

## 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	
<i>Reg. No. &amp; Name</i>	18BCE2113
<i>Team No.</i>	<b>18</b>
<i>Paper Title</i>	Detection of spam-posting accounts on Twitter
<i>Citation (APA style)</i>	Inuwa-Dutse, I., Liptrott, M., & Korkontzelos, I. (2018). Detection of spam-posting accounts on Twitter. <i>Neurocomputing</i> , 315, 496-511.

TABLE 2	
<b>Problem answered in this paper.</b> (1-2 lines)	The rapid growth in the volume of global spam is compromising the validity of social media data which is being used in research works.
<b>Detailed description about the problem</b> (5-8 lines)	There are numerous researches which use social media data as their main source of information. Websites like Facebook and Twitter are doing such a great work to connect the whole world. Users of these platforms freely generate and consume information leading to unprecedented amounts of data. But the rapid increase in social media spam contents questions the credibility of research. Online spamming comes in different forms such as malware dissemination, abusive content, fake news, and generating fake product reviews. This makes it difficult to check the legitimacy of the contents being posted. Thus, spamming makes utilizing data directly from social media platforms untrustworthy and may mislead the analysis to wrong conclusions due to unrepresentative data.
<b>Why that problem is chosen in this paper? Scope of the</b>	The quality of research work is on the verge of deterioration due to the boom in spamming on Twitter. Both spamming bots and users are flooding the social media with fake news, malwares, etc. The Spam Post Detection model is proposed to identify the accounts posting these spam

<b>problem and solution</b> (Refer Introduction) (5-8 lines)	tweets by the help of Honeypot dataset. Scikit-learn toolkit is used to conduct numerous experiments with the use of features like UPF, AIF, EwF and EbF.
<b>History of the problem.</b> (Refer Introduction) (8-10 lines)	Social media is a major phenomenon in this technology-driven era. There are tons of social media sites enabling population from all over the globe to connect with each other. It's estimated that over 2 billion people would be connected by 2020. Users of these social media sites are always generating unprecedented amounts of data. Information derived from social media has been utilized in health-care to support effective service delivery, in sport to engage with fans, in politics to track election processes, promote wider engagement with supporters and predict poll outcomes. But the substantial increase in spam questions the credibility of this data. It is estimated that every 1 in 200 social media posts is spam and the numbers are increasing. The spamming contaminates the social media data and results in biased results.
<b>List of the related/similar problems</b> (Refer Related work) – Describe each with proposed solutions	
<b>Related problem 1</b> – Describe (3-4 lines)	With the increasing popularity of Twitter, every news and events are being discussed over Twitter and a huge interaction takes place on hot topics in no time. This is being used as opportunity by spamming bots and users to grab attention and get maximum hits. To tackle this situation, a real-time method is proposed which scans the tweets for malicious URLs using various APIs.
<b>Paper in APA style</b>	Kamble, S., & Sangve, S. M. (2018, August). Real Time Detection of Drifted Twitter Spam Based on Statistical Features. In 2018 International Conference on Information, Communication, Engineering and Technology (ICICET) (pp. 1-3). IEEE.
<b>Related problem 2</b> – Describe (3-4 lines)	The increasing rate of spams on Twitter is resulting in impact on the real world. The unequal distribution between the spam and non-spam class is the major consideration in the development of this model. A fuzzy-based oversampling method is used to generate synthetic data samples. An ensemble learning approach is developed which provides better spam detection rates with imbalanced class distribution.
<b>Paper in APA style</b>	Liu, S., Wang, Y., Zhang, J., Chen, C., & Xiang, Y. (2017). Addressing the class imbalance problem in twitter spam detection using ensemble learning. Computers & Security, 69, 35-49.
<b>Related problem 3</b> – Describe (3-4 lines)	All the ML methods of spam detection use statistical features, which changes over time and thus decreases accuracy of ML models. A Lfun scheme is proposed which analyses over a million of spam and non-spam tweets. This scheme can detect changed spam tweets and incorporate them into classifier's training program. This significantly improves spam detection accuracy.

<b>Paper in APA style</b>	Chen, C., Wang, Y., Zhang, J., Xiang, Y., Zhou, W., & Min, G. (2016). Statistical features-based real-time detection of drifted Twitter spam. <i>IEEE Transactions on Information Forensics and Security</i> , 12(4), 914-925.
<b>Related problem 4 – Describe</b> <b>(3-4 lines)</b>	The cybercrime rate increasing in the form of spamming on sites like Twitter is addressed in this paper. An inductive learning method is applied for detection on spammers by applying Random Forest approach on extracted features.
<b>Paper in APA style</b>	Meda, C., Bisio, F., Gastaldo, P., & Zunino, R. (2014, October). A machine learning approach for Twitter spammers detection. In <i>2014 International Carnahan Conference on Security Technology (ICCST)</i> (pp. 1-6). IEEE.
<b>Related problem 5 – Describe</b> <b>(3-4 lines)</b>	Twitter is one of the main targets of spammers. The spamming techniques are evolving over time. A new framework is proposed to deal with the new spamming method, spam drift. It uses unsupervised ML to retain a real -time supervised tweet-level spam detection model in batch mode. It adaptively discovers and learns the patterns of new spam activities. This proposed work reduces spam drift problems.
<b>Paper in APA style</b>	Washha, M., Qaroush, A., Mezghani, M., & Sedes, F. (2019). Unsupervised Collective-based Framework for Dynamic Retraining of Supervised Real-Time Spam Tweets Detection Model. <i>Expert Systems with Applications</i> .
<b>What is the proposed solution in this paper for the problem chosen?</b> <i>(Refer Proposed work)</i> <b>(5-8 lines)</b>	Several features are selected for experimentation to build classification model. The main features are UPF, AIF, EwF and EbF. The dataset of tweets is derived from Honeypot dataset and experiments are done to generate different classification models: MaxEnt, Random forest, ExtraTrees, SVC, MLP and SVM +MLP. All the models are trained and then evaluated on the SPD <sub>automated</sub> dataset using 10 -fold cross-validation. This approach doesn't require historical tweets and detects the Spam posting accounts in real time.
<b>Architecture of the proposed solution.</b> <i>(Refer proposed work)</i> <b>Diagram</b>	<pre> graph TD     A[LSA on public spam dataset] --&gt; B([Set of relevant concepts from LSA])     B --&gt; C[LSA concepts used for data collection via Twitter API]     C --&gt; D[(Raw queried tweets)]     D -- "Validate data" --&gt; E[Filter data to only keep tweets that contain at least two LSA concepts.]     E --&gt; F[Tag tweet as spam]     F --&gt; G[Available data] </pre>

<b>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</b>	Spam Post Detection (SPD)
<b>List of existing algorithms used by the authors to complete the proposed work.</b> (1-2 lines for each algorithm)	Latent Semantic Analysis (LSA) : It is a relationship analyzing technique between the dataset and its content by producing a set of concepts.
<b>List of datasets used.</b> ( <i>Refer experimental evaluation/result discussion</i> ) (1-4 lines)	<ol style="list-style-type: none"> <li>1. Honeypot: It is used to study spam activity on twitter as well as for collecting SPD dataset using keywords.</li> <li>2. SPD<sub>automated</sub> : The automatically annotated spam-posts detection dataset.</li> <li>3. SPD<sub>manual</sub> : The manually annotated spam-posts detection dataset.</li> </ol>
<b>References/links to each of the dataset used in this paper (in APA style)</b>	1. <u>Honeypot</u> : K. Lee , B.D. Eoff, J. Caverlee , Seven months with the devils: a long-term study of content polluters on twitter, in: Proceedings of the Fifth International Conference on Weblogs and Social Media, Barcelona, Catalonia, Spain, 2011, pp. 185–192 .
<b>Why the above dataset(s) used?</b> ( <i>Refer experimental evaluation/result discussion</i> ) (3-4 lines)	The honeypot dataset is used because it's publicly available and is useful for studying spam activity on Twitter. It's also used to collect SPD datasets using keyword. Keywords extracted from honeypot dataset are used to retrieve large quantities of similar data. The SPD <sub>automated</sub> and SPD <sub>manual</sub> datasets comprises of tweets selected manually and automatically respectively which is then compared to check the legitimacy.
<b>List of equations that are very well applied in this problem domain</b>	<p>Equation 1: <math>TTR = (\text{unique tokens in } D) / (\text{tokens in } D)</math>  Description: Type-Token ratio (TTR) measures the richness of a lexicon in a dataset, D.</p> <p>Equation 2: <math>LD = (\text{words in } D \text{ excluding stop words}) / (\text{tokens in } D)</math>  Description: Lexical Density (LD) measures the linguistic complexity from the functional and content words in the dataset, D.</p>
<b>List of method(s)/metrics used to evaluate the proposed approach.</b> ( <i>Refer experimental evaluation/result discussion</i> ) (5-8 lines)	<ol style="list-style-type: none"> <li>1. F-score: It's the measure of test's accuracy.</li> <li>2. Precision: The fraction of results classified as positive, which are indeed positive.</li> <li>3. Recall: The fraction of all positive results detected.</li> <li>4. Accuracy: The percentage of predictions that were correct.</li> </ol>
<b>List of supporting tools/concepts</b> (3-4 lines)	Scikit-learn toolkit

<p><b>What are the similar approaches with which the proposed approach is compared?</b> (Refer experimental evaluation/result discussion)</p> <p><b>Explain each of these approach (3-4 lines)</b></p>	<p>Approach/method 1: <u>Machine Learning Approach</u>: To automate the process of spam detection on Twitter, most methods use ML. The feature selection for classifiers determine the efficiency of these models.</p> <p>Approach/method 2: <u>Blacklist Approach</u>: Most spammers using embedded links in their tweets, this approach works by detecting URL containing tweets which rely on the third-party blacklisting techniques.</p> <p>Approach/method 3:</p>
<p><b>How the results of proposed approach are compared with other similar approaches?</b> (Refer experimental evaluation/result discussion)</p>	<p>SPD delivers better performance as compared to Honeypot on both the datasets. It delivers improved effectiveness and efficiency. The lightweight version of SPD performs better than Honeypot when it is applied on Automated dataset but it lacks behind Honeypot on the Manual dataset. The performance is much worse in the Lightweight version as compared to Full SPD (as expected).</p>
<p><b>Advantages/merits of proposed solution in your view.</b> (Refer conclusion / result discussion / experimental evaluation)</p>	<ol style="list-style-type: none"> <li>1. It detects spam on twitter in real-time.</li> <li>2. It delivers high effectiveness and efficiency.</li> <li>3. The method is cost-sensitive as it also uses Random Forest.</li> </ol>
<p><b>Disadvantages/limitations of proposed solution in your view.</b> (Refer conclusion / result discussion / experimental evaluation)</p>	<ol style="list-style-type: none"> <li>1. The proposed approach only considers tweets posted in English language.</li> <li>2. The irrelevant symbols increase the lexical richness.</li> <li>3. The emoticons are not discarded in the dataset, which in turn confuses the classifier.</li> </ol>
<p><b>Future work as stated by authors</b> (Refer conclusion / result discussion / experimental evaluation)</p>	<ol style="list-style-type: none"> <li>1. Effect of recent increase in maximum length of tweet: The increase in the maximum characters allowed in a single tweet will make it harder for the spamming bots to generate lengthier tweets.</li> <li>2. Spam user's interaction: It's observed that the spam users are quite selective with the accounts they follow on twitter and they might have some pattern which could make them easier to spot.</li> </ol>
<p><b>Your one-page write-up about the paper</b></p>	
<p>Social networking sites have become a part of ourselves over time. Twitter is one of the most popular social networking sites where millions of people connect with each other publicly to share data and news. It results in a magnificent data flow over the platform</p>	

which is being used for several intelligent systems such as twitter sentiment analysis and recommendation systems. Due to huge database, twitter has also become main target for spamming activities like phishing, spreading malware, etc. This not only raises security issues but also generate waste resources.

It's estimated that over 2 billion people would be connected by 2020. Users of these social media sites are always generating unprecedented amounts of data. Information derived from social media has been utilized in health-care to support effective service delivery, in sport to engage with fans, in politics to track election processes, promote wider engagement with supporters and predict poll outcomes.

The Spam Post Detection model is proposed to identify the accounts posting these spam tweets by the help of Honeypot dataset. Scikit-learn toolkit is used to conduct numerous experiments with the use of features like UPF, AIF, EwF and EbF.

Through the Honeypot dataset, two more datasets are created SPD<sub>manual</sub> and SPD<sub>automated</sub> to do experimentation.

Several features are selected for experimentation to build classification model. The main features are UPF, AIF, EwF and EbF. The dataset of tweets is derived from Honeypot dataset and experiments are done to generate different classification models: MaxEnt, Random forest, ExtraTrees, SVC, MLP and SVM +MLP. All the models are trained and then evaluated on the SPD<sub>automated</sub> dataset using 10 -fold cross-validation. This approach doesn't require historical tweets and detects the Spam posting accounts in real time.

SPD delivers better performance as compared to Honeypot on both the datasets. It delivers improved effectiveness and efficiency.

The lightweight version of SPD performs better than Honeypot when it is applied on Automated dataset but it lacks behind Honeypot on the Manual dataset. The performance is much worse in the Lightweight version as compared to Full SPD (as expected).

The proposed method utilized an optimized set of radially available features. The model showed great values on experimentation and delivered great performance results. The method can be applied into real-time filtering applications like data connection pipelines for filtering out the contents which are not relevant in the scope. The combination of handcrafted features and features learnt in an unsupervised manner using word embeddings is shown to significantly improve baseline performance and to perform comparably to the best performing feature set using a smaller number of features.

#### **Your findings: (possible alternate for the solution proposed)**

- Recall values can be further improved by using unsupervised collective-based framework.
- Oversampling method can be used to deal with unequal distribution of spam and non-spam dataset.
- The proposed model can also use graph-based features to make the model more precise.

## 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**

<b>Reg. No. &amp; Name</b>	18BCE2113, Shaurya Choudhary
<b>Team No.</b>	<b>18</b>
<b>Paper Title</b>	Unsupervised collective-based framework for dynamic retraining of supervised real-time spam tweets detection model
<b>Citation (APA style)</b>	Washha, M., Qaroush, A., Mezghani, M., & Sedes, F. (2019). Unsupervised Collective-based Framework for Dynamic Retraining of Supervised Real-Time Spam Tweets Detection Model. Expert Systems with Applications.

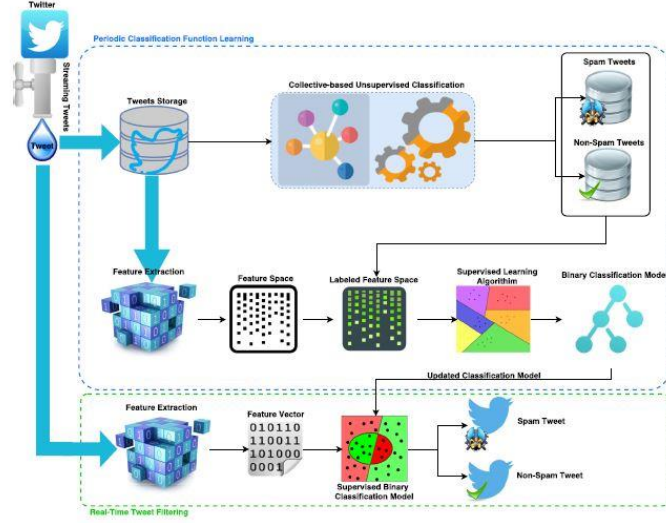
**TABLE 2**

<b>Problem answered in this paper.</b> (1-2 lines)	Real-time detection of spam tweets on Twitter by retraining a spam detection model on spam drift.
<b>Detailed description about the problem</b> (5-8 lines)	Social networking sites have become a part of ourselves over time. Twitter is one of the most popular social networking sites where millions of people connect with each other publicly to share data and news. It results in a magnificent data flow over the platform which is being used for several intelligent systems such as twitter sentiment analysis and recommendation systems. Due to huge database, twitter has also become main target for spamming activities like phishing, spreading malware, etc. This not only raises security issues but also generate waste resources.
<b>Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction)</b> (5-8 lines)	Twitter is one of the main targets of spammers. They are compromising the credibility of the information as well as threatening the interests of the user. The spamming techniques are evolving over time. A new framework is proposed to deal with the new spamming method, spam drift. It uses unsupervised ML to retain a real -time supervised tweet-level spam detection model in batch mode. It adaptively discovers and learns the patterns of new spam activities. This

	proposed work reduces spam drift problems.
<b>History of the problem.</b> ( <i>Refer Introduction</i> ) (8-10 lines)	Social media is a major phenomenon in this technology-driven era. There are tons of social media sites enabling population from all over the globe to connect with each other. It's estimated that over 2 billion people would be connected by 2020. Users of these social media sites are always generating unprecedented amounts of data. Information derived from social media has been utilized in health-care to support effective service delivery, in sport to engage with fans, in politics to track election processes, promote wider engagement with supporters and predict poll outcomes. But the substantial increase in spam questions the credibility of this data. It is estimated that every 1 in 200 social media posts is spam and the numbers are increasing. The spamming contaminates the social media data and results in biased results.
<b>List of the related/similar problems</b> ( <i>Refer Related work</i> ) – Describe each with proposed solutions	
<b>Related problem 1</b> – Describe (3-4 lines)	All the ML methods of spam detection use statistical features, which changes over time and thus decreases accuracy of ML models. A Lfun scheme is proposed which analyses over a million of spam and non-spam tweets. This scheme can detect changed spam tweets and incorporate them into classifier's training program. This significantly improves spam detection accuracy.
<b>Paper in APA style</b>	Chen, C., Wang, Y., Zhang, J., Xiang, Y., Zhou, W., & Min, G. (2016). Statistical features-based real-time detection of drifted Twitter spam. <i>IEEE Transactions on Information Forensics and Security</i> , 12(4), 914-925.
<b>Related problem 2</b> – Describe (3-4 lines)	Huge information available on Social Network sites draw attention of Cyber criminals. To prevent these crimes from microblogging platform like Twitter, spam filtering plays a major role. The proposed techniques use both (hybrid) content and graph-based features for identification of spammers.
<b>Paper in APA style</b>	Mateen, M., Iqbal, M. A., Aleem, M., & Islam, M. A. (2017, January). A hybrid approach for spam detection for Twitter. In 2017 14th International Bhurban Conference on Applied Sciences and Technology (IBCAST) (pp. 466-471). IEEE.
<b>Related problem 3</b> – Describe (3-4 lines)	The interest of a user is exploited by the spammers when they use platforms like Twitter to promote their websites and spread malware in unethical ways. Adaptive data classification is proposed by the use of commercial URL-based security tool and spam word lists. The dataset is analyzed by Naïve Bayes algorithm to detect spam.
<b>Paper in APA style</b>	Dangkesee, T., & Puntheeranurak, S. (2017, November). Adaptive Classification for Spam Detection on Twitter with Specific Data. In 2017 21st International Computer Science and Engineering Conference (ICSEC) (pp. 1-4). IEEE.



<b>Related problem 4 – Describe</b> <b>(3-4 lines)</b>	<p>The increasing rate of spams on Twitter is resulting in impact on the real world. The unequal distribution between the spam and non-spam class is the major consideration in the development of this model. A fuzzy-based oversampling method is used to generate synthetic data samples. An ensemble learning approach is developed which provides better spam detection rates with imbalanced class distribution.</p>
<b>Paper in APA style</b>	<p>Liu, S., Wang, Y., Zhang, J., Chen, C., &amp; Xiang, Y. (2017). Addressing the class imbalance problem in twitter spam detection using ensemble learning. <i>Computers &amp; Security</i>, 69, 35-49.</p>
<b>Related problem 5 – Describe</b> <b>(3-4 lines)</b>	<p>The quality of research work is on the verge of deterioration due to the boom in spamming on Twitter. Both spamming bots and users are flooding the social media with fake news, malwares, etc. The Spam Post Detection model is proposed to identify the accounts posting these spam tweets by the help of Honeypot dataset. Scikit-learn toolkit is used to conduct numerous experiments with the use of features like UPF, AIF, EwF and EbF.</p>
<b>Paper in APA style</b>	<p>Inuwa-Dutse, I., Liptrott, M., &amp; Korkontzelos, I. (2018). Detection of spam-posting accounts on Twitter. <i>Neurocomputing</i>, 315, 496-511.</p>
<b>What is the proposed solution in this paper for the problem chosen? (Refer Proposed work)</b> <b>(5-8 lines)</b>	<p>The model can be divided into 2 parts: (i) Real-time tweet filtering model, (ii) Periodic classification model learning. A design of online collective-based spam tweets classification framework is proposed which makes use of unsupervised ML to automatically provide latest datasets after annotation. This produces updated supervised classification models. The model then predicts the spamming behavior by correlating social spammers. The first part of the model develops feature vectors and stream them to the trained classification model to predict class label of the tweets. The second module stores these tweets and frequently creates a newly labeled training dataset using unsupervised ML. Finally, a classical supervised learning method is applied to the new labeled feature space to build a binary classification model to replace the current classifier model.</p>

<p><b>Architecture of the proposed solution.</b> (Refer proposed work) <b>Diagram</b></p>	 <p>The diagram illustrates the architecture of the proposed solution for spam tweet detection. It starts with a Twitter icon and a 'Stream Tweets' input. The data flows into a 'Tweets Storage' unit. From there, it branches into two main paths. The top path, labeled 'Periodic Classification Function Learning', involves 'Collective-based Unsupervised Classification' and 'Supervised Learning Algorithms' to create a 'Binary Classification Model'. The bottom path, labeled 'Real-Time Tweet Filtering', involves 'Feature Extraction' to create a 'Feature Vector' (e.g., 010110, 110011, 101000, 0001) which is then processed by a 'Supervised Binary Classification Model' to output 'Spam Tweet' and 'Non-Spam Tweet' results. An 'Updated Classification Model' is also shown, indicating a feedback loop for dynamic retraining.</p>
<p><b>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</b></p>	<p>Unsupervised collective based-framework (for dynamic retraining of real-time spam tweet detection model)</p>
<p><b>List of existing algorithms used by the authors to complete the proposed work.</b> (1-2 lines for each algorithm)</p>	<p>Unsupervised ML: Unsupervised learning is the training of machine using information that is neither classified nor labeled and allowing the algorithm to act on that information without guidance.</p>
<p><b>List of datasets used.</b> (Refer experimental evaluation/result discussion) (3-4 lines)</p>	<ol style="list-style-type: none"> <li>1. Ground-truth: A large dataset of tweets directly observed through different ways like manual inspection, clustering and blacklists.</li> </ol>
<p><b>References/links to each of the dataset used in this paper (in APA style)</b></p>	<ol style="list-style-type: none"> <li>1. Chen, Zhang, Chen et al., 2015; Hu et al., 2014; Hu et al., 2013; Sedhai &amp; Sun, 2017a; Thomas, Grier, Song et al., 2011; Wu, Liu et al., 2017</li> </ol>
<p><b>Why the above dataset(s) used?</b> (Refer experimental evaluation/result discussion)</p>	<p>A ground truth dataset is used because it is obtained from direct observation and the other datasets provided by other publishers only contain target object ID's which in most cases have already been suspended by Twitter. So, to have a large reliable dataset, different methods like</p>

<b>(3-4 lines)</b>	blacklists and clustering is used.
<b>List of equations that are very well applied in this problem domain</b>	Equation 1: $UA = (Time_{now} - Time_{creation}) / 864 * 10^5$ Description: It is used to determine the age of the twitter account.
<b>List of method(s)/metrics used to evaluate the proposed approach.</b> (Refer experimental evaluation/result discussion) <b>(5-8 lines)</b>	<ol style="list-style-type: none"> <li>1. Precision: The fraction of results classified as positive, which are indeed positive.</li> <li>2. Recall: The fraction of all positive results detected.</li> <li>3. F-measure: It's the measure of test's accuracy.</li> <li>4. Accuracy: The percentage of predictions that were correct.</li> </ol>
<b>List of supporting tools/concepts</b> <b>(3-4 lines)</b>	NMF: Non-negative Matrix Factorization is used as a tool for unsupervised method, to infer communities' structure due to its problem clustering capabilities. Weka Tool: It comprised of machine learning algorithms for data mining purposes.
<b>What are the similar approaches with which the proposed approach is compared?</b> (Refer experimental evaluation/result discussion) <b>Explain each of these approach</b> <b>(3-4 lines)</b>	<p>Approach/method 1: <u>Honeypot Approach</u>: It is a supervised learning method which requires an initial user-labelled dataset, it does not work efficiently with spam drift issues.</p> <p>Approach/method 2: <u>Machine Learning Approach</u>: To automate the process of pam detection on Twitter, most methods use ML. The feature selection for classifiers determine the efficiency of these models.</p> <p>Approach/method 3: <u>Blacklist Approach</u>: Most spammers using embedded links in their tweets, this approach works by detecting URL containing tweets which rely on the third-party blacklisting techniques.</p>
<b>How the results of proposed approach are compared with other similar approaches?</b> (Refer experimental evaluation/result discussion)	The proposed model produces very high recall values while giving low precision values. In contrary, the classical and asymmetric methods gave high precision value with low recall values. It is argued that recall value possess much more importance in spam tweet detection as the information wrongly marked are spam can be covered by another similar tweet on the subject but at the same time high recall values will provide high quality and relevant information.
<b>Advantages/merits of proposed solution in your view.</b> (Refer conclusion / result discussion / experimental evaluation)	<ol style="list-style-type: none"> <li>1. The proposed system detects spam in real time.</li> <li>2. It deals very efficiently with spam drift.</li> <li>3. It delivers high recall values.</li> </ol>
<b>Disadvantages/limitations of proposed solution in your view.</b> (Refer conclusion / result discussion / experimental	<ol style="list-style-type: none"> <li>1. Low precision value obtained by proposed system.</li> <li>2. The system did not address the growth of collected training dataset.</li> <li>3. Very old non-spam data is still stored, which cause waste of resources.</li> </ol>

evaluation)	
<b>Future work as stated by authors</b> <i>(Refer conclusion / result discussion / experimental evaluation)</i>	<ol style="list-style-type: none"> <li>1. Introducing other tweet content features.</li> <li>2. Study the effect of feature engineering methods.</li> <li>3. Testing other clustering methods.</li> <li>4. Reducing the effect of class imbalance dataset.</li> <li>5. Handling the growth of the collected training dataset.</li> </ol>
<b>Your one page write-up about the paper</b>	
<p>Social media is a major phenomenon in this technology-driven era. There are tons of social media sites enabling population from all over the globe to connect with each other. It's estimated that over 2 billion people would be connected by 2020. Users of these social media sites are always generating unprecedented amounts of data. Information derived from social media has been utilized in health-care to support effective service delivery, in sport to engage with fans, in politics to track election processes, promote wider engagement with supporters and predict poll outcomes. But the substantial increase in spam questions the credibility of this data. It is estimated that every 1 in 200 social media posts is spam and the numbers are increasing. The spamming contaminates the social media data and results in biased results.</p> <p>Websites like Facebook and Twitter are doing such a great work to connect the whole world. Users of these platforms freely generate and consume information leading to unprecedented amounts of data. But the rapid increase in social media spam contents questions the credibility of research. Online spamming comes in different forms such as malware dissemination, abusive content, fake news, and generating fake product reviews. This makes it difficult to check the legitimacy of the contents being posted. Thus, spamming makes utilizing data directly from social media platforms untrustworthy and may mislead the analysis to wrong conclusions due to unrepresentative data.</p> <p>The proposed model uses ground truth dataset which comprises of tweets directly observed through different ways like manual inspection, clustering and blacklists. The model can be divided into 2 parts: (i) Real-time tweet filtering model, (ii) Periodic classification model learning. A design of online collective-based spam tweets classification framework is proposed which makes use of unsupervised ML to automatically provide latest datasets after annotation. This produces updated supervised classification models. The model then predicts the spamming behavior by correlating social spammers.</p> <p>The first part of the model develops feature vectors and stream them to the trained classification model to predict class label of the tweets. The second module stores these tweets and frequently creates a newly labeled training dataset using unsupervised ML. Finally, a classical supervised learning method is applied to the new labeled feature space to build a binary classification model to replace the current classifier model.</p> <p>As spammers are becoming smarter with advancement in technology, it's getting tougher for the classical models to deal with the new Spam Drift patterns.</p>	

This model updates itself over time with latest datasets and has no limitations whatsoever. The only problem that is visible right now is that it keeps old non-spam data in its database which keeps on consuming storage resource and serve no purpose whatsoever.

**Your findings: (possible alternate for the solution proposed)**

- The proposed model can also use graph-based features to make the model more precise.
- Public Twitter APIs can be used for decreasing the processing time for certain types of tweets.
- SVM + MLP model can be used as input for training dataset.

## 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**

<b>Reg. No. &amp; Name</b>	18BCE2113
<b>Team No.</b>	<b>18</b>
<b>Paper Title</b>	An intelligent system for spam detection and identification of the most relevant features based on evolutionary Random Weight Networks
<b>Citation (APA style)</b>	Faris, H., Ala'M, A. Z., Heidari, A. A., Aljarah, I., Mafarja, M., Hassonah, M. A., & Fujita, H. (2019). An intelligent system for spam detection and identification of the most relevant features based on evolutionary random weight networks. Information Fusion, 48, 67-83.

**TABLE 2**

<b>Problem answered in this paper.</b> (1-2 lines)	Detection of spam emails using Genetic Algorithm and Random Weight Network.
<b>Detailed description about the problem</b> (5-8 lines)	With the boom in technology-driven era, email has become an inseparable part of billions of people around the globe. Every second, tons of data are flowing through it. This huge database of information makes it vulnerable as Cyber criminals get lured into it. The use of spam for exploiting user's interest is one of the most common way. The user's inbox is filled with junk messages, unethical promotions and malicious URLs.
<b>Why that problem is chosen in this paper? Scope of the problem and solution (Refer Introduction)</b> (5-8 lines)	The email has become an essential and popular platform for communication. It is used for both official and unofficial purposes. As a result, it's being targeted by spammers and users are exposed to security threats. An adaptable spam detection model based on GA-RWN is proposed for this issue with automatic identification capability. The system is then evaluated on several email corpora in such a way that it can identify the most relevant features of spam email. An

	automatic email spam detection in obtained after experimentation.
<b>History of the problem.</b> ( <i>Refer Introduction</i> ) (8-10 lines)	Email communication is now being used more than ever. It's prevalent and indispensable nowadays. Every second, tons of data are flowing through it. This huge database of information makes it vulnerable as Cyber criminals get lured into it. The threat of spamming is getting more serious. A survey revealed that 40% of emails were spam in 2006 and recently it has reached as high as 70%. The spam drift is making the problem even severe. Spammers are using different features for their spam messages and are evolving over time. Spamming is usually done with similar content and in large quantities. This makes filtering comparatively easy for the most part. The spam messages waste the valuable resources, including storage, bandwidth, and productivity.
<b>List of the related/similar problems</b> ( <i>Refer Related work</i> ) – Describe each with proposed solutions	
<b>Related problem 1</b> – Describe (3-4 lines)	Two ML techniques: ELM and SVM are widely used for classification problems. These two methods are compared with each other on the problem of email spam detection which uses effective classifiers. Using a popular email corpus for experimentation it is found that ELM is a lot faster while other hand SVM has more accuracy.
<b>Paper in APA style</b>	Olatunji, S. O. (2017, April). Extreme Learning machines and Support Vector Machines models for email spam detection. In 2017 IEEE 30th Canadian Conference on Electrical and Computer Engineering (CCECE) (pp. 1-6). IEEE.
<b>Related problem 2</b> – Describe (3-4 lines)	The main problem taken into consideration is the spam emails sent by botnets and to find its source and architecture. A detailed timeline of botnets is presented which shows events in their advancement over time. The botnets are categorized on basis of their nature of defense and method of detection. In this way a model is developed to detect spamming botnets.
<b>Paper in APA style</b>	Khan, W. Z., Khan, M. K., Muhaya, F. T. B., Aalsalem, M. Y., & Chao, H. C. (2015). A comprehensive study of email spam botnet detection. IEEE Communications Surveys & Tutorials, 17(4), 2271-2295.
<b>Related problem 3</b> – Describe (3-4 lines)	The proposed method is inspired from the near-duplicate similarity matching scheme where the emails are identified as spam through user feedback and then similar emails are identified as spam via matching. The Cosdes system uses novel email abstraction using HTML content which enables it to efficiently perform near-duplicate matching with progressive update content to cope up with the new spamming methods.
<b>Paper in APA style</b>	Tseng, C. Y., Sung, P. C., & Chen, M. S. (2010). Cosdes: a collaborative spam detection system with a novel e-mail abstraction scheme. IEEE transactions on knowledge and data engineering,



	23(5), 669-682.
<b>Related problem 4</b> – Describe (3-4 lines)	A model based on Genetic Programming combined with SMOTE is taken in used for the spam email detection. This method giver superior results than usual classifiers method by considering 4 main measures: accuracy, recall, precision and G-mean. This provides more effectiveness for spam detection on emailing platforms.
<b>Paper in APA style</b>	Habib, M., Faris, H., Hassonah, M. A., Sheta, A. F., & Ala'M, A. Z. (2018, November). Automatic Email Spam Detection using Genetic Programming with SMOTE. In 2018 Fifth HCT Information Technology Trends (ITT) (pp. 185-190). IEEE.
<b>Related problem 5</b> – Describe (3-4 lines)	Around 140 features are extracted from SpamAssassin dataset using developed tool to train the Machine Learning model for email spam detection. Extracted features provide better results than previous studies due to better evaluation of ML classifiers and thus provide improved results.
<b>Paper in APA style</b>	Faris, H., Ala'M, A. Z., & Aljarah, I. (2017, October). Improving email spam detection using content based feature engineering approach. In 2017 IEEE Jordan Conference on Applied Electrical Engineering and Computing Technologies (AEECT) (pp. 1-6). IEEE.
<b>What is the proposed solution in this paper for the problem chosen?</b> (Refer Proposed work) (5-8 lines)	Auto-GA-RWN system is applied in this work. It's a 5 stages method comprising of Feature extraction, FS, model development, evaluation & assessment, and feature importance analysis. Three email datasets and constructed in the first step followed by removal of irrelevant features. Then the predictive powers of the RWN network is evaluated. Further analysis of done to identify more influencing features. Feature selection, Auto-tuning of hidden neurons, and evolution of model, all tasks take place simultaneously. The model identifies the most relevant features and optimizes the config. of its core classifiers. The model then detects the spam emails based on its RWN.
<b>Architecture of the proposed solution.</b> (Refer proposed work) <b>Diagram</b>	<p>The diagram illustrates the architecture of the proposed Auto-GA-RWN system. It starts with a 'Training dataset' represented as a blue grid. This dataset is processed by a 'Reduce features' block, which outputs a 'Reduced dataset used to train RWN' shown as a yellow grid. The 'Reduced dataset' is then used for 'Decoding' to generate the 'Neurons part', represented by green boxes labeled <math>n_1, n_2, \dots, n_k</math>. The 'Features part', represented by blue boxes labeled <math>f_1, f_2, \dots, f_d</math>, is also used for 'Decoding'. The 'Reduced dataset' and the 'Neurons part' are used to 'Build RWN based on the decoded number of hidden neurons', which results in a neural network diagram with multiple layers of nodes.</p>



<b>Name of the approach as stated by the authors (if not, you try to give a name based on the concepts used)</b>	Auto-GA-RWN
<b>List of existing algorithms used by the authors to complete the proposed work.</b> (1-2 lines for each algorithm)	GA algorithm: The algorithm is used to generate high-quality solutions by relying on bio-inspired operators. The best chromosomes with highest fitness value are selected as classifier.
<b>List of datasets used.</b> (Refer experimental evaluation/result discussion) (3-4 lines)	<ol style="list-style-type: none"> <li>1. SpamAssassin: It's an open source anti-spam dataset which enables admins to classify and block emails.</li> <li>2. LingSpam: This dataset contains spam messages from Linguist list.</li> <li>3. CSDMC2010 Corpus: This dataset comprises of selected emails as training/test data for spam detection.</li> </ol>
<b>References/links to each of the dataset used in this paper (in APA style)</b>	<ol style="list-style-type: none"> <li>1. <a href="http://spamassassin.org/publiccorpus/">http://spamassassin.org/publiccorpus/</a></li> <li>2. <a href="http://www.aueb.gr/users/ion/data/">http://www.aueb.gr/users/ion/data/</a></li> <li>3. <a href="http://www.aueb.gr/users/ion/data/">http://www.aueb.gr/users/ion/data/</a></li> </ol>
<b>Why the above dataset(s) used?</b> (Refer experimental evaluation/result discussion) (3-4 lines)	The above datasets are used due to their availability as open source as well as their specific function which prove quite useful in managing email spams. The above datasets are precisely made for training purposes of ML models.
<b>List of equations that are very well applied in this problem domain</b>	<p>Equation 1: <math>\text{Fitness} = \alpha \text{Err} + \beta(f / F) + \gamma(n / N)</math>  Description: A fitness function is needed in metaheuristic algorithms to assess the quality of the generated possible solutions.</p> <p>Equation 2: <math>\text{Accuracy} = (\text{TP} + \text{TN}) / (\text{TP} + \text{FN} + \text{FP} + \text{TN})</math>  Description: The fraction of predictions that were correct.</p> <p>Equation 3: <math>G - \text{Mean} = (\text{Recall}_S \times \text{Recall}_H)^{1/2}</math>  Description: The procedure of the proposed Auto-GA-RWN algorithm.</p> <p>Equation 4: <math>\text{rel}(x_i, B) = \text{RefCount}(x_i, B) / B</math>  Description: Identification of most influencing features in the set of best models, B.</p>
<b>List of method(s)/metrics used to evaluate the proposed approach.</b> (Refer experimental evaluation/result discussion)	<ol style="list-style-type: none"> <li>1. Precision: The fraction of results classified as positive, which are indeed positive.</li> <li>2. Recall: The fraction of all positive results detected.</li> <li>3. Accuracy: The percentage of predictions that were correct.</li> </ol>

<b>(5-8 lines)</b>	
<b>List of supporting tools/concepts</b> <b>(3-4 lines)</b>	EMFET: E-mail Features Extraction Tool is utilized to convert email messages into features that can be processed by machine algorithms.
<b>What are the similar approaches with which the proposed approach is compared?</b> (Refer experimental evaluation/result discussion) <b>Explain each of these approach</b> <b>(3-4 lines)</b>	<p>Approach/method 1: SVM: It is a supervised ML algorithm which is used for classification problems. The classification is done by finding the hyper plane that differentiate the two classes very well where each data point is plotted as a point in n-dimensional space with value of each feature as value of the coordinate.</p> <p>Approach/method 2: Naive Bayes: It's a classification algorithm which works on Bayes theorem of probability and predict the class of unknown datasets.</p> <p>Approach/method 3: Classic GA: The algorithm is used to generate high-quality solutions by relying on bio-inspired operators. The best chromosomes with highest fitness value are selected as classifier.</p>
<b>How the results of proposed approach are compared with other similar approaches?</b> <i>(Refer experimental evaluation/result discussion)</i>	The proposed model shows more accuracy than SVM, Naive Bayes and nearest neighbor. On the other hand, SVM delivered slightly better Recall <sub>n</sub> and Precision <sub>s</sub> value with RWN as second. While RWN gave highest value for Recall <sub>s</sub> and Precision <sub>n</sub> The RWN model gave highest G-mean value. The performance varied slightly on different datasets but overall the Auto-GA-RWN gave the most promising results for spam email detection.
<b>Advantages/merits of proposed solution in your view.</b> <i>(Refer conclusion / result discussion / experimental evaluation)</i>	<ol style="list-style-type: none"> <li>1. Optimizes the core classifier with relevant features.</li> <li>2. Give better accuracy and recall values than similar methods.</li> <li>3. The classification process is automated.</li> </ol>
<b>Disadvantages/limitations of proposed solution in your view.</b> <i>(Refer conclusion / result discussion / experimental evaluation)</i>	<ol style="list-style-type: none"> <li>1. Unable to deal with imbalanced dataset classification.</li> <li>2. It's not as efficient as other spam detecting models.</li> <li>3. The system did not address the issue of growth of collected dataset.</li> </ol>
<b>Future work as stated by authors</b> <i>(Refer conclusion / result discussion / experimental evaluation)</i>	<ol style="list-style-type: none"> <li>1. Further research can be done on other solutions for the imbalanced classification tasks.</li> <li>2. Impact of the solutions on relevance of input features.</li> </ol>
<b>Your one page write-up about the paper</b>	

Email communication is now being used more than ever. It's prevalent and indispensable nowadays. Every second, tons of data are flowing through it. This huge database of information makes it vulnerable as Cyber criminals get lured into it. The threat of spamming is getting more serious. A survey revealed that 40% of emails were spam in 2006 and recently it has reached as high as 70%. The spam drift is making the problem even severe. Spammers are using different features for their spam messages and are evolving over time. Spamming is usually done with similar content and in large quantities. This makes filtering comparatively easy for the most part. The spam messages waste the valuable resources, including storage, bandwidth, and productivity.

As a result, it's being targeted by spammers and users are exposed to security threats. An adaptable spam detection model based on GA-RWN is proposed for this issue with automatic identification capability. The system is then evaluated on several email corpora in such a way that it can identify the most relevant features of spam email. An automatic email spam detection is obtained after experimentation.

Three datasets namely SpamAssassin, CSDMC2010, and LingSpam are used for the proposed model experimentation.

The proposed method can be divided into stages:

1. Feature Extraction: The three datasets are constructed based on SpamAssassin, CSDMC2010, and LingSpam. Then EMFET is used to convert these email corpuses into feature sets.
2. Feature Selection: False Spam method is executed in this step on the training part of the dataset to eliminate features which are not relevant.
3. Evaluation and Assessment: The RWN network is tested of its predictive powers in this method. The analysis is done on the matrices like Precision, accuracy, and recall.
4. Feature Importance Analysis: Further analysis is done to identify the most influencing features in the dataset. It helps in designing more accurate spam filters.

Feature selection, Auto-tuning of hidden neurons, and evolution of model, all tasks take place simultaneously. The model identifies the most relevant features and optimizes the config. of its core classifiers. The model then detects the spam emails based on its RWN. The proposed model shows more accuracy than SVM, Naive Bayes and nearest neighbor. On the other hand, SVM delivered slightly better Recall and Precision value with RWN as second. While RWN gave highest value for Recall and Precision. The RWN model gave highest G-mean value. The performance varied slightly on different datasets but overall the Auto-GA-RWN gave the most promising results for spam email detection.

The Auto-GA-RWN method is evaluated to find out that it can hit very promising figures and it is capable of updating its own classifier over time with most relevant features. The proposed model is very capable with very few limitations and is very application oriented in detection on spam emails over internet.

**Your findings: (possible alternate for the solution proposed)**

- SMOTE can be combined with the proposed model to make the results even better.
- Special features can be selected to identify the spamming bots and their source.
- ELM method can also be used to trade speed for a little accuracy.