B.Tech Project Report

IT-414

On

Image Based Spam Detection

In Online Social Networks (OSNs)

By

Vikash Chauhan (11510535)

Ankit Saini (11510520)

Deepak (11510519)

Under the Supervision of

Dr. B. B. GUPTA

DEPARTMENT OF COMPUTER ENGINEERING

NATIONAL INSTITUTE OF TECHNOLOGY

KURUKSHETRA - 136119, HARYANA (INDIA)

July 2018 – December 2018

CERTIFICATE

We hereby certify that the work which is being presented in this B.Tech. Minor Project (IT413) report entitled "**Image Based Spam Detection in Online Social Networks**", in partial fulfillment of the requirements for the award of the **Bachelor of Technology in Information Technology** is an authentic record of our own work carried out during a period from December 2017 to April 2018 under the supervision of **Dr. B. B. Gupta**, Assistant Professor, Computer Engineering Department.

The matter presented in this project report has not been submitted for the award of any other degree elsewhere.

Signature of Candidate

Vikash Chauhan (11510535)

Ankit Saini (11510520)

Deepak (11510519)

This is to certify that the above statement made by the candidates is correct to the best of my knowledge.

Date:

Signature of Supervisor

Dr. B. B. Gupta

Assistant Professor

TABLE OF CONTENT

LIST OF FIGURES

Abstract

Since, in this era of internet various kinds of Social Media are present like Facebook, Twitter, Whatsapp etc. Since, this mainstream media contribute a lot in sharing multimedia text and images. Text can be normal or as well as the spam but these textual approaches were used earlier for capturing sensitive data of users but at the present generation the scenario has changed drastically and now images are used for spreading the spam messages. Images can be of two types they can either be spam or ham. Spam images are global threat to the internet users at this present scenario because they not only captures the undue attention of people but also poses challenges in the form of theft of confidential information by redirecting the users to the other pages and then collecting their sensitive information. So, there is an urgent need to prevent those spam images otherwise the confidence of people will be lost in sharing those multimedia thorough the various social networking platforms. In this paper we provided in efficient method of detecting spam/ham images. Detection of spam images or the normal images can only be possible after extracting the features of images like compression ratio, total pixel, bit depth, height, width etc. A spam image consists of text in a greater percentage or some emoji whereas these are absent in ham image. Hence clear cut features are there which are used to classify that whether an image is spam or not. Moreover the accuracy of the classifier is also improved by choosing an optimal set of features which are given as a result of Genetic Algorithm. Different classifiers results are also compared. Once it is detected with a decent accuracy rate that a particular image is spam then preventive schemes are applied in order to develop a complete system which will help the users to get rid of spam images.

1 Introduction

We are living in the era of Internet; we can easily connect to different persons living in different countries at the same instance of time. Earlier modes of communication were the letters later telephone came into existence and then fax machine but in the recent time with the rise of internet electronic mail system had been developed and this email system is currently the most popular way of communicating each other. These communications brought the peoples closer to each other. Our geographical and economical borders are shrinking day by day. This is possible because of online social network (OSNs).Since there are a lot of options in OSN like Whatsapp, Facebook, and Twitter which contains images so it is obvious that some of them are spams. At the present scenario Facebook monthly active users number is around 2 billion [1] so it is quite obvious that such a large collection of personal data of users is continuously being under threat from the hackers as well as the false companies policies. An attacker uses the spam as a primary method of inculcating the private details of peoples. The first spam which ever recorded were email spams which contains text as their spammer contents and these spammer contents contains text like winning lottery notifications and notifications of the type that your system is under threat and you should install a particular software in order to get rid of that particular threat. But the main area of working of spammers comes after the users click that particular link that is they install some Trojan horses, adware or viruses to our systems and the user does not know about them. Now a days there is a lot of work on classifying the text as spam or not but work on detecting spam on images is quite limited and hence, as a result many different types of spams are coming day by day. Spam causes a lot of harm to the clients who are using the internet for browsing the online social networking sites. Spam in not the only threat in OSN many other vulnerabilities are also present and some of these includes malware, social bots etc. Due to spams either the confidential information of the people is lost by redirecting them to the other links and then stealing their information or the valuable time span of people is lost since some time is wasted on observing the advertisements. So it should be of utmost priority to detect and prevent spams. Image spammers are the messages which are sent in large batches and each batch contains the visually same message and this is one of the reasons that why detecting a spam image with a decent accuracy rate is still difficult. Earlier Image Spammers like honey-pots can detect some of the spam images but these are not reliable.

Now in order to develop an OSN image based spam detection system it should satisfy the certain requirements: extensibility, high accuracy rate and high efficiency. Earlier, spammers like honeypot work on low level features of images and these low level features include shape, texture and color. The limited number of features which are going to be analyzed in OSN spam detection is the foremost and only reason for its lower accuracy rate. Later OCR technique is used in addition with the features of images so that we can improve the efficiency and spam assassin [2] is the best example to prove that efficiency of detecting spam images has been increased to a far greater extent as compared to the previous approaches of extracting only limited number of features.

Since lot of multimedia content is available and for their protection our researchers has already developed solutions which includes digital oblivion, steganalysis and water marking.

1.1The System Architecture

Architecture of email based spam detection system:

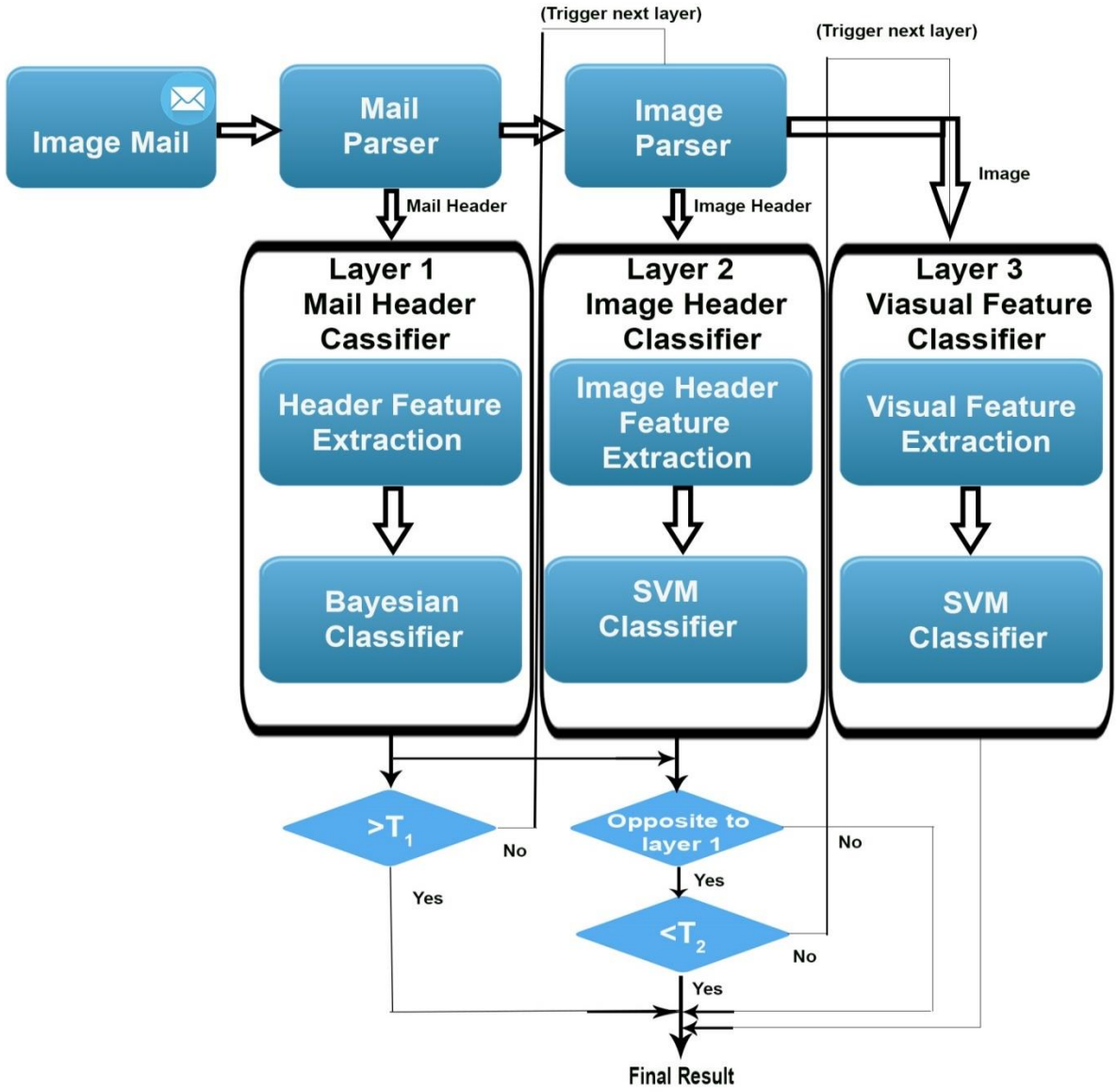

Figure 1 Previous architecture of email based spam detection

Image based spam detection is simply an advancement of email based spam detection system. Different image spam techniques have its own strength and weaknesses but at the present scenario a multilayer image spam filtering system [3] is used which is shown below in figure 1.

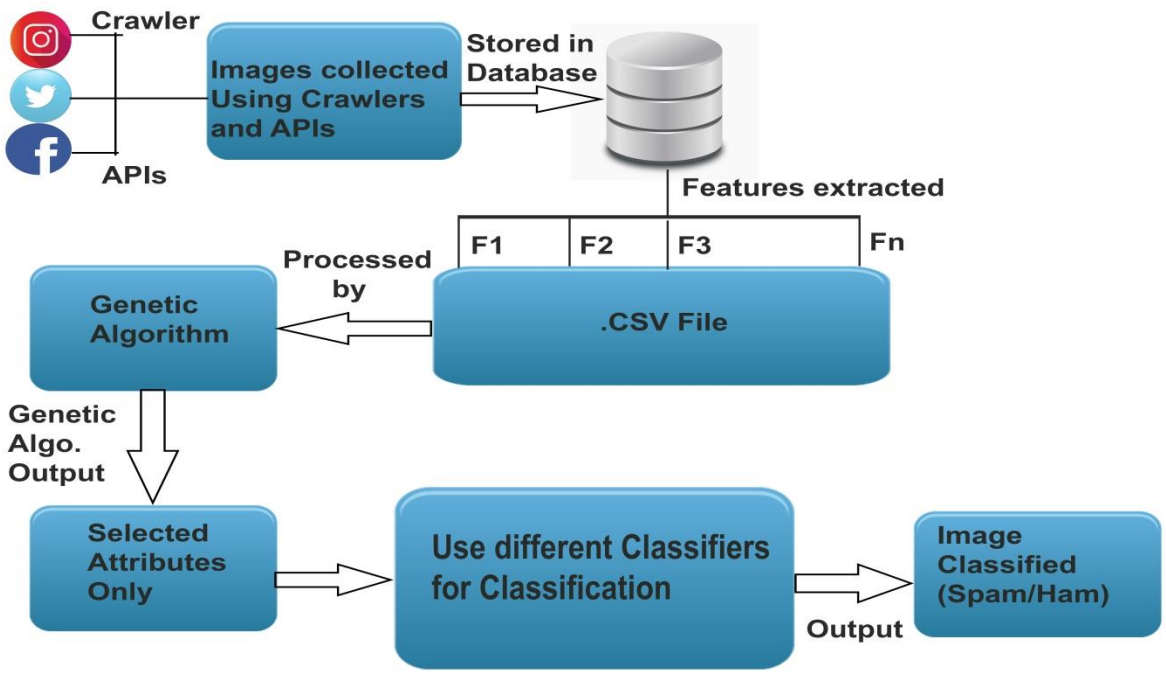

Figure 2 System Architecture

Now a day's spammer's uses image based spam to fool the previously implemented spam detection systems. Therefore, in this project we will develop an image classifier, which classifies images as Spam, non-spam or ham. Here, Spam images refer to images that have large amount of text, memes, cartoons etc. and non-spam images in general are clicked pictures of humans, landscapes, animals etc. Image Spam is the type of spam where the textual spam message is embedded into images. A spam image contains HTML code which contains link of the other websites which further leads to stealing of user's personal data. An effective spam system consists of the following three parameters namely accuracy rate, efficiency, extensibility and all of these features are already contained in text based spammers but since image spammers are quite new as compared to text based spammers so they are lacking these characteristics and there is great scope of innovation in image spammers.

1.2What leads to Spam Image entering into System?

There can be a lot of causes of these and majority of it includes the following:

I. Inadequate Permission
II. Sending Email with Spam content
III. Unauthenticated ID of sender
IV. Not appealing to the receipt
V. Legal Violations

2 Literature Survey:

Authors in [4], a technique based on multiple features is analyzed, where the classification uses inference system which is difficult to perceive. The accuracy of the technique used in this paper is around 85%.Authors in [5], relies on global image features, including color and gradient histograms, with probabilistic boosting trees used for classification. The accuracy of the technique used in this paper is around 94%. Here, they apply clustering at the server side to separate the ham and the spam images. At the client side, feature based methods are used for classification. The existing system extracts a large feature set, which results in a relatively high computational complexity. The work is based on low-level image features. According to their experimental results, simply analyzing basic file properties detects almost 80% of image spam. By using histogram analysis, the existing systems are able to remove about 84% of the spam images, with a minimal false positive rate. Authors in [6], combines fingerprint technique and big data processing to detect the spam emails. Different machine learning techniques are used there The architecture of this particular model is shown in Figure (1). In [7], the author used a technique in which they already maintained the keyword privacy in the encrypted cloud data, and suggested a method to solve the problem of effective fuzzy keyword search. In the other works that study data protection, the proposed signature is used for hidden the identity of data owner.

Author in [8] proposed a model of filtering the spam images and their algorithm work in stages. That particular model is based on Naïve Bayes Classifier which finally identifies the spam email by recognizing their behavior. Author in [9] suggested a model Mail Avenger which works on the network layer and the basis of their model is on blacklist, whitelist and greylist. Blacklist contains the DNS of source which we are sure about and them to be blocked definitely for sure. Whitelist simply means that the DNS is legitimate and need not to be blocked at all. Greylist simply the DNS names which we are not sure about that whether they are legitimate or not .Author in [10] determines a model in which image spam are determined by extracting the text from the image by using OCR but later the text is modified by some tricky methods (CAPTCHA). The accuracy rate of this particular model is decreased as compared to the previous model. There are two types of image features which are high level and low level features. The high level feature composes of file size, file format, file name and many more [11].Low level features comprises of texture, shape, colors and many other [11].Author in [12]

combined a method in which both low level and high level features are studied simultaneously. Then after thorough studying of both decision tree classifier is used and this method definitely increased the accuracy rate of determining the spam images. Authors in [13] suggested a approach in which features are studied collectively in comparison to giving importance only to a single feature. They combined features like aspect ratio, image area, compression ratio, bit depth etc. Authors in [17] proved that making a histogram will increase the efficiency a lot because a lot of features can be extracted even from the histogram like RGB values and entropy of the whole image. The author combined the histogram with the previously used features by the other authors and hence as a result the accuracy improved a lot because numbers of features which are used to train the dataset are increased. Authors in [18] proved a way by taking the partial images step by step. Once the image is taken along the X-axis and a partial image is generated which is referred as Sobel-X and in the other step partial image along the Y-axis is generated which is termed as Sobel-Y then Laplacian correction is used in which both the axis are combined together.

3 Technical Contributions:

Since, there is a rapid growth in online social network like Facebook, Instagram, Whatsapp, Twitter, etc. Therefore, the chances of these websites or apps being affected by spam i.e. text or images are more. Already existing system are able to detect text based spam easily and conveniently but there is still lack of system which will identify that a particular image is spam. Till the present date there is not an efficient and accurate method to detect an image spam. Therefore, in this project we will design and develop a framework which will distinguish between spam and ham images with a decent accuracy rate. Our framework will include detection of spam images through attributes of images like percentage of text in an image, RGB values of image, etc. using various machine learning techniques. Once we identified image as a spam then we can further use our knowledge to prevent these spam images to some extent.

3.1 Dataset Building:

Since, this is a recent concept of detecting the images as spam/ham so dataset is not available on the Google directly and it should be built so that we can classify every image as spam/ham based on machine learning. APIs and Google chrome extension (Download Album) are used for making the dataset. Also various social media platforms need to be selected for downloading the spam/ham images. In this particular project, we downloaded images from Google, Facebook, Twitter and some other few sites and finally a dataset of 14,929 images is made.

3.2 Extracting features from the image:

Since images can't be used directly so features must be extracted from the images and these features are the ones who will provide the final base of classifying an image into spam/ham.

Various features which are used in making the project successful are as follows:

Sr. No.	Features	Explanation	References
1	Width	It is the horizontal expansion of the image	[13]

2	Height	It is the vertical expansion of the image	[13]
3	Image Size	It is the actual size of the image file. Size of image file in KB is used. These features are extracted easily without coding or decoding	[13]
4	Aspect Ratio	Aspect ratio is the ratio of image width to image height	[13]
5	Image Area	Image area is multiplication of height of the image to its width	[13]
6	Compression Ratio	Spammers often change the original size of the image because spammers send spam images in bulk and the size of the original image is large so they reduce the image size. Compression ratio is the ratio of original size of the image to the changed image done by spammers. Or it is the ratio of image size to image area	[13]
7	Bit Depth	Number of unique colors present in the image's color palette in terms of 0's and 1's or bits is defined as the bit depth of the image	[13]
8	Total Pixel	An image is made by numbers of pixels. Total pixels are the total number of pixels available in the image	[13]
9	Saturation	It is the intensity of the colors in the image. The pure red color is fully saturated with the value of 1. Tints of red have the value less than 1 and the color white has saturation value 0	[13]
10	Entropy	Entropy is defined as the measures of the degree of randomness in the image. If the value of entropy is	[14]

		high it means the image has more details else it has fewer details	
11	Average Color	An image is basically made up of three primary colors red, green and blue. An average color of the image is defined as the average of red, green and blue values of the image	[15]
12	Edge Detection	Determined by the point where the intensity of image just changes abruptly with the movement of next pixel	[16]
13	Color Histogram	Color distribution of an image is defined inside the graph. RGB values graph is known as histogram	[17]
14	Sobel X	Edges along the X-axis are defined. This results in a partial image	[18]
15	Sobel Y	Edges along the Y-axis are defined. This results in a partial image along Y-axis	[18]
16	Hue	Hue of color refers to which pure color it resembles. All the tints and shades of the color red have the same hue value. Red has value 0, yellow has value 1/6, green has value 1/3 and so on	New
17	Value	The value is the lightness of the image. It means how light or dark the image is. The black color has the value 0 and white has the value 1	New
18	Laplacian	Both the axis of images are combined inside the Laplacian	New
19	Entropy of Red	Degree of Randomness of red color in the image.	New

	Color	Higher the degree means details are in the image.	
20	Entropy of Blue Color	Blue color randomness which is present inside the image is defined as entropy of blue color.	New
21	Entropy of Green Color	Degree by which green color vary inside the image is termed as entropy of green color.	New

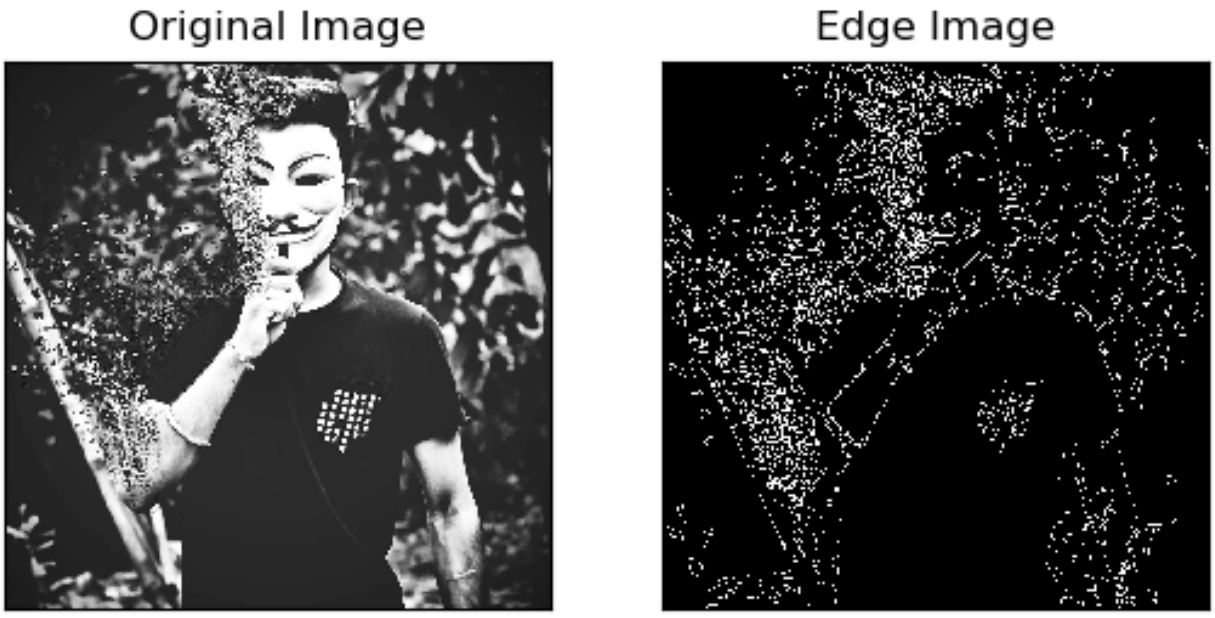

Fig. 3 Edges of the image.

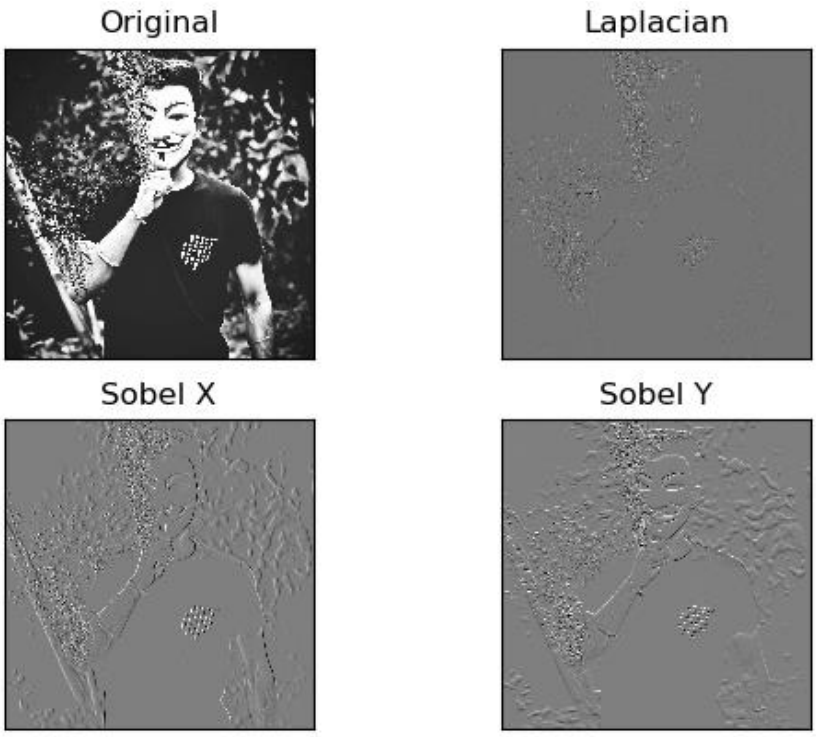

Figure 4 Sobel X, Sobel Y and Laplacian of the image.

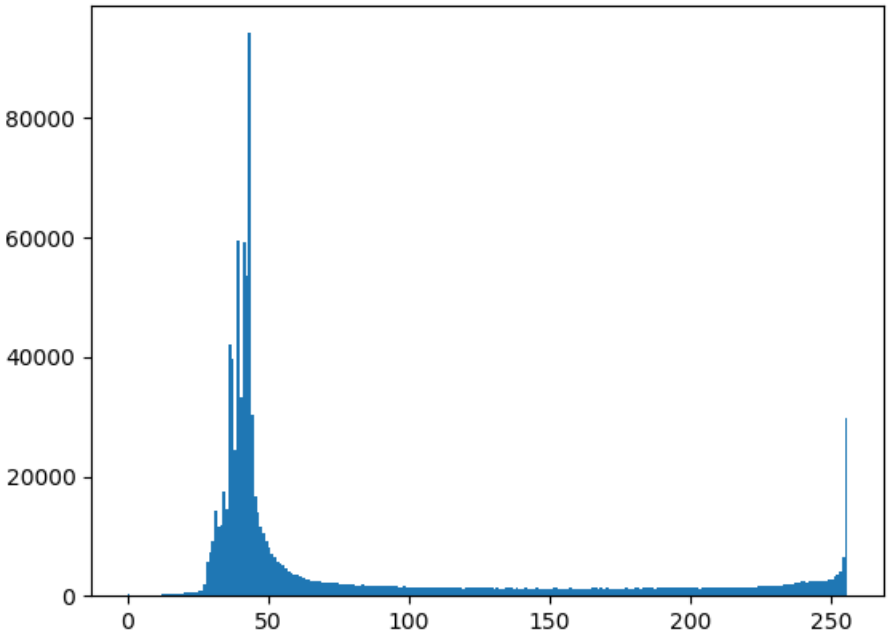

Figure 5 Histogram of the image.

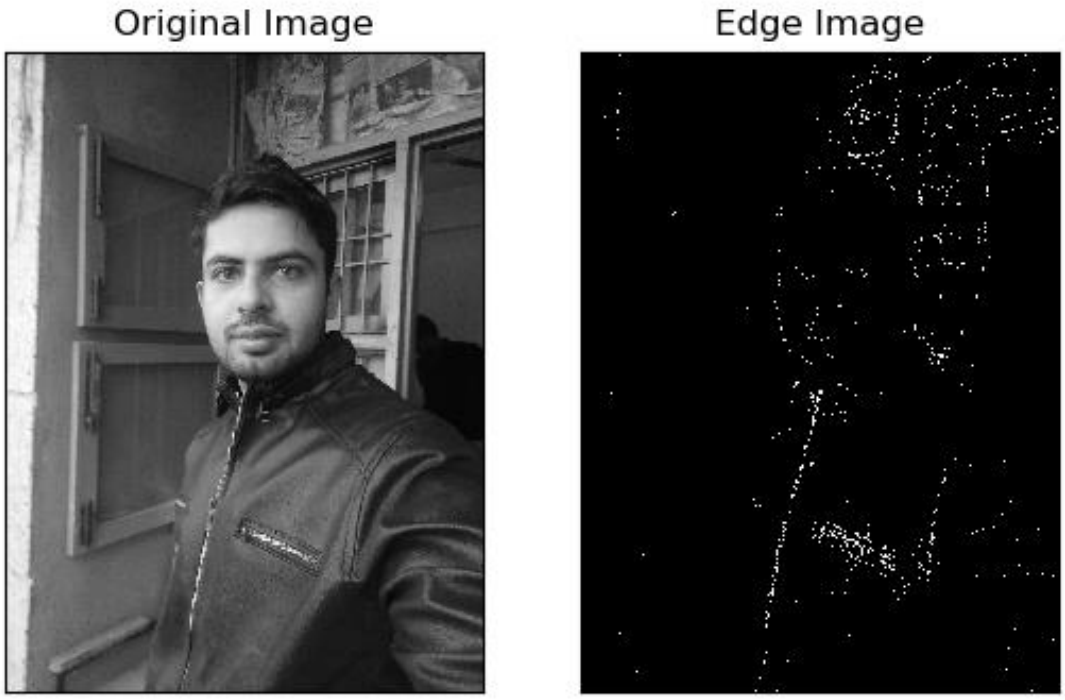

Figure. 6 Edges of the image.

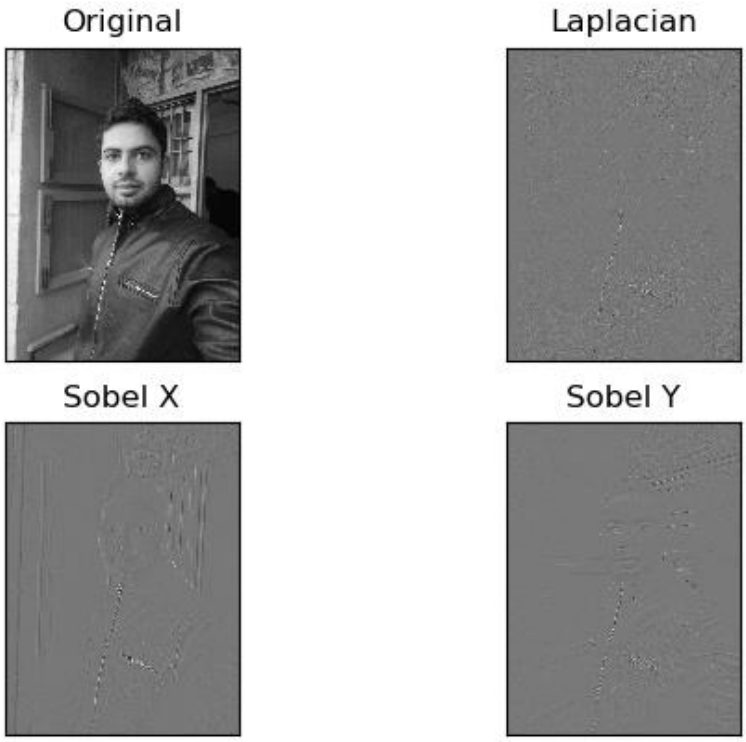

Figure 7 Sobel X, Sobel Y and Laplacian of the image.

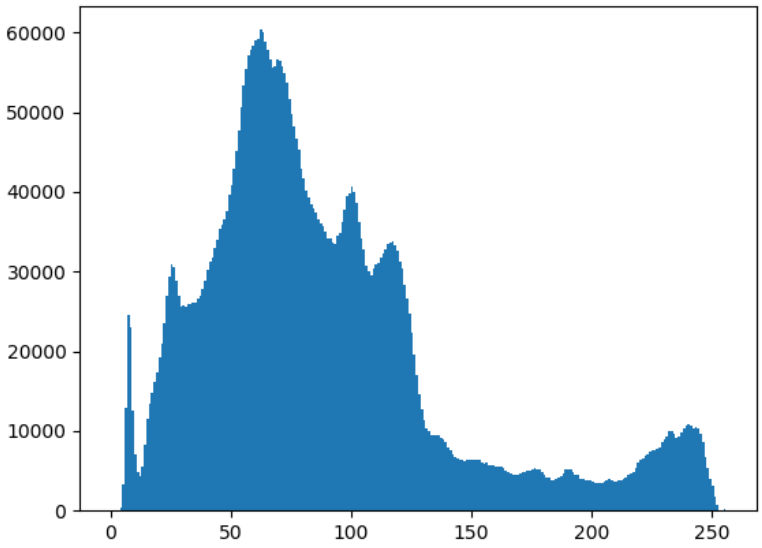

Figure 8 Histogram of the image.

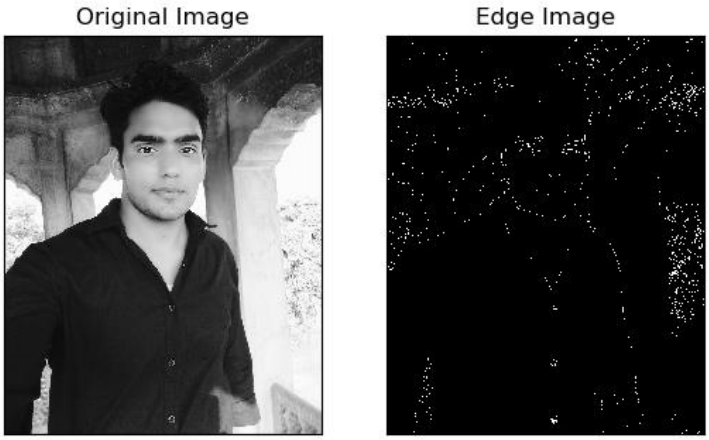

Figure. 9 Edges of the image.

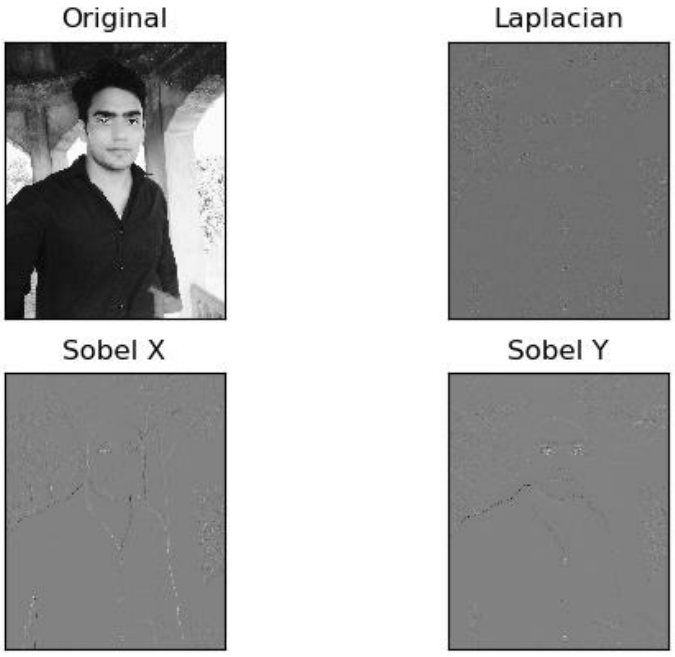

Figure 10 Sobel X, Sobel Y and Laplacian of the image.

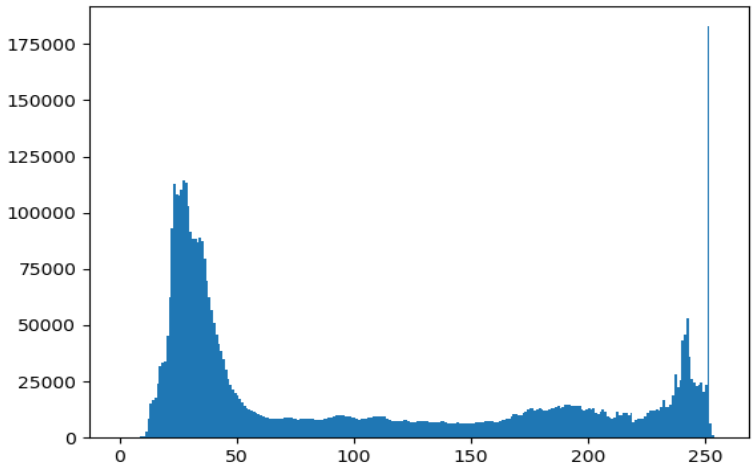

Figure 11 Histogram of the image.

3.3 Processing With Genetic Algorithm:

Genetic Algorithm is based upon the Charles Darwin principle which says, It is not the strongest of species that survives, not the most intelligent but the species which is most responsive to the change".

The same concept which exists in nature is used in the Genetic Algorithm. It basically selects out some features from the existing population then with the evolution of next generation there might be two cases to occur that is either the already existing population got vanished or the new population which evolved is mutated.

```
# Launch genetic algorithm

pop, log = algorithms.eaSimple(pop, toolbox, cxpb=0.5, mutpb=0.2, ngen=numGen, stats=stats, halloffame=hof, verbose=True)

#pop    = a list of individuals
#toolbox= a toolbox that contains the evolution operators
#cxpb   = the probablility of mating 2 individuals
#mut    = the probablility of mutating an individual
#ngen    = the number of generation
#stats   = the statistics object that is updated inplace
#HallOfFame = the halloffame object whcih contains the best individuals
#verbose = whetehr to log or to not to log statistics
```

Figure 12 Genetic Algorithm Function

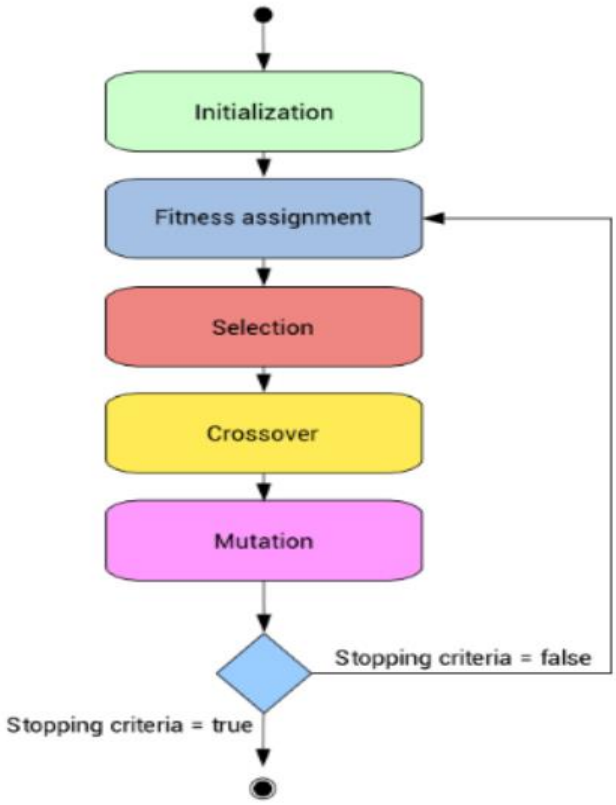

Figure 13 Genetic Algorithm Steps

Steps of Genetic Algorithm

Step1: Initialization

Since our population will contain will individuals and each individual is assigned a particular feature as in the real life .It is assigned a value 0 or 1.0 means that a particular feature is dropped and 1 means that particular feature is present.

Step2: Fitness Function

It defines how a particular function is fit in order to be transferred to the next generation. It is similar to the concept that some weights are there and some profit is assigned to each value and only the individual/item having the maximum profit along

with the minimum weight is transferred to next generation.

Step3: **Selection**

Since we are having all the individuals with the fitness values of chromosomes so we choose the best chromosomes that are fittest in comparison to other chromosomes.

Step4: **Cross Over**

Parents are selected in the crossover step and they are going to produce a new individual/offspring in Cross Over step

Step5: **Mutation**

Since during the production of next generation sometime the traits of produced offspring differs from the parent and this is termed as the mutation.

Genetic Algorithm results in a arrays of features which guarantees maximum efficiency production of results.

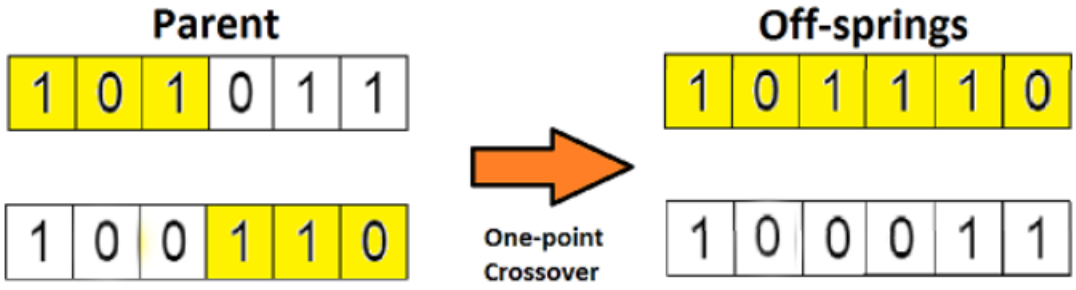

Figure 14 Crossover.

Figure 15 Mutation.

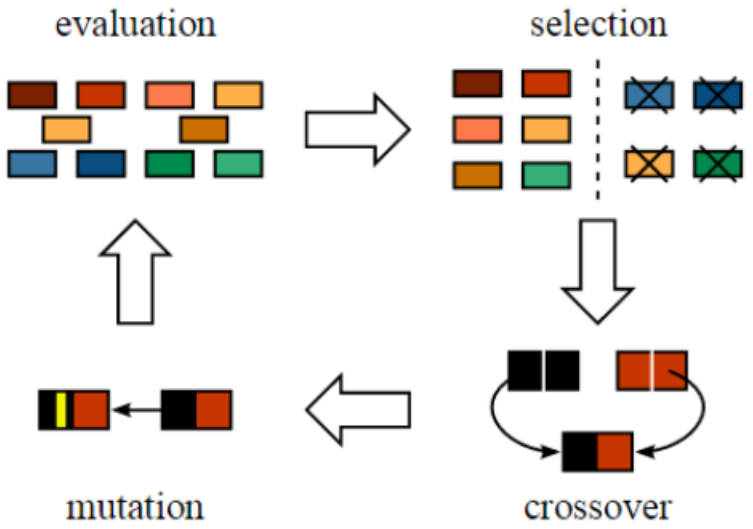

Figure 16 Complete Genetic Algorithm Process

3.4 Processing with Classifiers:

Now the optimal feature subset is processed using the various classifiers and the final result which shows the accuracy percentage is compared for the following classifiers.

I. Decision Tree Classifier

II. Random Forest Classifier

III. Naïve Bayes Classifier

IV. Bagging Classifier

V. Logistic Regression Classifier

VI. Gradient Booster Classifier

4 DATA FLOW DIAGRAM

4.1 LEVEL0 DFD:

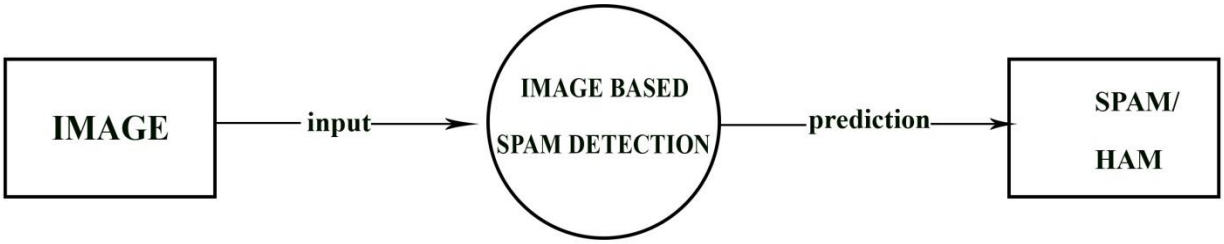

Figure 17 Level 0 DFD.

4.2 LEVEL1 DFD:

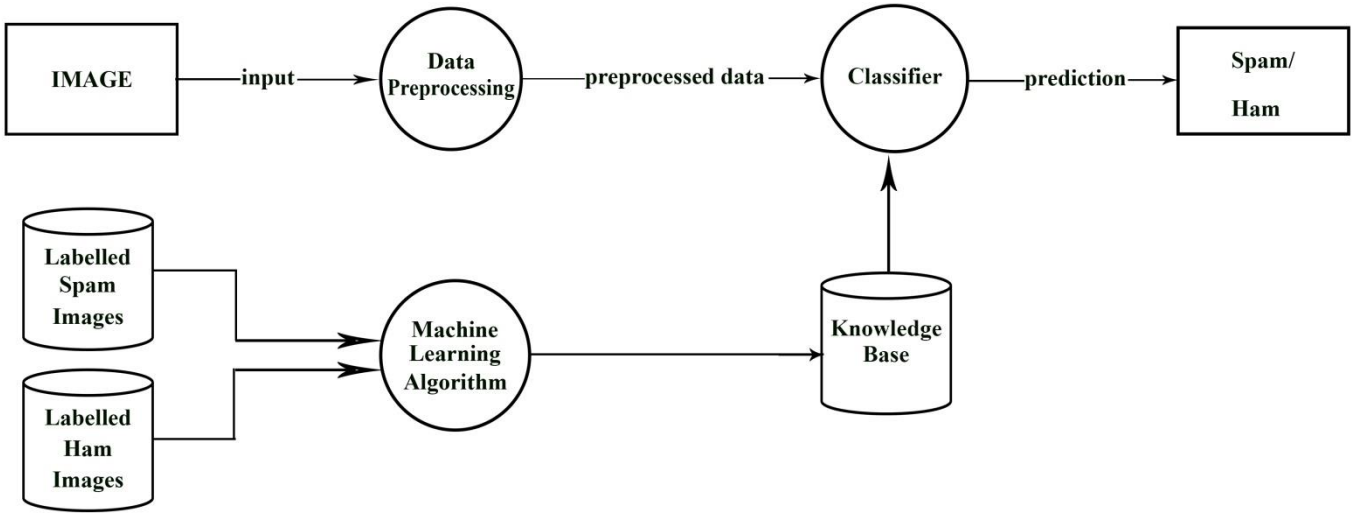

Figure 18 Level 1 DFD.

4.3 LEVEL2 DFD:

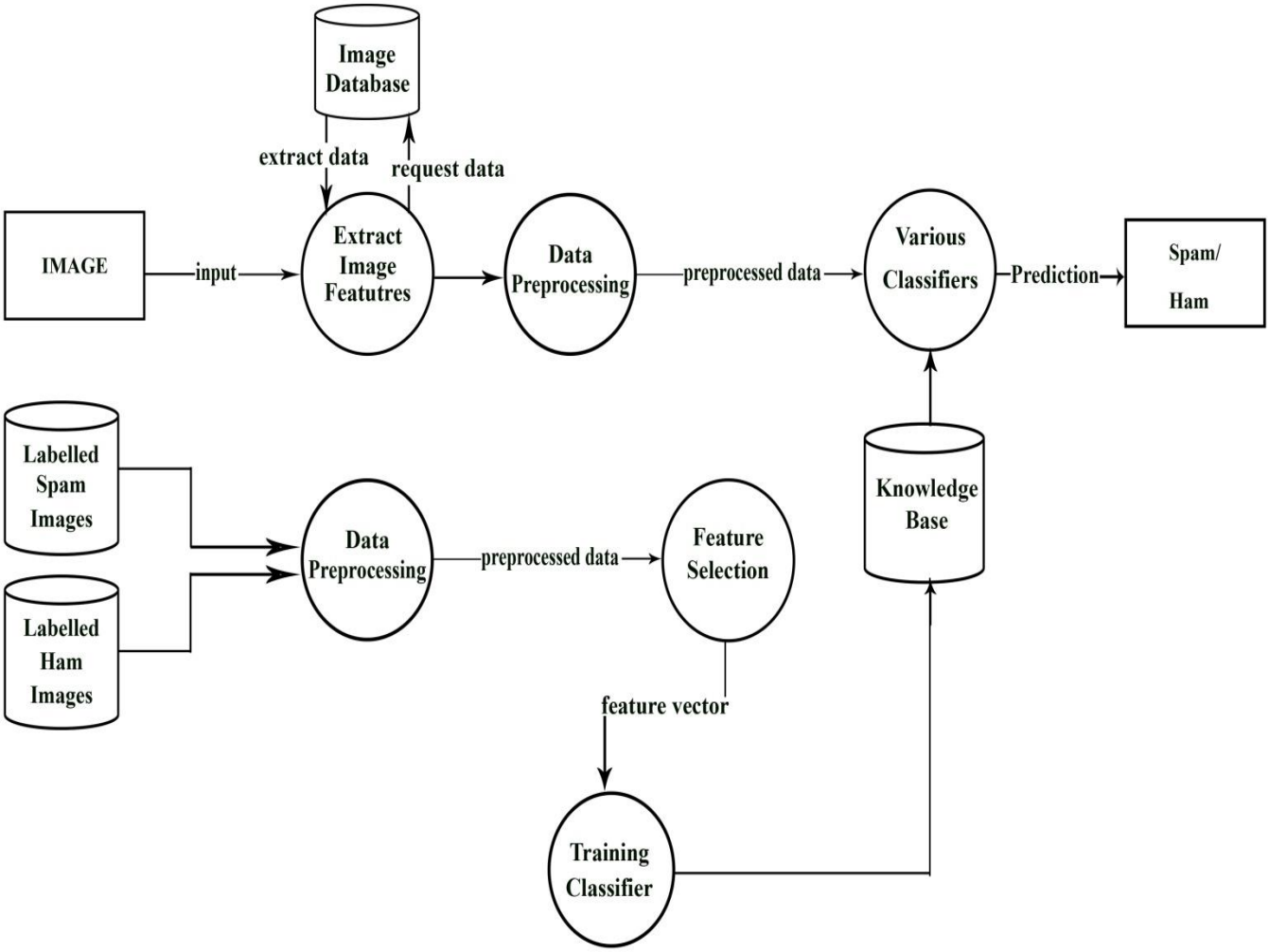

Figure 19 Level 2 DFD.

5 Observations and Results

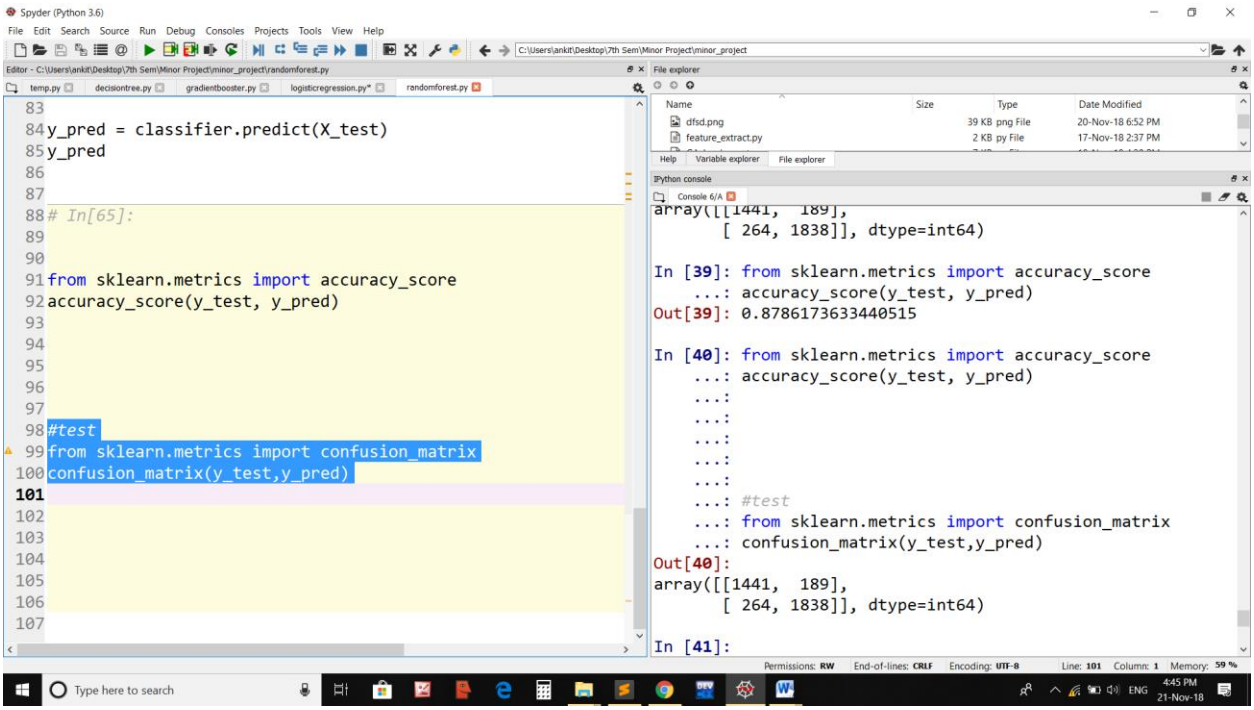

Figure 20 Random Forest Classifier without genetic algorithm.

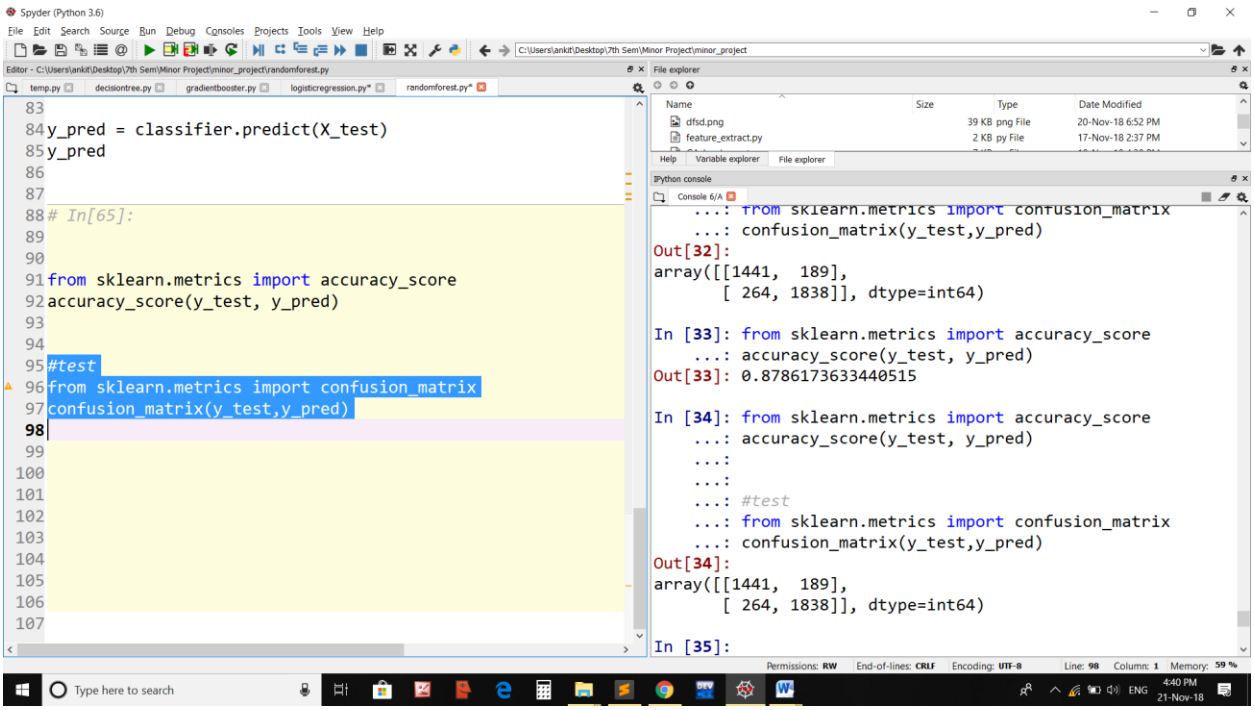

Figure 21 Random Forest Classifier with genetic algorithm.

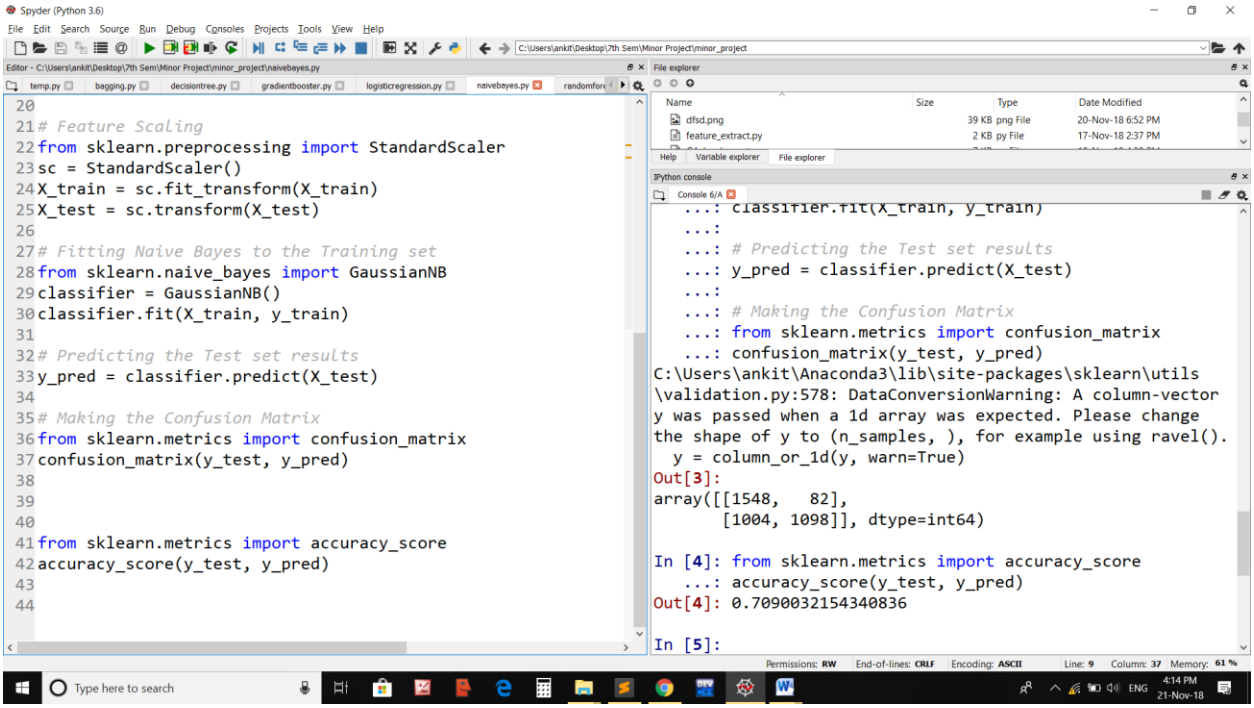

Figure 22 Naïve Bayes Classifier without genetic algorithm.

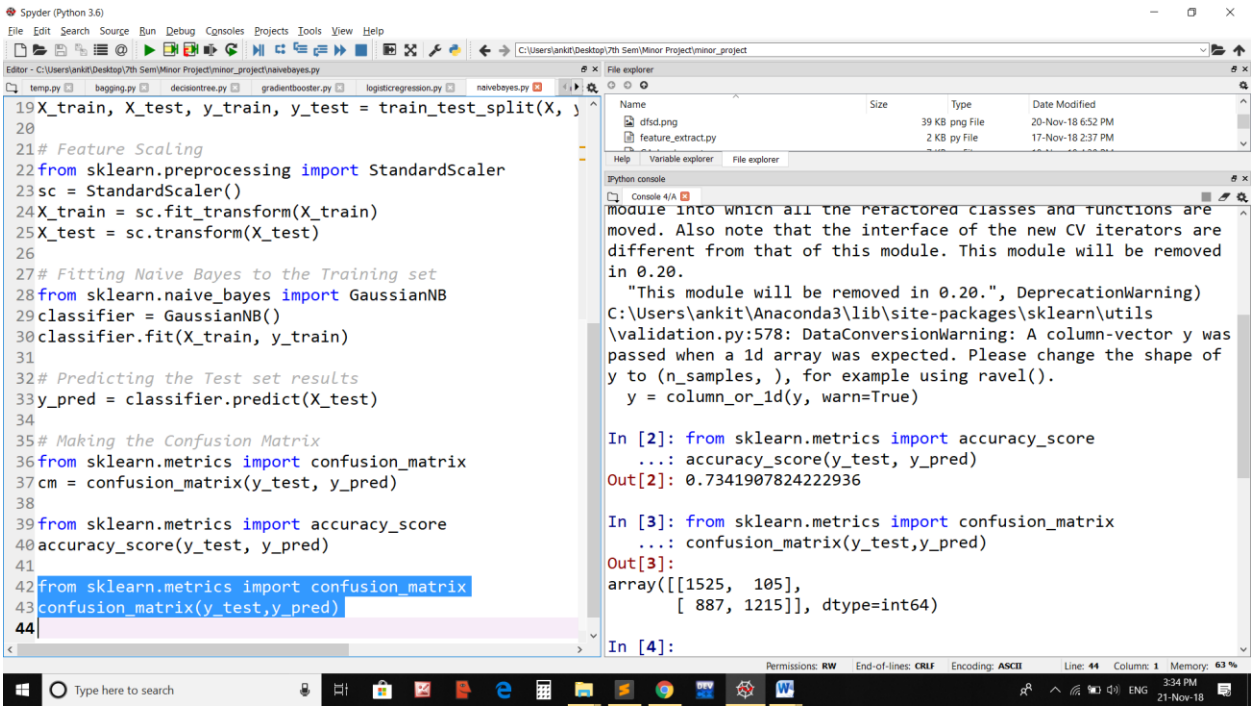

Figure 23 Naïve Bayes Classifier with genetic algorithm.

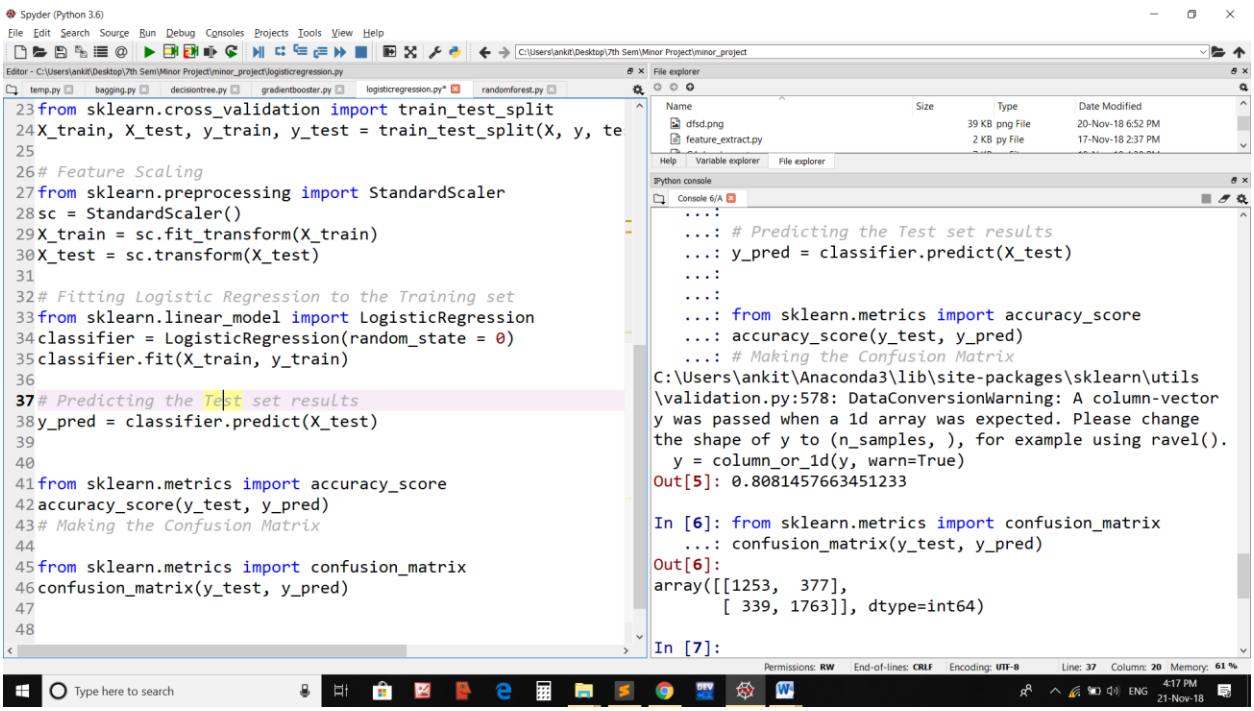

Figure 24 Logistic Regression Classifier without genetic algorithm .

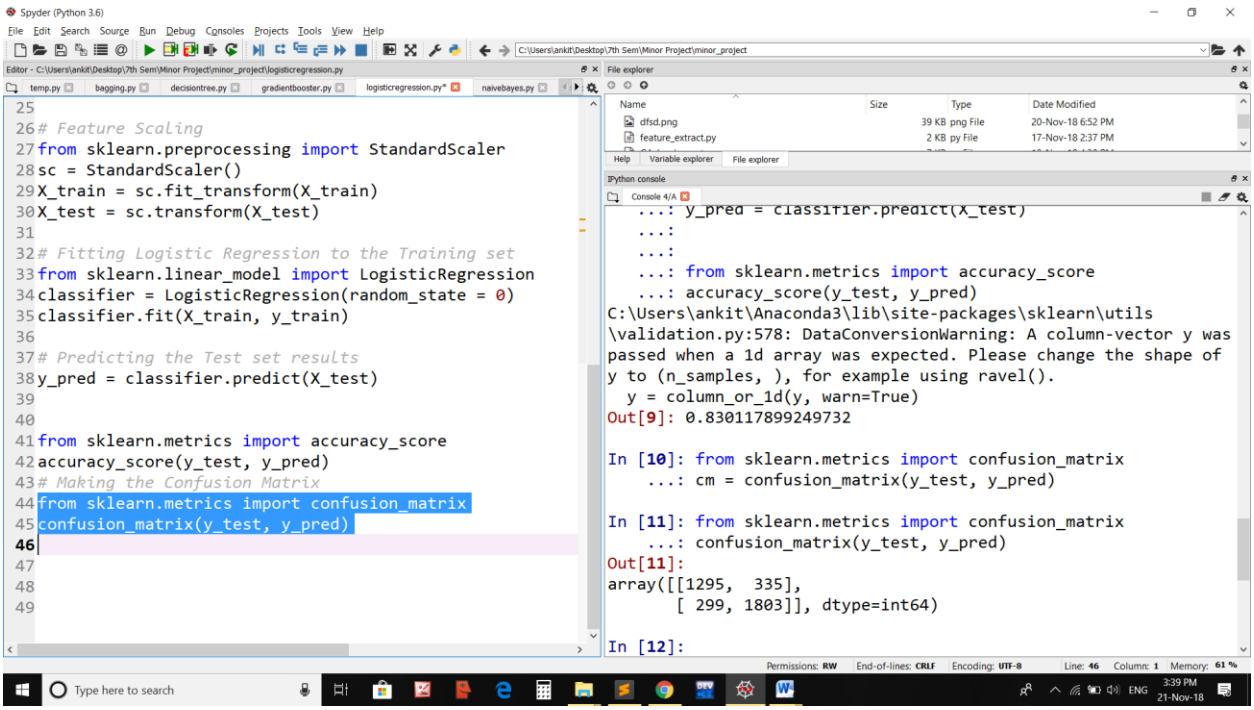

Figure 25 Logistic Regression Classifier with genetic algorithm .

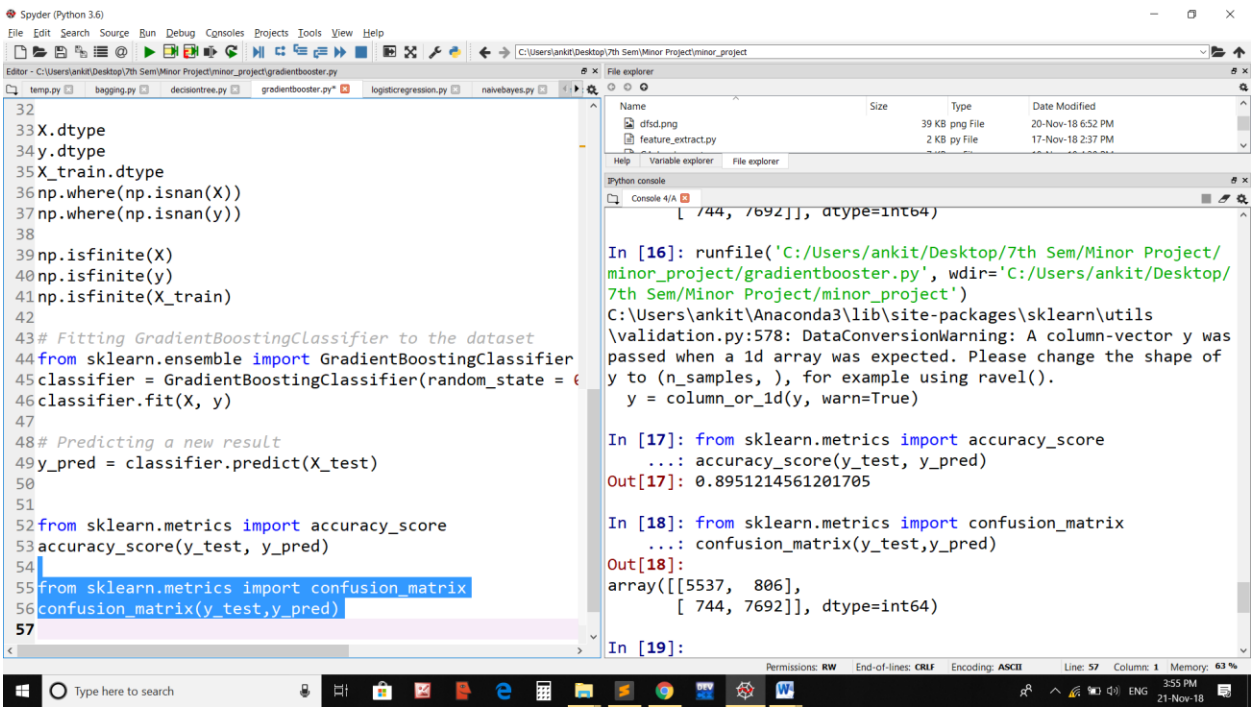

Figure 26 Gradient Booster Classifier without genetic algorithm.

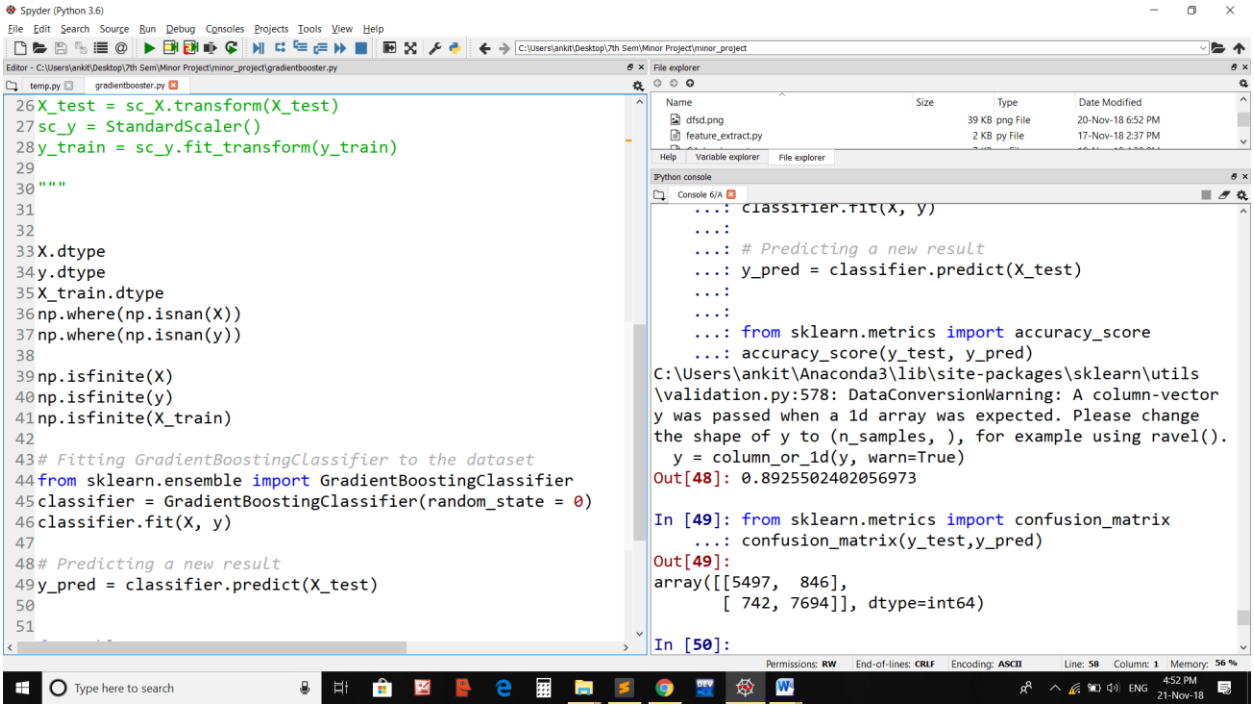

Figure 27 Gradient Booster Classifier with genetic algorithm.

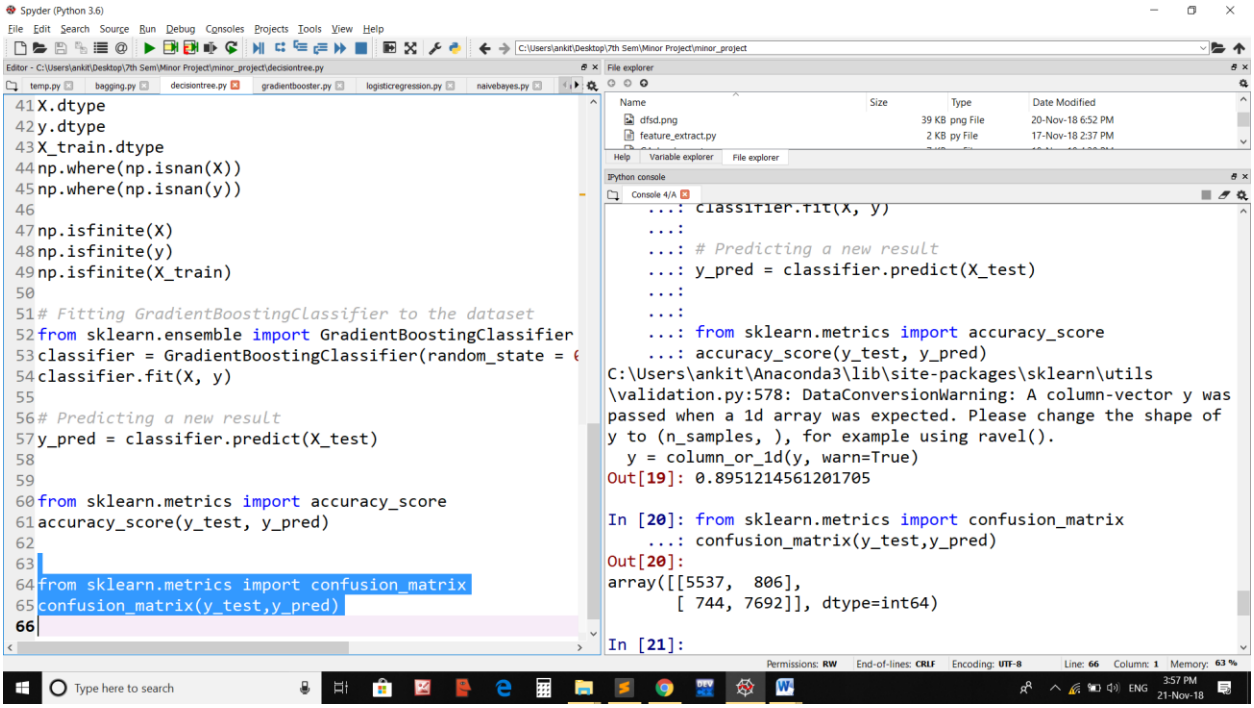

Figure 28 Decision Tree Classifier without genetic algorithm.

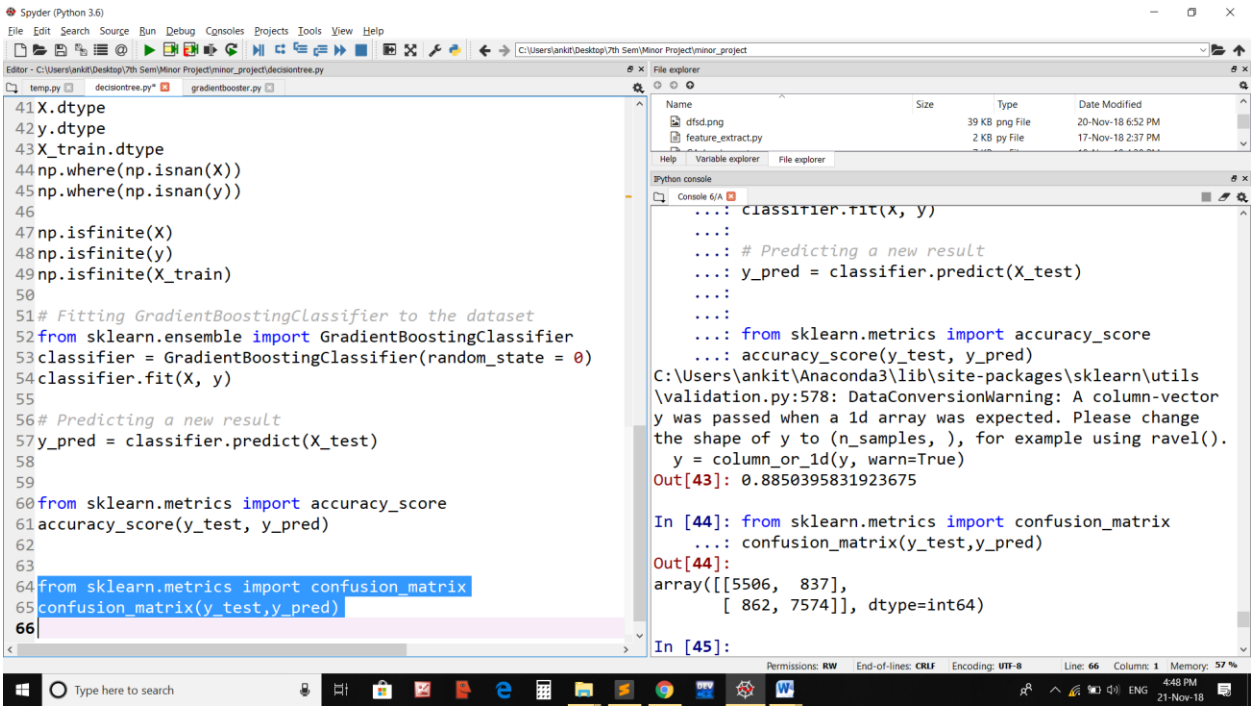

Figure 29 Decision Tree Classifier with genetic algorithm.

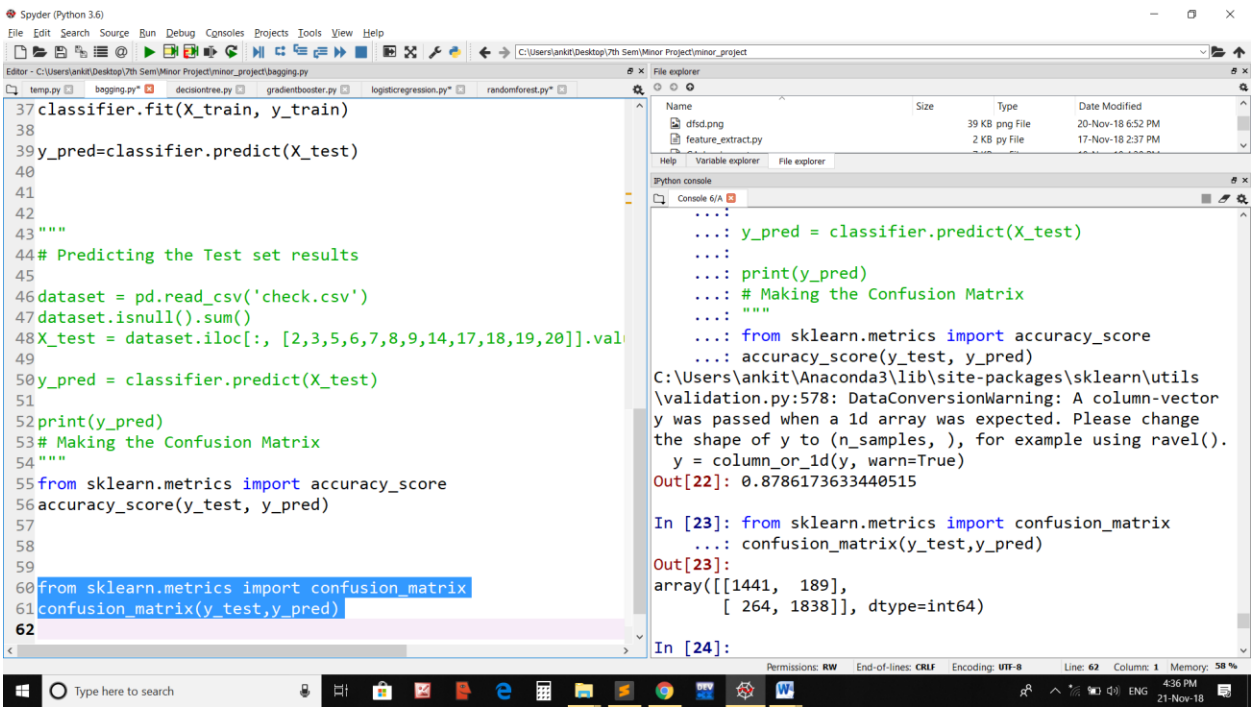

Figure 30 Bagging Classifier without genetic algorithm.

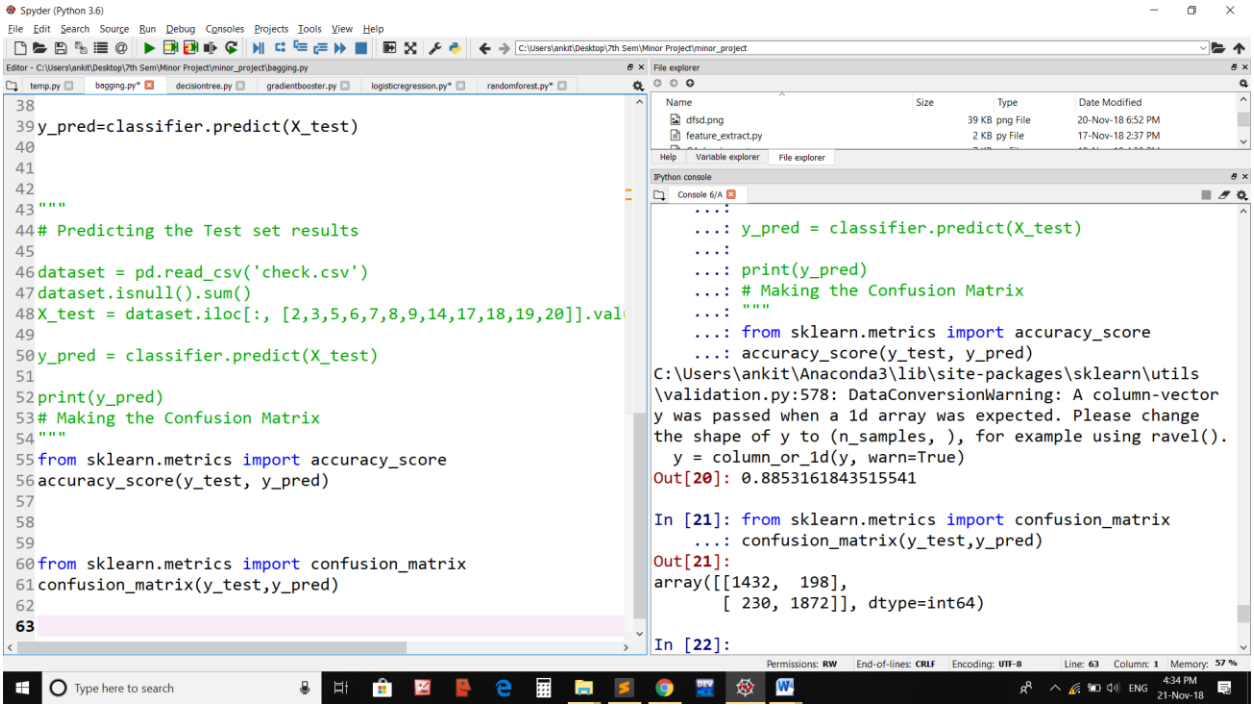

Figure 31 Bagging Classifier with genetic algorithm.

Sr. No	Classifier	Without Genetic Algorithm	With Genetic Algorithm
1	Random Forest Classifier	87.86%	87.86%
2	Decision Tree Classifier	89.51%	89.62%
3	Logistic Regression Classifier	80.81%	83.01%
4	Gradient Booster Classifier	89.51%	89.62%
5	Naïve Bayes Classifier	70.90%	73.41%
6	Bagging Classifier	87.86%	88.53%

Figure 32 Comparison Table among different classifier.

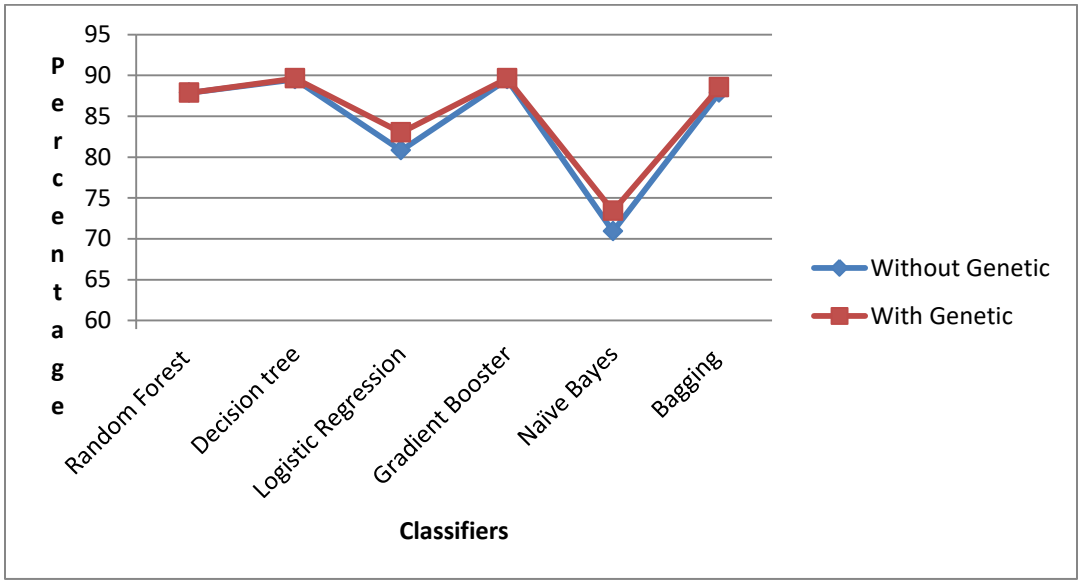

Figure 33 Visual camprison among different classifire.

6 Conclusion:

We in this project aimed at developing a technique which helps to detect with image as spam/ham with a decent accuracy rate. Firstly, all specific types of images as have some features in common .For instance, spam images have more text but these are lacking in the ham images. Hence, features are extracted from the images and a database a made in .csv form suing those features. An image based spam detection system is developed by using the optimal features of the .csv file which is nothing but simply the output of Genetic Algorithm. The optimal features guarantee the maximum accuracy rate because these are the tested possible combinations during the execution of Genetic Algorithm. Then various algorithms of machine learning are used to classify that whether image is spam/ham. It will be a mix of algorithms for more accuracy. Since different Classifiers are using different approaches to classify an image into a particular group so we can easily compare the accuracy rate of different classifiers. Hence, it will give an overall clarity that whether the image is spam/ham because result of all classifiers can be compared for that image.

7 Future /Works:

Since current system is lacking the GUI so an Google extension or an android app can be made which will be very handful to use rather than the command line mode. Some other features can also be added which will definitely improve the accuracy. Since this particular system is only detecting a spam image so it should be combined with the system of prevention to prevent that spam images from entering into our computer system. That preventive should be basically based on the system of exploiting the vulnerabilities of the spam images. Once the detection system is combined with the prevention system than it will result into a complete GUI based software which will ease out our sharing of information in the various social networking sites like Facebook, Twitter, and Whatsapp etc.

8 Software Requirements:

Programming languages: Python and its framework.

IDE Spyder (Python IDE)

Machine learning algorithms

References:

1 Fire, M., Goldschmidt, R., & Elovici, Y. (2014). Online social networks: threats and solutions. IEEE Communications Surveys & Tutorials, 16(4), 2019-2036.

2 Gargiulo, F., Penta, A., Picariello, A., & Sansone, C. (2008, September). Using heterogeneous features for anti-spam filters. In Database and Expert Systems Application, 2008. DEXA'08. 19th International Workshop on (pp. 670-674). IEEE.

3 Annadatha, A., & Stamp, M. (2018). Image spam analysis and detection. Journal of Computer Virology and Hacking Techniques, 14(1), 39-52.

4 Liu, T. J., Tsao, W. L., & Lee, C. L. (2010, August). A high performance image-spam filtering system. In Distributed Computing and Applications to Business Engineering and Science (DCABES), 2010 Ninth International Symposium on(pp. 445-449). IEEE.

5 Liang, H., Chen, Z., & Wu, J. (2018). Dynamic reputation information propagation based malicious account detection in OSNs. Wireless Networks, 1-14.

6 Parekh, P., Parmar, K., & Awate, P. (2018). Spam URL Detection and Image Spam Filtering using Machine Learning.

7 Chen, J., Zhao, H., Yang, J., Zhang, J., Li, T., & Wang, K. (2017). An intelligent character recognition method to filter spam images on cloud. Soft Computing, 21(3), 753-763.

8 Ming, L., Yunchun, L., & Wei, L. (2007, November). Spam Filtering by stages. In Convergence Information Technology, 2007. International Conference on (pp. 2209-2213). IEEE.

9 Mail Avenger, 2006. http://www.mailavenger.org

10 Fumera, G., Pillai, I., & Roli, F. (2006). Spam filtering based on the analysis of text information embedded into images. Journal of Machine Learning Research, 7(Dec), 2699-2720.

11 Krasser, S., Tang, Y., Gould, J., Alperovitch, D., & Judge, P. (2007, June). Identifying image spam based on header and file properties using C4. 5 decision trees and support vector machine learning. In Information Assurance and Security Workshop, 2007. IAW'07. IEEE SMC (pp. 255-261). IEEE.

12 He, P., Wen, X., & Zheng, W. (2009, June). A simple method for filtering image spam. In Computer and Information Science, 2009. ICIS 2009. Eighth IEEE/ACIS International Conference on (pp. 910-913). IEEE.

13 Wang, C., Zhang, F., Li, F., & Liu, Q. (2010, July). Image spam classification based on low-level image features. In Communications, Circuits and Systems (ICCCAS), 2010 International Conference on (pp. 290-293). IEEE.

14 Gao, Y., Choudhary, A., & Hua, G. (2010, March). A nonnegative sparsity induced similarity measure with application to cluster analysis of spam images. In Acoustics Speech and Signal Processing (ICASSP), 2010 IEEE International Conference on (pp. 5594-5597). IEEE.

15 Xu, Z., Wang, H. G., & Shao, Z. Z. (2009, December). Evaluation of image spam classification system based on AHP. In Computational Intelligence and Software Engineering, 2009. CiSE 2009. International Conference on (pp. 1-4). IEEE.

16 Jithesh, K., Sulochana, K. G., & Kumar, R. R. (2003, March). Optical character recognition (OCR) system for Malayalam language. In National Workshop on application of language technology in Indian languages (Vol. 6, p. 7).

17 Soranamageswari, M., & Meena, C. (2010, February). Statistical feature extraction for classification of image spam using artificial neural networks. In Machine Learning and Computing (ICMLC), 2010 Second International Conference on (pp. 101-105). IEEE.s

18 Woods, N. C., Longe, O. B., & Roberts, A. B. C. (2012). A sobel edge detection algorithm based system for analyzing and classifying image based spam. Journal of Emerging Trends in Computing and Information Sciences, 3(4), 506-512.

Appendix

As per the discussion with our project mentor, Dr. B.B. Gupta , this code is private and we might use this code for publication in future.

Signature of Supervisor

Dr. B.B. Gupta

Assistant Professor