) – Describe each with proposed solutions | |
| **Related problem 1** – Describe  **(3-4 lines)** | The paper talks about a novel approach to distinguish between the spammers and the non spammers. Rather than seeing the history of the users, they propose real time features in detecting spammers by taking the twitter account into consideration. They also concluded how the spammers evade the spam detection and also gave a data set that could help explain the spammer’s behavior. |
| **Paper in APA style** | Inuwa-Dutse, I., Liptrott, M., & Korkontzelos, I. (2018). Detection of spam-posting accounts on Twitter. *Neurocomputing*, *315*, 496-511. |
| **Related problem 2** – Describe  **(3-4 lines)** | This paper talks about using hashtags as a way to detect spammers. The spammers use the trending and popular tweets for their means. After conducting their research for 2 months on 14 million tweets, they created a dataset called HSpam14 which could be used for hashtag oriented spam research. |
| **Paper in APA style** | Sedhai, S., & Sun, A. (2015, August). Hspam14: A collection of 14 million tweets for hashtag-oriented spam research. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 223-232). ACM. |
| **Related problem 3** – Describe  **(3-4 lines)** | This paper talks about spammers who give fake reviews to either promote or demote product. It proposes a novel unsupervised approach called Author Spamicity Model (ASM) which focuses on the behavior of spammers. |
| **Paper in APA style** | Mukherjee, A., Kumar, A., Liu, B., Wang, J., Hsu, M., Castellanos, M., & Ghosh, R. (2013, August). Spotting opinion spammers using behavioral footprints. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining* (pp. 632-640). ACM. |
| **Related problem 4 –** Describe  **(3-4 lines)** | In the paper, they proposed to use WordVector in order to learn about the syntaxes in the spam tweets. They used different kinds of classifiers to achieve upto 30% of improvement in terms of F-measure. |
| **Paper in APA style** | Wu, T., Liu, S., Zhang, J., & Xiang, Y. (2017, January). Twitter spam detection based on deep learning. In *Proceedings of the australasian computer science week multiconference* (p. 3). ACM. |
| **Related problem 5 –** Describe  **(3-4 lines)** | This paper talks about the way the spammers evade the spam detection. They used the Crawl Twitter dataset to analyze the tricks by using classifiers like RF, DT etc. They found out that INB is the best classifier among all others. |
| **Paper in APA style** | Yang, C., Harkreader, R. C., & Gu, G. (2011, September). Die free or live hard? empirical evaluation and new design for fighting evolving twitter spammers. In *International Workshop on Recent Advances in Intrusion Detection* (pp. 318-337). Springer, Berlin, Heidelberg. |
| **What is the proposed solution in this paper for the problem chosen?** (*Refer Proposed work*)  **(5-8 lines)** | The CluStream assumes that the microclusters are too small and have symmetric distribution. It ignores the big microclusters. Therefore, they decided to consider the mean and boundary of the clusters to give more information. They suggested to replace the Euclidean distance with the set of INB classifiers. They used DenStream in their framework and called it as INB-DenStream. It is very similar to DenStream but has low complexity. |
| **Architecture of the proposed solution.** (*Refer proposed work*) **Diagram** |  |
| **Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)** | INB-DenStream |
| **List of existing algorithms used by the authors to complete the proposed work.**  **(1-2 lines for each algorithm)** | 1. DenStream Clustering   It is a density based streaming algorithm. It uses microclusters and outliers to analyze the data online quickly.  2. StreamKM++  It is based on core set idea and it uses randomised k-means++ algorithm.  3. Clustream  Clustream is an extension of BIRCH that works with the data streams that are changing. The interesting property of Clustream is that it works on the principle that the older clusters are deleted rather than the smaller ones. |
| **List of datasets used.** (*Refer experimental evaluation/result discussion*)  **(3-4 lines)** | They have used dataset that was already created. They used Dataset -I , -II , -III and -IV. The Datasets -I and -III contain the tweets that are selected randomly during a particular time continuously. The Datasets -II and -IV, on the other hand, are mutually exclusive tweets that are selected randomly. |
| **References/links to each of the dataset used in this paper (in APA style)** | 1. Chen, C., Zhang, J., Xie, Y., Xiang, Y., Zhou, W., Hassan, M. M., ... & Alrubaian, M. (2015). A performance evaluation of machine learning-based streaming spam tweets detection. *IEEE Transactions on Computational social systems*, *2*(3), 65-76. |
| **Why the above dataset(s) used?** (*Refer experimental evaluation/result discussion*)  **(3-4 lines)** | The four datasets used because they are very easy and simple to be dealt with. It is because the datasets were created by collecting the data through the twitter API and were retrieved in the form of JSON format. Therefore, it’s very easy to parse this dataset. |
| **List of equations that are very well applied in this problem domain** | Equation 1:    Description:  Each core microclusters is defined as a group of points like pi1,....pin and they have time stamps like Ti1,....,Tin. The weight(w), center(c) and radius(r) of these microclusters are defined using the formulas above. Here, when they talk about distance it is the Euclidean distance and the function is the fading function of the DenStream algorithm.  Equation 2:    Description:  P-microcluster is known as potential microclusters. They are the microclusters formed during the online phase of the DenStream method. The above equations are for finding the weighted linear and square summation of points and the weight of each microcluster. |
| **List of method(s)/metrics used to evaluate the proposed approach.** (*Refer experimental evaluation/result discussion*)  **(5-8 lines)** | 1. Purity   It checks the cluster quality.   1. Recall   It checks for the actual positives that it captured in all those that are labelled positive.   1. F1 measure   It is a measure to check the balance between precision and recall. They used specific version called F1-R. |
| **List of supporting tools/concepts**  **(3-4 lines)** | DenStream Clustering algorithm  INB classifiers  Infinite Latent Feature Selection (INFfs) |
| **What are the similar approaches with which the proposed approach is compared?**  (Refer experimental evaluation/result discussion)  **Explain each of these approach**  **(3-4 lines)** | Approach/method 1:  DenStream Clustering method  It is a density based streaming algorithm. It uses microclusters and outliers to analyze the data online quickly. It has two phases- the online phase and the offline phase. It also has a faded function that decreases the weight of each datapoint.  Approach/method 2:  StreamKM++  It is based on core set idea and it uses randomised k-means++ algorithm. This method also contains an offline phase and an online phase. In the online phase, it maintains the core sets by the core set tree bucket. In the offline phase, k-means++ is applied to the core set.  Approach/method 3:  Clustream is an extension of BIRCH that works with the data streams that are changing. The interesting property of Clustream is that it works on the principle that the older clusters are deleted rather than the smaller ones. |
| **How the results of proposed approach are compared with other similar approaches?** (*Refer experimental evaluation/result discussion*) | The DenStream Clustering method provided higher clustering purity compared to the method proposed because of the size of the microclusters. There seems to be undesirable results when the datasets are small. But when there are highly populated datasets, it was evident that the INB-DenStream method was much more accurate. DenStream, CluStream and StreamKM++ came in second, third and fourth place correspondingly in regards with F1-measure. There is also faster adaptation of the proposed method compared to others which eventually led to the most accurate results. |
| **Advantages/merits of proposed solution in your view.** (*Refer conclusion / result discussion / experimental evaluation*) | It was seen that INB-DenStream clustering will work better on highly populated datasets. Having temporal information also makes it easier for clustering through the method proposed. Using F1 measure, it could be seen that the time taken for clustering is lesser and the computational complexity is not high either. It is able to handle the streaming data in real time manner. This method takes in the cases where there is asymmetric distribution due to the size of big microcluster and also considers the boundary of the microclusters. Thus this method gives accurate clustering of data. |
| **Disadvantages/limitations of proposed solution in your view.** (*Refer conclusion / result discussion / experimental evaluation*) | There are undesirable results when the datasets are small and DenStream clustering seems to be having higher clustering than INB-DenStream method. Therefore, the results are not really convincing in the case of Dataset-II as it has less amount of data. |
| **Future work as stated by authors**  (*Refer conclusion / result discussion / experimental evaluation*) | The current clustering methods can be improved by an efficient real time Euclidean distance learning during the online phase. |
| **Your one page write-up about the paper** | |
| Due to the growing popularity of the microblogging sites like twitter, there have been more growth in spammers. Spammers on Twitter seem to be more dangerous than the mail spammers as they exploit the limitation on the characters of Twitter for their own purposes. Although there have been recent studies on machine learning for the supervised methods, there is little to none information regarding the unsupervised methods. This led to the need for a flexible method that takes in all the conventional methods’ properties but is more accurate.  The authors proposed INB-DenStream Clustering method as a way to make past this inaccuracy in the conventional methods. It acts similar to the DenStream method but the Euclidean distance used in the online phase with a set of INB classifiers. The main reason for replacing the Euclidean distance was to consider the microclusters that do not have symmetric distribution. The methods present at the moment only considered the mean which did not give a very accurate result. Through the method they proposed, they planned to take the mean as well as the boundary of the microclusters. Their method starts off similar as that of the DenStream Clustering method by calculating the Euclidean distance and classifying them into clusters. But by the second window of data, the population is checked with a minimum value called MinC and on excession, an INB classifier was assigned to that cluster. The INB classifier took the information like the mean and variance of the clusters. As the data comes in, the process is repeated until it exceeds a value called the SimThreshold and is assigned to the microcluster which has the INB with higher probability. The proposed method also ensures that the data does not take too much memory by only keeping important information. It also ensures adaptability by updating and retraining the INB classifiers as time goes by. This way the method ensures low computational complexity with low usage of memory.  The datasets used to test their method were Dataset-I, -II, -III and -IV. The Datasets -I and -III contain the tweets that were selected randomly during a particular time continuously. The Datasets -II and -IV, on the other hand, are mutually exclusive tweets that were selected randomly. They were also selected at the cost of losing temporal information. They used Purity, Recall and F1 measure as a way to test the methods. They compared their method with DenStream, StreamKM++ and Clustream clustering methods. It was evident that the INB-DenStream clustering method that they have proposed have come in first in F1-R measure. It has also been seen that the proposed method acted more robust than the DenStream clustering method in noisy environments. While checking the sensitivity of the method, the results were always either greater or similar to that of the DenStream clustering method. Although, there has been no evident success in the case of Dataset-II, it was mostly due to the small information regarding the cluster. There are undesirable results when the datasets are small and DenStream clustering seems to be having higher clustering than INB-DenStream method.  In conclusion, the method proposed has shown evident improvement to the other methods it was compared to. Although there are drawbacks to this method when the microcluster is very small, the improvements shown outshine these drawbacks. | |
| **Your findings: (possible alternate for the solution proposed)**   * Neural Network based approach can be used for spam detection. * DCStream clustering method can be used. * More importance to the Feature Selection based clustering can help give accurate results. | |

|  |  |
| --- | --- |
| **TABLE 1** | |
| ***Reg. No. & Name*** | 17BCE2208- Varshitha Chennamsetti |
| ***Team No.*** | **18** |
| ***Paper Title*** | A Neural Network-Based Ensemble Approach for Spam Detection in Twitter |
| ***Citation*** *(APA style)* | Madisetty, S., & Desarkar, M. S. (2018). A neural network-based ensemble approach for spam detection in Twitter. *IEEE Transactions on Computational Social Systems*, *5*(4), 973-984. |

|  |  |
| --- | --- |
| **TABLE 2** | |
| **Problem answered in this paper.**  **(1-2 lines)** | Detection of spam at tweet level through a robust approach that involves neural networks |
| **Detailed description about the problem**  **(5-8 lines)** | Spam tweets are more dangerous than the spam mail. The main attraction of these spam tweets is their hashtags. The spammers choose the most popular and trending hashtags as a way to lure the users. These tweets may lead the user to phishing or getting their personal details taken. Currently, users can report the spam tweet and the spam account but the spammers can simply create new accounts. This leads to a requirement for a tweet level spam detector. |
| **Why that problem is chosen in this paper? Scope of the problem and solution** (*Refer Introduction*)  **(5-8 lines)** | There has been spam for years at tweet level in twitter. They are really dangerous than the spam mail. Although, the twitter users can report spammers and spamming accounts, the spammers can continue their activity by creating new accounts. This where we need a tweet level spam detection. This paper proposed a combination of five CNNs model and one feature based model via neural networks regarding this problem. Here, neural networks work as a meta-classifier. |
| **History of the problem.** (*Refer Introduction*)  **(8-10 lines)** | Twitter is one of the most popular social networking sites that is used for microblogging. It is to no surprise that it became a target of the spammers. They use various methods to spam on twitter. Mainly they post URLs that either redirect the users to malicious pages or that get the users’ personal details. The main attraction for these spammers is the popular hashtags that are trending through which they gain the users’ attention. But Twitter detects spammers and blocks them. The users can also mark the tweets and spam accounts as spam. But the spammers can create new accounts to continue their spamming. Research and studies show that the spam at tweet level is more dangerous than the spam in the mails. But there has been less work related to detecting spam at tweet level which became a reason for the proposal of this paper. |
| **List of the related/similar problems** (*Refer Related work*) – Describe each with proposed solutions | |
| **Related problem 1** – Describe  **(3-4 lines)** | This paper talked about spam filtering at email level. They checked the four filters of the Naive Bayesian filtering technique. They analysed it based on its effectiveness and also suggest a derivative filter based on the number of tokens. |
| **Paper in APA style** | Androutsopoulos, I., Koutsias, J., Chandrinos, K. V., Paliouras, G., & Spyropoulos, C. D. (2000). An evaluation of naive bayesian anti-spam filtering. *arXiv preprint cs/0006013*. |
| **Related problem 2** – Describe  **(3-4 lines)** | This paper sees the classification of messages in three ways, as opposed to the binary classification- spam or non spam. The main advantage of this proposed method is that the users will get an accurate feedback about the messages. |
| **Paper in APA style** | Zhou, B., Yao, Y., & Luo, J. (2010, May). A three-way decision approach to email spam filtering. In *Canadian Conference on Artificial Intelligence* (pp. 28-39). Springer, Berlin, Heidelberg. |
| **Related problem 3** – Describe  **(3-4 lines)** | The paper talks about spam detection by analyzing the behaviour of the spammers. They conducted a study for seven months and found 36,000 spammers on twitter. They analyzed the behaviour of these spammers by checking the follower-followee relationship, link payloads etc. |
| **Paper in APA style** | Lee, K., Eoff, B. D., & Caverlee, J. (2011, July). Seven months with the devils: A long-term study of content polluters on twitter. In *Fifth International AAAI Conference on Weblogs and Social Media*. |
| **Related problem 4 –** Describe  **(3-4 lines)** | This paper talks about statistical analysis as a way to detect spam. The analysis sees only the content and not the user’s details in detecting spam. The analysis was taken on the dataset created by them through machine learning techniques on the popular hashtags. |
| **Paper in APA style** | Martinez-Romo, J., & Araujo, L. (2013). Detecting malicious tweets in trending topics using a statistical analysis of language. *Expert Systems with Applications*, *40*(8), 2992-3000. |
| **Related problem 5 –** Describe  **(3-4 lines)** | This paper talks about using hashtags as a way to detect spammers. The spammers use the trending and popular tweets for their means. After conducting their research for 2 months on 14 million tweets, they created a dataset called HSpam14 which could be used for hashtag oriented spam research. |
| **Paper in APA style** | Sedhai, S., & Sun, A. (2015, August). Hspam14: A collection of 14 million tweets for hashtag-oriented spam research. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 223-232). ACM. |
| **What is the proposed solution in this paper for the problem chosen?** (*Refer Proposed work*)  **(5-8 lines)** | They proposed a solution by creating a neural network algorithm, a feature based model and an ensemble. They used words as the feature representation of tweets. The CNNs use word embeddings in their proposed method. The feature based model used user-based, content-based, and n-gram features. The proposed ensemble method uses five CNNs model and one feature based model. They have also used a neural network based meta classifier for the binary classification. |
| **Architecture of the proposed solution.** (*Refer proposed work*) **Diagram** |  |
| **Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)** | There has been no particular name given to the approach they followed. The name would probably be 5CNN1FB as the method proposed contains 5 CNNs and one feature based model. |
| **List of existing algorithms used by the authors to complete the proposed work.**  **(1-2 lines for each algorithm)** | 1. Random forest algorithm   It is the most popular classification algorithm. It is also used for regression purposes. It works on the principle that the more number of trees, the more robust it is.   1. SVM algorithm   It is also a machine learning algorithm used for classification and regression. The main objective of this method is to find the hyperplane in N-dimensional features that classifies the data points distinctively.   1. Skip-gram method   In this method, we use a target word to find the context. Although this method is slower than CBOW, its performance is better. |
| **List of datasets used.** (*Refer experimental evaluation/result discussion*)  **(3-4 lines)** | They have used two datasets to check their proposed method. One of them is a subset of HSpam14 data set. They took 1 million tweets in the starting of this dataset. This subset is then split into a ratio of 2:1 and then classified either as spam or non spam. The second dataset 1KS10KN and it was not a balanced dataset. |
| **References/links to each of the dataset used in this paper (in APA style)** | 1. Sedhai, S., & Sun, A. (2015, August). Hspam14: A collection of 14 million tweets for hashtag-oriented spam research. In *Proceedings of the 38th International ACM SIGIR Conference on Research and Development in Information Retrieval* (pp. 223-232). ACM. 2. Wang, B., Zubiaga, A., Liakata, M., & Procter, R. (2015). Making the most of tweet-inherent features for social spam detection on twitter. *arXiv preprint arXiv:1503.07405*. |
| **Why the above dataset(s) used?** (*Refer experimental evaluation/result discussion*)  **(3-4 lines)** | The first dataset HSpam14 is used because each data was a tweet and it was already classified as either spam or non spam. The second dataset was used because it contained a big imbalance in the ratio of spam to non spam. The data was mostly non spam. |
| **List of equations that are very well applied in this problem domain** | Equation 1:    Description:  They considered words as the feature of tweets. So each tweet vector (Tv) is a concatenation of the word embeddings such as w1,w2,…..,wn.  Equation 2:    Description:  This is one of the features used in the feature based model. The reputation score can be seen as a way to detect spamming accounts. The spammers will have lesser reputation score than the normal users. |
| **List of method(s)/metrics used to evaluate the proposed approach.** (*Refer experimental evaluation/result discussion*)  **(5-8 lines)** | Accuracy, precision, recall, F-Measure, execution time, and area under curve (AUC) are the metrics that have been used by them. CNN+Twitter Glove, CNN+Google news, CNN+Edinburgh, CNN+H Spam are some of the methods used which have different dimensions of embeddings space. There is also feature based model, the proposed model and other models from reference works that have been used to compare the results. |
| **List of supporting tools/concepts**  **(3-4 lines)** | Random Forest Algorithm  SVM Algorithm  word2vec |
| **What are the similar approaches with which the proposed approach is compared?**  (Refer experimental evaluation/result discussion)  **Explain each of these approach**  **(3-4 lines)** | Approach/method 1:  SVM algorithm-  It is also a machine learning algorithm used for classification and regression. The main objective of this method is to find the hyperplane in N-dimensional features that classifies the data points distinctively.  Approach/method 2:  CNN+Twitter Glove, CNN+Google news, CNN+Edinburgh, CNN+H Spam-  These are the methods used to compare with the proposed method. They are a combination of CNN with Twitter Glove word embeddings, Google news corpus word2vec embeddings, Edinburgh Twitter corpus word2vec embeddings, HSpam14 Twitter corpus word2vec embeddings and random embeddings resspectively.  Approach/method 3:  Feature based model-  This method uses specific features to distinguish between the spammers and non spammers. It uses user-based, content-based, and n-gram features. |
| **How the results of proposed approach are compared with other similar approaches?** (*Refer experimental evaluation/result discussion*) | Although, the precision of the feature based method is lower than that of the deep learning techniques, the value is comparatively higher than that of the literature methods. The proposed ensemble method took in the CNNs and feature based method that contained accuracy, precision and F-Measure metrics with better performance. Since there are more features in the model, more execution time is taken. The proposed method gives better or similar performance with the baselines for both the balanced and the imbalanced dataset. |
| **Advantages/merits of proposed solution in your view.** (*Refer conclusion / result discussion / experimental evaluation*) | The proposed method gives better or similar performance with the baselines for both the datasets with a higher precision comparatively. Since the method proposed contains both the feature based and the deep learning methods, spammers cannot usually learn the patterns for the spam detection. So the method is robust. |
| **Disadvantages/limitations of proposed solution in your view.** (*Refer conclusion / result discussion / experimental evaluation*) | The precision of the feature based method is lower than that of the deep learning techniques in the case of HSpam14. The proposed method has to work with multiple stages and larger dimensions of the feature vectors. The execution time is also greater than all other methods as it considers all the individual methods in the ensemble. |
| **Future work as stated by authors**  (*Refer conclusion / result discussion / experimental evaluation*) | The author stated that further work should be done in regards to deep learning techniques. The performance of deep learning techniques could be improved by taking more information about the tweets or users. |
| **Your one page write-up about the paper** | |
| The paper talks about spam detection at tweet level. Although there have been ways to mark an account or tweet as spam, spammers could just recreate new accounts and pollute the content on twitter. So there is definite and urgent need for spam detection at tweet level. Spammers try to use trending and popular hashtags to lure the users. They use URLs along with these hashtags which redirect to malicious sites or get their personal details. This paper aims to give an efficient neural network based ensemble method to detect spam at tweet level.  They proposed neural network based ensemble method, deep learning and feature based techniques to detect spam at tweet level. They have considered word embeddings as the feature of tweets. Their proposed CNN method used various word embeddings with different dimensions. They have used the dense representation of the word vectors because of low computational speed and generalisation power. It is also preferred as there is a correlation between the features. There are five layers present in the CNN architecture proposed in this paper. They are input layer, convolution layer, pooling layer, hidden layer, and an output layer. The training of the word embeddings was done using the skip-gram method. The feature based model used user-based, content-based, and n-gram features. The ensemble method combined the five CNNs and one feature based model. Each CNN was trained with word embeddings like Twitter Glove, Google news corpus word2vec, Edinburgh Twitter corpus word2vec, HSpam14 Twitter corpus and random embeddings with different dimensions. The random forest algorithm is applied to the features in this ensemble method to detect spam.  The dataset used for evaluating this method are HSpam14 and 1KS10KN. While HSpam14 contained balanced ratio of spam to non-spam, 1KS10KN contained imbalanced ratio of tweets. It had more number of non-spam tweets. After calculating the metrics, it could be seen that the proposed ensemble method performed better than most of the literature methods. The precision of the proposed method was greater than that of the literature methods, mostly due to the features chosen by the authors proved to give more accurate results. There is a minute drawback in regards with the execution time taken by the proposed method. The proposed method takes more execution time than the other methods. It seemed to show better performance for a smaller and unbalanced dataset like 1KS10KN rather than a big and balanced dataset like HSpam14. But all in all, it shows better performance than all the base methods and has great robustness when spammers try to fool the algorithm.  This proposed method could be more efficient if there is a better feature representation. The data taken for the deep learning methods was only tweets with no additional information. The performance of the deep learning techniques could be better with additional information. | |
| **Your findings: (possible alternate for the solution proposed)**   * Semi Supervised Spam Detection (S3D) framework can be used to detect spam at tweet level. * Clustering method could be used to detect spam at tweet level. * More importance to deep learning techniques should be given by finding additional information. | |

|  |  |
| --- | --- |
| **TABLE 1** | |
| ***Reg. No. & Name*** | 17BCE2208- Varshitha Chennamsetti |
| ***Team No.*** | **18** |
| ***Paper Title*** | Few Are as Good as Many: An Ontology-Based Tweet Spam Detection Approach |
| ***Citation*** *(APA style)* | Halawi, B., Mourad, A., Otrok, H., & Damiani, E. (2018). Few are as good as many: an Ontology-based tweet spam detection approach. *IEEE Access*, *6*, 63890-63904. |

|  |  |
| --- | --- |
| **TABLE 2** | |
| **Problem answered in this paper.**  **(1-2 lines)** | There is a limitation on the information access. There needs to be a method to reduce the reliability on private and follower/followee relationship information in the detection of spam. |
| **Detailed description about the problem**  **(5-8 lines)** | There are many spam detection methods today but they all have limitations. They mostly use behavioural and statistical methods to detect spam. But this would mean we need to have the details like public information and the relationship information such as follower-followee ratio. The problem comes here because there is restricted access to such information and the cost for accessing it is high. This led to the need for a novel ontology approach that overcomes the limitations. |
| **Why that problem is chosen in this paper? Scope of the problem and solution** (*Refer Introduction*)  **(5-8 lines)** | The followers data is unavailable which makes the follower/followee method impossible to find out. Spammer’s are also learning about the different features to escape from the spam detection. There is also a restriction in using user’s information without their consent. Therefore, there is a need to find an accurate method to find spam in the public messages without using restricted information. They propose a method where the tweet messages are compared with ontologies in order to classify whether its spam or not. |
| **History of the problem.** (*Refer Introduction*)  **(8-10 lines)** | Twitter has become unreliable for the researchers as there has been an increase of spammers. There are also a lot of limitations while finding a solution to reduce spam content. There is a restricted access to the Twitter APIs and the metadata that the process becomes so expensive. The followers data is also unavailable which makes the follower/followee method impossible to find out. Spammer’s are also learning about the different features to escape from the spam detection. There is also a restriction in using user’s information without their consent. Therefore, there is a need to find an accurate method to find spam in the public messages. Although ontology methods have been used to find spam in long messages like mails, it has never been used in the case of tweets. |
| **List of the related/similar problems** (*Refer Related work*) – Describe each with proposed solutions | |
| **Related problem 1** – Describe  **(3-4 lines)** | In this paper, the statistical features are considered such that they change over time. Therefore, they proved that the machine learning classifiers are not that accurate to detect spammers. The paper then gave an alternative approach called the Lfun scheme where they analyzed the changed tweets and used them to create classifiers. |
| **Paper in APA style** | Chen, C., Wang, Y., Zhang, J., Xiang, Y., Zhou, W., & Min, G. (2016). Statistical features-based real-time detection of drifted Twitter spam. *IEEE Transactions on Information Forensics and Security*, *12*(4), 914-925. |
| **Related problem 2** – Describe  **(3-4 lines)** | This paper sees an importance in the URL. Spammers mostly use URLs to redirect the users and get their personal details. It is more dangerous than the spam mail. Therefore, they proposed a method called WarningBird for the URL detection in twitter which focuses on the URL redirects. |
| **Paper in APA style** | Lee, S., & Kim, J. (2013). Warningbird: A near real-time detection system for suspicious urls in twitter stream. *IEEE transactions on dependable and secure computing*, *10*(3), 183-195. |
| **Related problem 3** – Describe  **(3-4 lines)** | This paper mainly focuses on the user’s account information. They consider the follower-followee relationship and also the content of the account like the photo, biography etc. The behaviour of the account can seem suspicious if the follower-followee relationship is not balanced. |
| **Paper in APA style** | Sumner, C., Byers, A., Boochever, R., & Park, G. J. (2012, December). Predicting dark triad personality traits from twitter usage and a linguistic analysis of tweets. In *2012 11th International Conference on Machine Learning and Applications* (Vol. 2, pp. 386-393). IEEE. |
| **Related problem 4 –** Describe  **(3-4 lines)** | This paper talks about ontology based approach in detecting the spam mails. They have used two levels of ontology based spam filters- a global ontology filter and a customized ontology filter. The customized ontology filter was mostly based on the user’s background. |
| **Paper in APA style** | Youn, S. (2014). SPONGY (SPam ONtoloGY): Email classification using two-level dynamic ontology. *The Scientific World Journal*, *2014*. |
| **Related problem 5 –** Describe  **(3-4 lines)** | This paper talks about the statistical characters as a way to find the spammers. They collected the characteristics from 98 socialbots to analyse. They understood and analysed the profiles of the socialbots such as age, gender etc. |
| **Paper in APA style** | Fazil, M., & Abulaish, M. (2017, January). Why a socialbot is effective in Twitter? A statistical insight. In *2017 9th International Conference on Communication Systems and Networks (COMSNETS)* (pp. 564-569). IEEE. |
| **What is the proposed solution in this paper for the problem chosen?** (*Refer Proposed work*)  **(5-8 lines)** | The proposed method only focuses on the content of the tweets and not on the user’s details. They proposed a method where the tweet messages are checked with the ontologies to classify them into spam or non-spam tweets. They followed seven unidirectional ways to create the ontology. They fed some text to create the ontologies and make groups. They used only three main themes: sports, technology and politics. |
| **Architecture of the proposed solution.** (*Refer proposed work*) **Diagram** |  |
| **Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)** | Few are as good as many is the way they described their approach. |
| **List of existing algorithms used by the authors to complete the proposed work.**  **(1-2 lines for each algorithm)** | 1. NLTK model   It is a python package that contains the most commonly used NLP algorithms.   1. Cosine vector similarity   It is a metric to measure the similarity between two documents. It is advantageous as two documents can be similar even if they have high Euclidean distance.   1. Co-occurrence model   It checks to see how much the redundancy happens based on their mention in the same text. |
| **List of datasets used.** (*Refer experimental evaluation/result discussion*)  **(3-4 lines)** | The dataset used was raw and unstructured data from the date 05-2013 to 08-2013. They found these tweets from an online archive. They prepared the data to be tested from this raw data by themselves. They clustered the cleaned data into groups based on the hashtags. |
| **References/links to each of the dataset used in this paper (in APA style)** | 1.<https://archive.org/download/archiveteam-twitter-stream-2013-05> |
| **Why the above dataset(s) used?** (*Refer experimental evaluation/result discussion*)  **(3-4 lines)** | The four tweet files taken from archive.org are good to be structured into ontologies as it is raw and unstructured data. It is an ideal dataset to test for handling big data. They also contain data of different types of topics with different hashtags and different time zones. It is ideal for an ontology study. |
| **List of equations that are very well applied in this problem domain** | Equation 1:    Description:  The above formula is used for the probabilistic generation of ontologies. |
| **List of method(s)/metrics used to evaluate the proposed approach.** (*Refer experimental evaluation/result discussion*)  **(5-8 lines)** | They have used accuracy as a measure. They compared the ontologies with a random set of data which contain varied values of token similarity threshold. They have also found the efficiency of the spam detection by comparing them with a random set of data which contain varied values of token similarity threshold. This method will give an idea about the false positives and false negatives in the spam detection. |
| **List of supporting tools/concepts**  **(3-4 lines)** | text2onto  NLTK  Co-Occurrence Model  Cosine vector Similarity  Word Net |
| **What are the similar approaches with which the proposed approach is compared?**  (Refer experimental evaluation/result discussion)  **Explain each of these approach**  **(3-4 lines)** | Approach/method 1:  Follower-followee network analysis-  It is a comparison between the number of followers the user has and the number of people the user is following. This ratio can be seen as a way to detect spamming accounts. The spammers will have lesser ratio than the normal users.  Approach/method 2:  Message to message model-  It uses NLTK model, Cosine vector similarity and Co-occurrence model to analyse the given data. These models mostly evaluate the long texts such as email messages. But these statistical methods give inaccurate results. |
| **How the results of proposed approach are compared with other similar approaches?** (*Refer experimental evaluation/result discussion*) | The other methods require the information of the relationship between the users but the proposed method only concentrates on the content of the tweets. The threshold when is lower achieves more false positives. Different topics can produce different results of accuracy. The ontology based approach proposed by them outperformed all other message to message models present in today’s time. |
| **Advantages/merits of proposed solution in your view.** (*Refer conclusion / result discussion / experimental evaluation*) | The proposed method does not just rely on the private and relationship information for spam detection which major spam detection methods today require. It had better performance than most of the message to message techniques available right now. There is also a lot of reduction in false positives and false negatives. |
| **Disadvantages/limitations of proposed solution in your view.** (*Refer conclusion / result discussion / experimental evaluation*) | It was seen that different types of ontologies produce different results of accuracy. The lower the threshold, the larger the number of false positives are. |
| **Future work as stated by authors**  (*Refer conclusion / result discussion / experimental evaluation*) | The authors stated that the work on few are as good as many approach can be used for finding the data that is pretty accurate. |
| **Your one page write-up about the paper** | |
| The paper starts off talking about the limitations of today’s spam detection methods for twitter. The main limitation is that there is limited access to information such as the follower-followee relationship. There is a need for an approach that focuses on the content of the tweet rather than the user’s account information. Although there have been statistical approaches like the message to message model, they don’t produce accurate results. So there is a need for a novel approach that’s much more accurate.  This gave the need for the message to ontology evaluation approach. In this approach, first the data preparation is taken place. The data the authors decided to use is raw and unstructured data. It was selected from online archives. They were then divided into a structured format by filtering and cleaning. The redundant data and the unnecessary data like hyperlinks were deleted to get maximum accurate results. In the message to ontology evaluation approach, the messages are compared with the ontologies to search for the spam.They followed seven unidirectional ways to create the ontology. To get the ontologies, the model was filled with some themes. They have also included hashtags in detection of spam as it is one of the main attractions for the spammers. They use popular and trending hashtags to get the user’s attention. In the approach that they have suggested, they have included ontology extraction and ontology generation using probabilistic methods. The probabilistic values are added to increase the accuracy of the spam detection. In the generation of the ontology, they have used two algorithms- data driven discovery algorithm and natural language processing. The data driven discovery algorithm has been chosen as it is closely related to the context of the tweets. If any changes are made to these contexts, the change could be seen all through the ontology making it really flexible and accurate. They have used text2onto for the purpose of natural language processing. They have used it as it’s open source and it overcomes all the limitations that other models cannot overcome.  The data was divided mainly into themes such as sports, technology and politics. They have then conducted experiments on these different groups which have different values for token similarity threshold. They compared the ontologies with a random set of data which contain varied values of token similarity threshold. They have also found the accuracy and efficiency of the spam detection by comparing them with a random set of data which contain varied values of token similarity threshold. This method will give an idea about the false positives and false negatives in the spam detection.  After conducting these experiments, they’ve concluded on the experiments.They have seen that the lower thresholds have more chances of false positives. The different themes seemed to be showing different results of accuracy. Finally,It was evident that few are as good as many as the lower the similarities, the more accurate the ontologies become. | |
| **Your findings: (possible alternate for the solution proposed)**   * Neural network ensemble method can be used. * Semi Supervised Spam Detection (S3D) framework can be used to detect spam at tweet level * Clustering method can be used. | |