# Hyper Parameter Optimization using Genetic Algorithm on Machine Learning Methods for Online News Popularity Prediction

Ananto Setyo Wicaksono[1], Ahmad Afif Supianto[2]

Department of Informatics
Faculty of Computer Science, Brawijaya University
Malang, Indonesia

*Abstract*—Online news is a media for people to get new information. There are a lot of online news media out there and a many people will only read news that is interesting for them. This kind of news tends to be popular and will bring profit to the media owner. That's why, it is necessary to predict whether a news is popular or not by using the prediction methods. Machine learning is one of the popular prediction methods we can use. In order to make a higher accuracy of prediction, the best hyper parameter of machine learning methods need to be determined. Determining the hyper parameter can be time consuming if we use grid search method because grid search is a method which tries all possible combination of hyper parameter. This is a problem because we need a quicker time to make a prediction of online news popularity. Hence, genetic algorithm is proposed as the alternative solution because genetic algorithm can get optimal hypermeter with reasonable time. The result of implementation shows that genetic algorithm can get the hyper parameter with almost the same result with grid search with faster computational time. The reduction in computational time is as follows: Support Vector Machine is 425.06%, Random forest is 17%, Adaptive Boosting is 651.06%, and lastly K - Nearest Neighbour is 396.72%.

*Keywords—Hyper parameter; genetic algorithm; online news; popularity; machine learning*

## I. INTRODUCTION

The news is information about what is happening in the world. This information is useful to people, for example as the topic of conversation, decision making, etc. News can be obtained by several ways, such as printed media, television, and online. Nowadays, online media is one of the most accessed media to obtain news [1].

Online news is a popular media to get information because a lot of people is using internet on their gadget [2]. The ability to comment and share the news is one of the appeals of the online news. Usually, people tend to share news they think interesting. The number of shares is one way to measure the popularity [3]. After reading the news, people don't read the same news contained in media news. This makes unshared online news not read by people [4]. If the news is not read by many people, an advertiser will not insert their advertisement and without advertisement, the online news media will lose its source of income. To solve it, it is necessary to know whether a news will be popular or not before its' publication [2].

Online news popularity prediction can be solved by one of the data mining techniques that are classification [5]. Classification is a way to discover the class of a certain input data using a certain rule. Support Vector Machine (SVM) [2], Random Forest (RF) [2] [6], Adaptive Boosting (AdaBoost) [2], K-Nearest Neighbour (KNN) [2], Naive Bayes (NB) [2] [7] was used to make a classification of online news popularity. The result was most of them can only get a result in 60% accuracy. To be able to get a more reliable result in online news popularity prediction, it is needed to produce a method that can make a result better than the previous research. This unsatisfactory prediction performance may occur because of the characteristic of the online news data.

In online news, there are a lot of features that can influence the amount of the popularity. That redundant feature usually makes the result of the prediction worse [8]. That is why feature reduction is one way to raise the result. Feature reduction is to use reduced feature instead of a full feature of the dataset. Feature reduction can be achieved by using feature selection and feature extraction. The two of them are methods to reduce the feature used, but the difference is, in feature selection, we select a number of features from the entire feature to be used in classification process, while in feature extraction, all of the features are converted into a new number of features that is less than the original feature. In [5], feature selection method is used to increase the result of the online news. Information gain, correlation-based, gain ratio, learner-based selection method was proposed and between them, the learner based achieved the best accuracy that is 63%. Even if it is able to get that result, when it comes to very high feature data, feature selection can have a lower performance [9]. For feature extraction, Principal Component Analysis (PCA) was used in [10]. After using PCA, the reduced feature then used as the feature of machine learning methods to do classification task for online news popularity prediction. The result is the SVM can achieve up to 69% accuracy. But, the weakness of using PCA is the feature became uninterpretable and it is necessary to define the optimal number of the principal component to get the optimum result [11]. These methods were methods to increase the performance of the machine learning methods from the attribute used.

Even without feature reduction, machine learning methods' performance can be increased, and one of the methods we can

use is a hybrid algorithm. The hybrid algorithm is combining algorithm with another algorithm to get a better result [12]. In [13] hybrid SVM-RF was proposed to increase the performance of the online news popularity prediction. This implementation indeed gives a better performance by achieving up to 73% accuracy. While it can give a better performance, the downside of using hybrid algorithm is the difficulty in implementation and choosing the combination of the algorithm. If the combination of the algorithm is not right, it will give worse performance than the original algorithm.

Actually, there is a simpler method that can be used to enhance the performance of machine learning methods that is by tuning its hyper parameter. Hyper parameter is parameters inside machine learning methods. Each machine learning method has a different set of hyper parameter. In order to change this hyper parameter, we can just manually change the value until the result is satisfactory. However, manually tuning hyper parameter is a tedious task because there is a lot of possible combination that can be used. That's why automatic hyper parameter tuning is used to find the optimal value.

Grid Search can be used to automatically find the optimal hyper parameter. Grid Search is an optimization algorithm that searches all possible combination in the search space. In [2], Grid Search was used to optimize the hyper parameter of the machine learning methods. It can make the performance of machine learning algorithm enhanced and the best result out of all the machine learning method is RF with 67% accuracy. Grid Search also comes with a downside that is, the computational time is slow [14]. To overcome this problem, another optimization method can be used to replace this method. One of the methods is Genetic Algorithm.

Genetic Algorithm is an optimization algorithm that can get near the optimal value of a function [15]. The time needed to get the optimal value is also plausible. In [14], SVM with genetic algorithm hyper parameter optimization was used to try several free dataset and the result is SVM with a genetic algorithm can get the near optimal hyper parameter using a lot faster computational time.

Hence, a genetic algorithm is proposed in this research to replace the grid search method. This is the novelty of this research because genetic algorithm hasn't been used to determine hyper parameter in online news data. The purpose of using genetic algorithm is to get a faster computational time for determining the popularity of online news because a lot of news are created every day and speed is necessary in online news media.

## II. Material and Methods

Several steps are needed in order to make a popularity estimation of online news. The online news texts need to be translated into several attributes. Then, in order to enhance the result of the prediction, the hyper parameter is tuned using a genetic algorithm. After that, machine learning methods with optimal hyper parameter are used to classify whether online news is popular or not.

### A. Data for Online News Popularity Prediction

The data used in this research was online news popularity dataset downloaded from UCI machine learning website[1]. The dataset consists of 39797 instances and 61 attributes. It consists of many online news articles from Mashable news service [2]. These articles are articles published from January 2013 until January 2015. This many attribute and instances make it a challenging dataset to use because it is a lot of data and the higher the amount of data, the longer the computational time needed. This dataset is the summary information of the news that is needed in order to know the popularity. For example, the number of positive words, negative words, the number of images, the number of shares, etc.

The target of this dataset is the number of shares. If the number of shares reaches a certain threshold, then the news is considered as popular, otherwise, it is not popular. The number of shares can be a way to determine popularity because people won't share an article they don't like to other people. The number of shares is used as popularity measurement. In this research, the minimum number of share of news considered as popular is 1400 shares [2].

### B. Machine Learning Methods

Machine learning is a method to make a computer to have the ability to learn. Statistical methods are often used to achieve it. There are a lot of things that can be learned by using machine learning methods, one of them is classification. Until now, several machine learning methods are already developed to get a better result on learning. Machine learning methods can be used to solve a classification problem, clustering problem, and forecasting problem.

*1) Support Vector Machine (SVM)*: SVM is a machine learning method for regression and classification [14]. SVM as classifier will make a hyper plane that separates the data into several classes. SVM has C, gamma, and kernel as hyper parameter. SVM is a popular machine learning methods used in many classification problems. SVM can be used to classify many things like in [16] as tumour classifier. In [17] SVM is used to classify poetry. [18] Used SVM in bank direct marketing problem.

*2) Random Forest (RF)*: Random Forest is a machine learning method using several decision trees. It can be used as regressor or classifier. Random forest initializes a number of the tree via randomization technique [6]. The hyper parameter of Random Forest is the number of trees, the depth of the tree, etc. If it is tuned, it can achieve higher result than using a default hyper parameter.

*3) K-Nearest Neighbor (KNN)*: KNN is a machine learning method that uses neighbouring data to get the result. KNN can be used as regressor or classifier. As a classifier, KNN will assign new data to a class that is nearest to its K neighbour. K is the number of the neighbour.

---

[1] https://archive.ics.uci.edu/ml/datasets/online+news+popularity

In KNN, firstly, the distance of new data will be measured against all of the data in the dataset. There are various algorithms to measure the distance of the data, such as Euclidean distance and Manhattan distance. Then, after the distance of the new data and all of the data is known, several data with the shortest distance with the new data will determine what the label of the data is. The new data will be assigned to a label with n data with the shortest distance. The number of n is equal to the number of K.

The number of K will affect the result of KNN. KNN is a simple yet effective classification method used in many classification problems [16] [19].

*4) Adaptive Boosting (AdaBoost)*: AdaBoost is a classification method that combines several weak classifier methods to achieve better result [2]. The weak classifier will be trained into the dataset, and then the weight is assigned to them until all classifier has optimal weight and best prediction result. AdaBoost uses only a significant feature in the dataset as training. It makes AdaBoost's accuracy result have a higher accuracy Adaptive boosting tend to have an over fitting problem and sensitive to noisy data and outliers.

*5) Grid Search*: Grid Search is a widely used method for hyper parameter optimization [16]. Grid Search is a hyper parameter optimization method that uses brute force in order to find the best hyper parameter. It is more guaranteed to find the optimal hyper parameter because Grid Search will try all possible combination within a set of parameters. [20] and [21] is the usage of grid search for hyper parameter optimization.

*6) Genetic algorithm*: The genetic algorithm is optimization algorithm inspired by the process of evolution [17]. A chromosome is the representation of solution in the genetic algorithm. Then, it uses crossover and mutation to generate the new solution. Crossover is a mechanism of combining two chromosomes into one chromosome. The mutation is a process to get a new solution by changing one chromosome. The solution will then be evaluated with the objective function and the solution that didn't fit the criteria will be dropped. The solution kept will continue the process of crossover and mutation and evaluation until the stop condition is met. The stop condition can be a number of iteration and a certain time limit.

The genetic algorithm is a versatile algorithm which can be used in several problems. In [22] the authors introduce an approach to multilingual single-document extractive summarization where summarization is considered as an optimization or a search problem which is solved by using genetic algorithms. In [23] genetic algorithm has been applied for effective personalize web search-based on clustered query sessions. The genetic algorithm also used for term-weighting learning in term of text classification [24].

The genetic algorithm can be used as a method to find an optimal hyperparameter of machine learning methods. For example in [25], Genetic Algorithm is used as hyperparameter tuning of a fuzzy rule to classify facial expression. The purpose is to make the fuzzy method to get a better classification result.

*C. Hyperparameter Optimization Methods*

Hyper parameter is a parameter that is necessary for machine learning methods to make a classification. Each machine learning methods have different hyper parameters. Choosing the right parameters can make a significant difference in prediction results. That's why it is important to make a tuning in hyper parameter instead using the default parameters of machine learning. Determining these hyper parameters can be done manually by trying all of the possible value. But, doing that is time-consuming because the number of possible combination is very large. That's why optimization algorithm to automatically find the optimal hyper parameter such as Grid Search is often used.

*D. Proposed Method*

The implementation of this research is shown in Fig. 1. Firstly the dataset needs to be downloaded from UCI Machine Learning website. The format file of the dataset is Comma Separated Values (CSV). This dataset don't have any missing value and all of the data is already in number, so a preprocessing method to fill the missing value and data conversion is not necessary. Almost all of the attributes will be used in the experiment which is 58 attributes. These attributes are the attributes that influence the popularity of the online news.

This dataset needs to be labelled first in order to do a classification because this dataset didn't have a label determining whether it is popular or not. The data in the dataset that can be used as popularity measurement is the number of the shares. If the number of shares reaches 1400 or more, then it is considered as popular, otherwise, it is not popular. This label is important because, in the methods that will be used, the data need to be a categorical data or in this research a binary data because there are only two labels for the data. The new data will be classified into these labels.

After the dataset was labeled, the next step is to split the dataset into two parts. The first part is the training dataset and the second part is testing dataset. The splitting ratio used in this research is 70% for training and 30% for testing. This was done using the Scikit Learn library in python [26]. Scikit Learn is a collection of library specialized at handling machine learning problem. From here on, until the evaluation, Scikit Learn library was used in the implementation.

Hyperparameter optimization comes after that. It is to determine the value of hyperparameter in machine learning automatically. Here, the genetic algorithm was proposed to replace the grid search method. The genetic algorithm will be used before the classification process to make the classification result of machine learning better. It is necessary to define several things before we can use a genetic algorithm, such as the chromosome, crossover rate, mutation rate, the number of iteration, etc. The chromosome of this algorithm is the hyperparameter of the machine learning. Each of the machine learning has different hyperparameter to optimize. In SVM, the chromosome of genetic algorithm will be Gamma, C, and kernel, in AdaBoost, they are number of estimator and

learning rate, then in Random Forest, the parameters are number of decision tree used, minimum sample leaf, and minimum weight fraction of leaf , and finally the number of K is the hyperparameter to optimize in KNN method. The crossover rate used is 0.5 and the mutation rate is 0.1. The tournament will be used as evaluation method to determine the next generation of the population. Lastly, the number of iteration used is 10 iteration. To implement this method, we use evolutionary algorithm search CV in Scikit-learn with population size 50 and generation number 1000.

Online News Dataset

↓

Splitting Dataset

↓

Genetic Algorithm Optimization

↓

Training the Model

↓

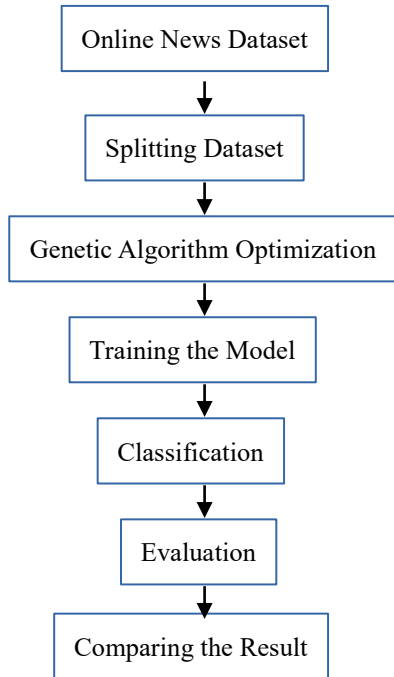Classification

↓

Evaluation

↓

Comparing the Result

Fig 1.    Diagram of the Proposed Method.

The next part is training the machine learning methods. The result of the genetic algorithm is an optimized hyper parameter of machine learning. The machine learning method using optimized hyper parameter will be trained using training data. We use Scikit-learn library for this. After that, the classification will be done using testing data in order to predict the popularity of online news. The result will be evaluated using several methods. They are accuracy, and the time needed to find the hyper parameter. The accuracy is obtained by using accuracy score library in Scikit-learn. The most important evaluation is the time because the proposed method is alternative hyper parameter tuning algorithm that has faster computational time.

### III.    RESULT AND DISCUSSION

The experiment in this research is comparing the accuracy and computation time of online news popularity prediction by using grid search and genetic algorithm. The result of the experiment is in Table 1 for implementation using Grid Search and Table 2 is Genetic Algorithm implementation. The time in this experiment is in second.  Fig. 2 is the chart of the evaluation of prediction using optimized hyper parameter. The evaluation method that is used in this research is the accuracy of machine learning methods using optimized hyper parameter.

The methods used these experiments are SVM, RF, AdaBoost, and KNN. Fig. 3 is the chart of the computational time needed for Grid Search and Genetic Algorithm to solve it. The time measurement is in second.

From the result of the experiment, the accuracy, of Grid Search and Genetic Algorithm is almost the same. Genetic Algorithm generates several solutions in one iteration. To get a new solution, Genetic Algorithm uses crossover and mutation method. Crossover is a combination of two solutions while mutation is a modification of one solution. This way of getting a new solution is derived from a process of evolution. The child is usually better than the parent, that's why it can be assumed that a combination or modification of solution can make a better solution result.

When it comes to computational time, Genetic Algorithm can achieve a much better result. From Fig. 3, the difference of computational time can be clearly seen. This make the prediction of online news popularity become faster. In Support Vector Machine, the time improvement to obtain the optimal hyper parameter is 7481 seconds. In the Random forest, 112 seconds is the time improvement result.  Adaptive Boosting has 6758 seconds improvement, and lastly, K - Nearest Neighbour is improved by 242 seconds. The entire machine learning methods has significant time improvement. This happened because genetic algorithm did not search all possible hyper parameter.

TABLE I.        HYPER PARAMETER GRID SEARCH RESULT

| Methods | Grid Search | |
| --- | --- | --- |
| | Accuracy | Time |
| SVM | 0.53 | 9241 |
| RF | 0.67 | 771 |
| AdaBoost | 0.66 | 7796 |
| KNN | 0.58 | 303 |

TABLE II.        HYPER PARAMETER GENETIC ALGORITHM RESULT

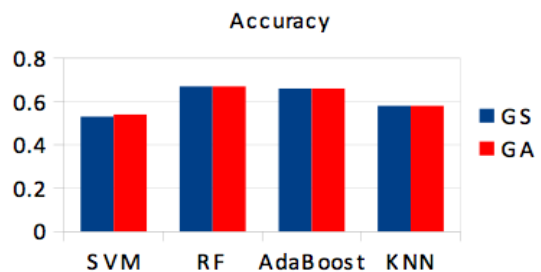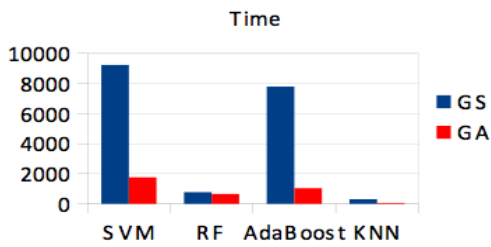| Methods | Genetic Algorithm | |
| --- | --- | --- |
| | Accuracy | Time |
| SVM | 0.54 | 1760 |
| RF | 0.67 | 659 |
| AdaBoost | 0.66 | 1038 |
| KNN | 0.58 | 61 |

Fig 2.    Accuracy of Prediction.

Fig 3.    Computational Time.

Genetic Algorithm uses crossover and mutation to get a new solution in each iteration. In each iteration, Genetic Algorithm will get a better solution. The iteration in Genetic Algorithm will be executed until a stop condition is met, such as the execution time and the number of iteration. This method makes Genetic Algorithm can get better hyper parameter when the iteration ends without trying all possible combination to get the best result. This makes Genetic Algorithm can have a faster computational time. For the next work, we can try deep learning to make an online news popularity prediction and use genetic algorithm to get the hyper parameter of deep learning methods for a better accuracy.

## IV.  CONCLUSIONS

The rapid usage of the internet makes online news become a popular source to obtain information. It is important to measure the popularity of online news prior to its publication. To solve this problem, we can use machine learning methods such as SVM, Random Forest, etc. The accuracy of machine learning methods' classification can be increased by tuning its hyper parameter. In this research, a genetic algorithm is proposed as hyper parameter tuning. The experiment is implemented using Scikit Learn library and the data used is the online news dataset downloaded from UCI machine learning site. This dataset has 39797 instances and 61 attributes.

Based on the experiment, it can be concluded that genetic algorithm can produce an optimal hyper parameter for machine learning with a reasonable amount of time. This happen because genetic algorithm can search for a better solution without trying all possible solution. It makes Genetic Algorithm a better replacement for Grid Search when the dataset that needs to be processed is very large.

## REFERENCES

[1]  L. H. Cheeks and A. Gaffar, "A Social Influence Model for Exploring Double Subjectivity through News Frames in Online News," p. 11, 2017.

[2]  K. Fernandes, P. Vinagre, and P. Cortez, "A Proactive Intelligent Decision Support System for Predicting the Popularity of Online News," in Progress in Artificial Intelligence, vol. 9273, F. Pereira, P. Machado, E. Costa, and A. Cardoso, Eds. Cham: Springer International Publishing, 2015, pp. 535–546.

[3]  Q. Hu, G. Wang, and P. S. Yu, "Public Information Sharing Behaviors Analysis over Different Social Media," 2015, pp. 62–69.

[4]  A. Tatar, M. D. de Amorim, S. Fdida, and P. Antoniadis, "A survey on predicting the popularity of web content," Journal of Internet Services and Applications, vol. 5, no. 1, Dec. 2014.

[5]  W. Alswiti and A. Rodan, "Features Selection Effect on Predicting the Popularity of Online News," Proceedings of the New Trends in Information Technology (NTIT-2017), p. 5, Apr. 2017.

[6]  R. Shreyas, D. Akshata, B. Mahanand, B. Shagun, and C. Abhishek, "Predicting popularity of online articles using Random Forest regression," 2016, pp. 1–5.

[7]  A. O. Adeleke, N. A. Samsudin, A. Mustapha, and N. Nawi, "Comparative Analysis of Text Classification Algorithms for Automated Labelling of Quranic Verses.," International Journal on Advanced Science, Engineering and Information Technology, vol. 7, no. 4, p. 1419, Aug. 2017.

[8]  M. Bahrololum, E. Salahi, and M. Khaleghi, "Machine Learning Techniques for Feature Reduction in Intrusion Detection Systems: A Comparison," 2009, pp. 1091–1095.

[9]  S. Srivastava, N. Joshi, and M. Gaur, "A Review Paper on Feature Selection Methodologies and Their Applications," p. 5.

[10]  H. Ren and Q. Yang, "Predicting and Evaluating the Popularity of Online News," p. 5.

[11]  S. Khalid, T. Khalil, and S. Nasreen, "A survey of feature selection and feature extraction techniques in machine learning," 2014, pp. 372–378.

[12]  F. Mar'i and A. A. Supianto, "Clustering Credit Card Holder Based on Billing Payment using Improved K-Means with Particle Swarm Optimization," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 5, no. 6, pp.737-744, Nov 2018.

[13]  A. Kathal and M. Namdev, "Correlation Enhanced Machine Learning Approach based Online News Popularity Prediction," p. 6.

[14]  I. Syarif, A. Prugel-Bennett, and G. Wills, "SVM Parameter Optimization using Grid Search and Genetic Algorithm to Improve Classification Performance," TELKOMNIKA (Telecommunication Computing Electronics and Control), vol. 14, no. 4, p. 1502, Dec. 2016.

[15]  M. Z. Sarwani, A. Rahmi, and W. F. Mahmudy, "An Adaptive Genetic Algorithm for Cost Optimization of Multi-Stage Supply Chain," vol. 9, no. 2, p. 6.

[16]  W. Li, X. Xing, F. Liu, and Y. Zhang, "Application of Improved Grid Search Algorithm on SVM for Classification of Tumor Gene," International Journal of Multimedia and Ubiquitous Engineering, vol. 9, no. 11, pp. 181–188, Nov. 2014.

[17]  Rang, "Poetry Classification Using Support Vector Machines," Journal of Computer Science, vol. 8, no. 9, pp. 1441–1446, Sep. 2012.

[18]  I. Oktanisa and A. A. Supianto, "A Comparison of Classification Techniques in Data Mining for Bank Direct Marketing," Jurnal Teknologi Informasi dan Ilmu Komputer (JTIIK), vol. 5, no. 5, pp. 567-576, Oct 2018.

[19]  A.A. Soebroto and A.A. Supianto, "Development of Decision Support Systems for Selection of Bali Cow Superior Seed Using the K-Nearest Neighbor Method," Journal of Environmental Engineering and Sustainable Technology, vol. 2, no. 1, pp.49-57, Jul 2015.

[20]  P. Lameski, E. Zdravevski, R. Mingov, and A. Kulakov, "SVM Parameter Tuning with Grid Search and Its Impact on Reduction of Model Over-fitting," in Rough Sets, Fuzzy Sets, Data Mining, and Granular Computing, vol. 9437, Y. Yao, Q. Hu, H. Yu, and J. W. Grzymala-Busse, Eds. Cham: Springer International Publishing, 2015, pp. 464–474.

[21]  L. Lin, Z. Xiaolong, Z. Kai, and L. Jun, "Bilinear Grid Search Strategy Based Support Vector Machines Learning Method," p. 8.

[22]  M. Litvak, M. Last, and M. Friedman, "A New Approach to Improving Multilingual Summarization Using a Genetic Algorithm," in: ACL, The Association for Computer Linguistics, pp. 927–936, 2010.

[23]  S. Chawla, "Application of genetic algorithm and backpropagation neural network for effective personalize web search-based on clustered query sessions," International Journal of Applied Evolutionary Computation (IJAEC), vol. 7, no. 1, pp. 33-49, 2016.

[24]  H. J. Escalante, M. A. García-Limón, A. Morales-Reyes, N. Graff, M. Montes-y-Gómez,E. F. Morales, and J. Martínez-Carranza, "Term-weighting learning via genetic programming for text classification," Knowledge-Based Systems, 83, pp.176-189, 2015.

[25]  A. Jamshidnezhad, "A Classifier Model based on the Features Quantitative Analysis for Facial Expression Recognition," p. 4.

[26]  F. Pedregosa et al., "Scikit-learn: Machine Learning in Python," Machine Learning in Python, p. 6.