International Journal of Electrical and Computer Engineering (IJECE)
Vol. 11, No. 3, June 2021, pp. 2275~2284
ISSN: 2088-8708, DOI: 10.11591/ijece.v11i3.pp2275-2284

Performance analysis of sentiments in Twitter dataset using SVM models

Lakshmana Kumar Ramasamy[1], Seifedine Kadry[2], Yunyoung Nam[3], Maytham N. Meqdad[4]
[1]Hindusthan College of Engineering and Technology, Coimbatore, Tamil Nadu
[2]Department of Mathematics and Computer Science, Faculty of Science, Beirut Arab University, Beirut, Lebanon
[3]Department of Computer Science and Engineering, Soonchunhyang University, Asan, South Korea
[4]Al-Mustaqbal University College, Hillah, Babil, Iraq

Article Info

Article history:

Received Dec 26, 2019
Revised Oct 8, 2020
Accepted Nov 2, 2020

Keywords:

Machine learning
Sentiment analysis
SVM model
Twitter dataset

ABSTRACT

Sentiment analysis is a current research topic by many researches using supervised and machine learning algorithms. The analysis can be done on movie reviews, twitter reviews, online product reviews, blogs, discussion forums, Myspace comments and social networks. The Twitter data set is analyzed using support vector machines (SVM) classifier with various parameters. The content of tweet is classified to find whether it contains fact data or opinion data. The deep analysis is required to find the opinion of the tweets posted by the individual. The sentiment is classified in to positive, negative and neutral. From this classification and analysis, an important decision can be made to improve the productivity. The performance of SVM radial kernel, SVM linear grid and SVM radial grid was compared and found that SVM linear grid performs better than other SVM models.

Corresponding Author:

Yunyoung Nam
Department of Computer Science and Engineering
Soonchunhyang University
Asan 31538, South Korea
Email: ynam@sch.ac.kr

1. INTRODUCTION

Sentiment analysis [1] is a current ongoing research in the field of text mining analysis. Many people post their views, opinion and ideas in unstructured format. The views are taken from the views of public, customer, social media, entertainment, sports, climate analysis and Industrial organization. Millions and billions of people and public are using social network websites such as Facebook, Twitter, Google plus and so on. The social media [2] generates a huge volume of sentiment data in the various forms such as tweet id, status updates, reviews, author, content, tweets type and tweets status update. As the data size is going larger and larger, it is necessary to analyze and categorize the sentiment reviews or opinion of the various people to predict. Machine learning techniques [3] is one of the frequently used techniques in sentiment data analysis to classify the tweets or author comments in the form of positive, negative and neutral based on the tweet data. SVM is a classifier algorithm which separates hyperplane [4-6]. It is defined as;

$$f(x) = wt^T x + b \tag{1}$$

where wt is the weight factor and b denotes bias of the function.

The optimal hyperplane of SVM is defined as;

$$|wtTx + b| = 1 \quad (2)$$

In (2), x is a training set. The distance is calculated as;

$$Distance = |wt^T x + b| / ||wt|| \quad (3)$$

The margin is represented as *M*, is the distance with closeness.

$$M = 2/||wt|| \quad (4)$$

The SVM kernel is a function defined as a dot product of data inputs. The frequently used kernel functions are;

- Linear
- Polynomial
- RBF and
- Sigmoid

The linear kernel function is;

$$Kernel\ (X_i\ , X_j)\ =\ X_i\ . X_j \quad (5)$$

The research paper is prepared and organized as: The detailed background study was discussed in section 2. Section 3 focuses methodology used in this work, section 4 provides the experimental results and conclusion is given in section 5.

2. REVIEW OF LITERATURE

Nadia *et al.* [7] developed a method which classifies the sentiment of Twitter data with the use of lexicons and classifier groups. Different public Twitter sentiment data are experimented using SVM, Naive Bayes, Random Forest, multinomial Naive Bayes and logistic regression to expand the classification accuracy. Medhat *et al.* [8] survey the recent techniques used in analysis of sentiments and 54 articles were classified and summarized. By analysing the article, author was given clear picture about sentiment classification and feature selection were used in the field of research. To solve the sentiment classification SVM and Naive Bayes were commonly used in the machine learning algorithms.

Ramasamy *et al.* [9] discusses the responsibility of text cleaning or pre-processing in sentiment analysis data, and the experimental results were done using support vector machines (SVM) with appropriate feature selection. Bifet *et al.* [10] describes the usage of Twitter data provided by Firehouse API, provides all messages user which are publicly available in internet or real-time. The author analyzed the data set with the algorithms

- Multinomial Naive Bayes
- Stochastic gradient descent and
- Hoeffding tree.

Nurulhuda *et al.* [4] developed a method to classify the Twitter based sentiment using PCA. This is combined with SentiWordNet lexicon based method and included with SVM. Shiyang [11] proposed a sophisticated neural network approach for analyzing Twitter data to perform sentiment analysis. Abinash *et al.* [5] compares the experimental result of Naive Bayes and SVM on polarity movie dataset. The training and testing data is classified based on positive and negative review by applying the algorithms. Albert and eibe, [12] discuss the challenges faced during the sentiment classification of Twitter dataset flow and the author proposed the method called sliding window based kappa value statistics to evaluate the time changing based data streams. Tan [13] performed the sentiment classification on Chinese documents by applying the numerous feature or attribute selection techniques such as IG, CHI, DF, and MI and other classification techniques KNN, SVM, Naive Bayes, and centroid classifier. Ahmed proposed [6] the variable selection method entropy by using weighted genetic algorithm (EWGA) and the experimental result was done with the benchmark movie dataset. The results were compared with SVM and EWGA to indicate the raise in the level of performance. The author worked [14] with the data mining classification algorithms such as;

- SVM
- Maximum entropy
- Naive Bayes

And achieved the accuracy of 80% and above and experiment was done using tweets with emoticon data. The pre-processing is must in order to increase the accuracy. The author [15] proposed the method to repeatedly detect the sentiment analysis of Twitter data and the experimental result shows the more abstract features are captured, in the removal of noisy and more robust about biased. The author [16] observed the sentiment using Twitter dataset and used the POS specific prior polarity features and tree kernel.

The author [17] contributed the novel investigation of stacked SVM based classification techniques to categorize the some of the user attributes and they performed the complete analysis of components and features. The author proposed [18] the technique to take out the Twitter data related to influenza using Twitter API and classify the influenza patients using the support vector machine. The author portrays the Twitter data which shows the real world, and natural language processing methods was applied. The author extracted tweets which has most needful information. The author proposed [3] the sliding window kappa statistic to estimate the time-changing data flow and survey the Twitter dataset using the machine algorithms for various data flows.

The author [19] presented the machine learning to analysis the sentiment data based review. Blog and forum texts data available in the WWW. The dataset is manually classified based on the tweet statements as the negative, positive and neutral. The author [2] investigates the responsibility of text pre-processing in sentiment analysis and experimental result explain with the suitable feature selection and demonstration using SVM. The author proposed [20] a new methodology for sentiment analysis for Facebook dataset to mine the user's sentiment categories and to notice major emotional modification. The author presents [1] the experimental result by comparing the SVM algorithm and ANN concerning for the document-level to analyze sentiment. The author contributes [21] the use of two different classifier approaches such as neutral-polar classifier and positive negative or polarity classifier for the experimental analysis

3. METHODOLOGY

Support vector machines (SVM) is a supervised machine learning method. This is associated closely with learning algorithms which analyze data in the data set used for classification. SVM model separates and constructs hyper plane. This hyper plane can be used for classification. SVM is a maximum margin classifier. The mapping is done through kernel functions. Some of the kernel functions used in this work are given. The linear kernel functions is used to deal large space data vectors and it is represented as;

$$k(x, x') = (x, x') \tag{6}$$

The Gaussian radial basis function is a general purpose kernel which is;

$$k(x, x') = exp(-sigma||x - x^2) \tag{7}$$

The radial basis function is denoted as;

$$K(x_i, x_j) = exp(-\gamma|x_i - x_j|2) \tag{8}$$

An important issue of SVM classifier is the selection of various parameters. The parameter tuning influences the effectiveness and the efficiency of the classifier. The cost parameter C finds the penalty for misclassifications. The frame work of proposed work is given in Figure 1.

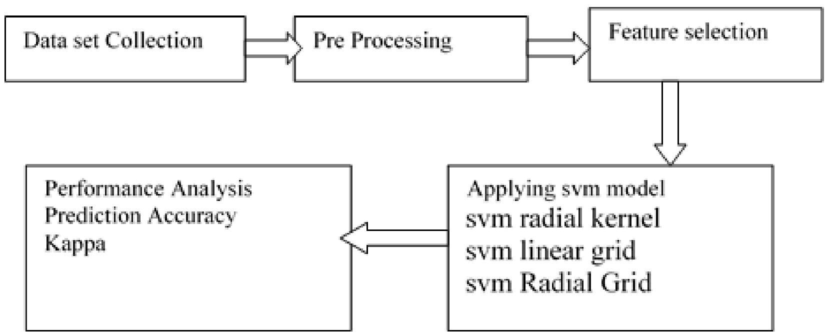

Figure 1. Proposed frame work

4. EXPERIMENTAL RESULTS

There are three different classes of sentiments in the Twitter data set. They are:
- Positive
- Negative
- Neutral

Positive sentiments refers to positive thinking nature of the person. The positive emotion sentiments are surprise, love, affection, happiness, joy and smile. The positive tweet reviews, opinions and sentiments will create a happy environment and good for individual as well as for the society and country. Negative Sentiments reflects negative nature of the individual. The negative sentiments will provide sadness, worry, jealous, and hate. The negative sentiments of person will not be happy and it affects the society. Neutral Sentiments reflects no participation. The person is not satisfied and could not take any decision. This is also very important for analysis. This emotion will also affect the society without creating more impacts.

The Twitter data is obtained from publically available internet. The 1000 tweets were taken for analysis and split into 70% Training set and 30% test set. The Twitter dataset is in the form of CSV file. CSV file is looking like a table structured spreadsheet program of Microsoft Excel. A tweet is a user's opinion and is expressed emotionally by different people. The Twitter dataset used in this work is labelled into three classes viz. neutral, negative and positive. The data may be redundant, inconsistent, unwanted blank spaces, special characters and non-related information. The pre-processing is essential to clean the data and bring into structured one. The pre-processing [19] was performed;
- Removed all punctuations, @, _ symbols and numbers
- Removed sequence of repeated characters
- Replaced all the emotions with their Sentiment value.
- Replaced all missing values with NAN
- Removed stop words
- Removed unnecessary white spaces

The machine learning algorithms is used in sentiment analysis for the purpose of classification. The supervised machine learning classification is made for sentiments in the data set. The data set is divided into training set and test set in machine learning algorithms, there are many algorithms which acts main role in sentiment analysis.
- Naive Bayes
- Maximum entropy
- Support vector machines

SVM and RBF kernel support vector machine is no probabilistic statistical analysis algorithm which is used to separate data linearly and nonlinearly. Here dataset is a set of attributes with its values. $Dataset = \{Xi, Yi\}$, Xi is set of data present in the dataset as tuple and Yi is class label of data in the form of tuples. Class labels are 0, 2 and 4 for neutral, negative and positive category respectively. The objective of SVM is to classify or separate the sentiments as neutral, negative and positive training by finding $n - 1$ hyper plane. Quadratic programming (QP) problem is essential to solve linear data using SVM model. A classifying hyper plane is written as: $wt * x + b = 0$, w being weight vector of n number of attributes, x is set of data present in the dataset as tuple and b is bias. A linear classifier has the form $f(x) = wt^T x + b$.

Feature selection is a selection of relevant variable or feature available in the dataset. It is also known as attributing selection. This is the initial process that is carried out before model construction. Identifying relevant feature using feature selection technique simplifies the task. The strongly correlated attribute provides the idea of classification on what basis. Like feature selection, feature extraction can also be used. Feature extraction generates new features from existing features, whereas feature selection creates and returns a new subset of the existing features. If the dataset contains many features and few samples, then there is a necessary to select features. The boruta feature selection method was performed in this work. The advantage of feature selection is
- Models can be simplified
- Training time is minimized
- The dimensionality nuisance is minimized
- Over fitting is reduced

Boruta is a feature selection method was proposed to find whether the feature is either strong or weak. Gsub() function was used to replace any unwanted expression into wanted one. Missing values were checked and are replaced by NAN. Blank spaces are cleaned. The format of boruta package is same as the format of linear regression lm method. Boruta provides a well significant call of variables in a data set. There are 5 attributes, 3 attributes were rejected and 2 attributes were confirmed. Some tentative attributes will be created by boruta package which is also having importance when the original features are not able to take any decision. Plotting of the boruta variable is shown in importance chart. The attributes of data set were given in x axis and its importance in y axis. The minimum average score and maximum Z score of attributes are tentative attributes generated by boruta feature selection package. The tentative attributes will be categorized as either confirmed or rejected by comparing the median Z score of the feature attributes. The feature selection of attributes is shown in Figure 2.

The random forest feature selection algorithm is functioning through *rfFuncs* option was performed. The top one attribute out of five attributes was selected by this method. The *rfetrain* was plotted and is shown in Figure 3. The order of importance of attributes in feature selection by boruta is shown in Figure 4.

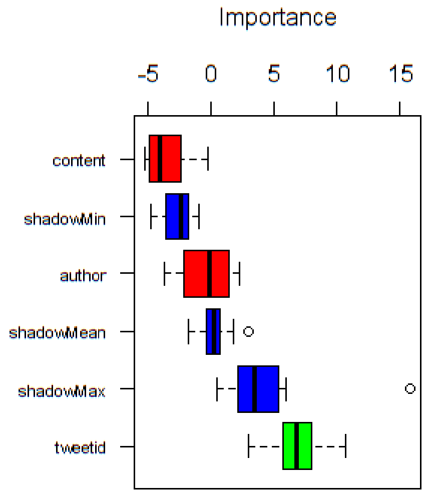

Figure 2. Feature selection of attributes

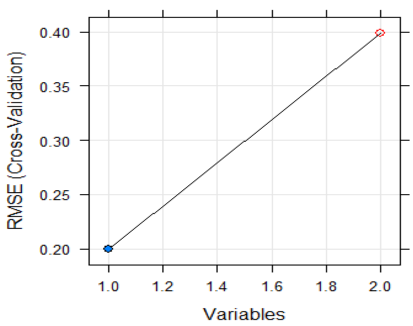

Figure 3. Random forest selection

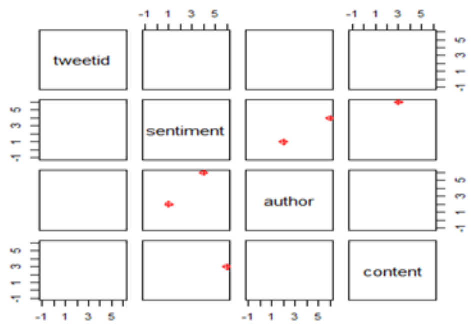

Figure 4. Random forest selection of attributes

SVM model linear kernel was implemented to classify the sentiments in the Twitter data set. Feature selection was performed by boruta and two attributes confirmed important which are tweetid and sentivalue. SVM model was plotted with linear kernel method and is shown in Figure 5. The confusion matrix is used to calculate the accuracy and summarize the performance of data. It gives clear idea about the efficiency of the model. It provides user with how much is achieved and how much went wrong. The confusion matrix of above model is shown in Figure 6.

The performance of the above model was not satisfied, the improper prediction was performed in the SVM model, and therefore the minimum accuracy of 15% was achieved. So, the tuning of parameter is necessary for improving the performance. The next SVM model of c classification with radial based kernel was implemented. SVM model c classification with radial kernel was applied for the same 1000 tweets in the Twitter data set with c type classification. This model classifies the data set sentiment value as 578 negative sentiments, 247 neutral sentiments and 175 positive sentiments which are shown in Table 1. This model has 701 training samples and 299 testing samples with four predictors and three classes. The 10 fold cross validated was performed for three times for the sample sizes 631. The tuning parameter c made constant with an assigned value of 1. The accuracy and kappa values are shown in the Table 2.

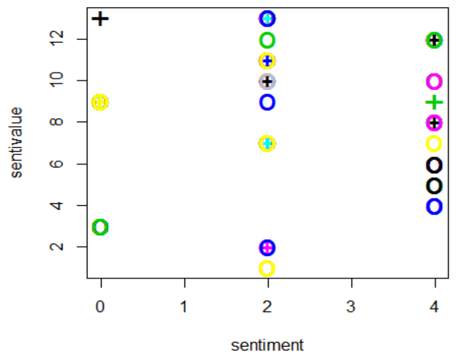

Figure 5. SVM classification

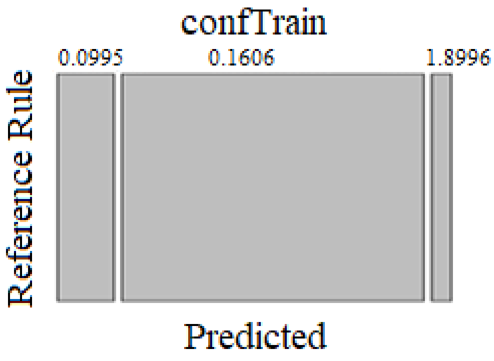

Figure 6. Confusion matrix of SVM

Table 1. Classification table with radial kernel

Pred	Negative	Neutral	Positive
	578	247	175

Table 2. Classification prediction accuracy

Accuracy	Kappa
0.8497393	0.7063484

Kappa processes the % of data set values in the diagonal element of the table. These diagonal values are adjusted for the agreement either accept or reject. Kappa value is calculated by the observed level of agreement Table 3;

$$K_0 = K_{11} + K_{22} \quad (9)$$

This value needs to be compared to the value that you would expect if the two raters were totally independent,

$$K_e = K_1K_1 + K_2K_2 \quad (10)$$

The Kappa value is defined to be;

$$Kappa = K_0 – K_e/1 - K_e \quad (11)$$

The kappa result can figure out. The predicted accuracy of three classes sentiment counts were given in the table and are shown in Table 4.

Table 3. Kappa agreement

Range	Agreement
< 0.20	Poor
0.20-0.40	Fair
0.40-0.60	Moderate
0.60-0.80	Good
0.80-1.00	Very good

Table 4. Prediction accuracy

Pred	0	2	4
Neutral	78	0	0
Negative	0	168	45
Positive	0	1	7

The predicted accuracy of repeated cross validation and cost is shown in Figure 7. The accuracy is 84% and the kappa is 70%. The kappa value indicates that there is a good agreement between individuals. SVM_linear_grid is another model performed to increase the accuracy with 76 samples, 4 predictors and 3 classes. The resampling of dataset is cross-validated with 10 fold and repeated 3 times with of sample sizes 69, 67 and 68. The accuracy and kappa values of SVM_linear_grid is shown in Table 5 and Figure 8.

Accuracy is normally used to identify the optimal model using the high score value. The final and last value used for the model was $c = 0.01$. The accuracy is improved from 84% to 89%. The accuracy for various c values and cross validation is shown in figure. SVM linear grid plot was applied in the same data set. The accuracy of predicted classification with SVM linear grid method is 89%.

Table 5. Prediction accuracy of SVM_linear_grid

C	Accuracy	Kappa
0.00	NaN	NaN
0.01	0.8936508	0.7973095

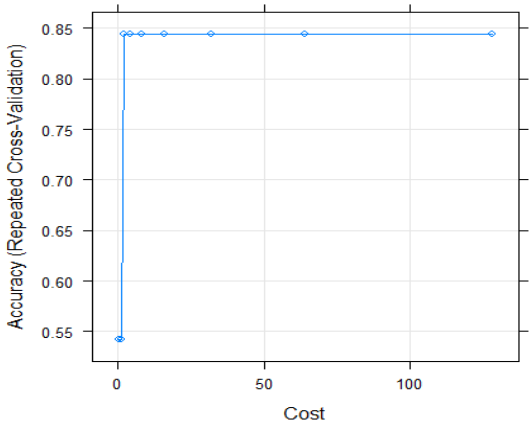

Figure 7. SVM radial kernel classification accuracy

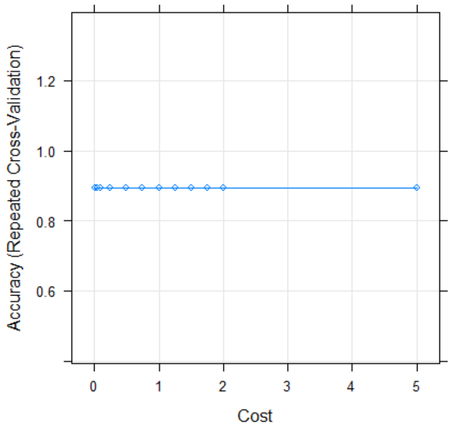

Figure 8. SVM linear grid accuracy

The setup for implementing support vector machines with radial basis function kernel uses 76 samples, 4 predictors and 3 classes of resampling of 10-fold cross validation and is repeated 3 times was applied for varying cost value. The Tuning parameter 'sigma' was held constant at a rate of 0.003044357. The final values used in this model were $sigma = 0.003044357$ and c=2. The accuracy obtained was 84%. The accuracy was plotted and is shown in Figure 9.

SVM_radial_grid Figure 10 method is applied with 76 samples of 4 predictors and 3 classes. SVM with radial basis function kernel summary sample size is 69, 68, 69, 68 and 68. The prediction Accuracy of SVM radial grid is shown in Tables 6 and 7. The values used for the model were sigma as 0.01 and c as 1.5. The accuracy 84% was maintained. The performance of various SVM model Statistics by three Classes is given in the Table 8.

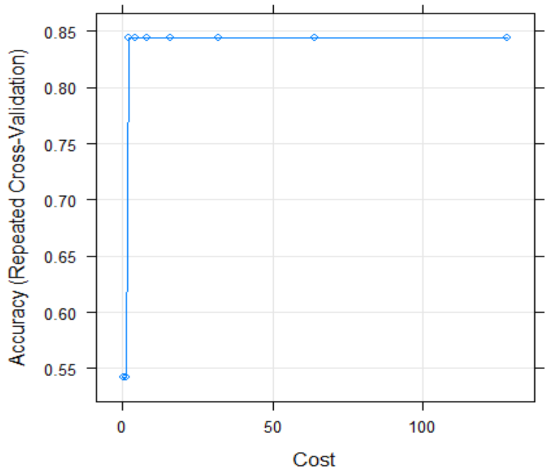

Figure 9. SVM radial basis function kernel accuracy

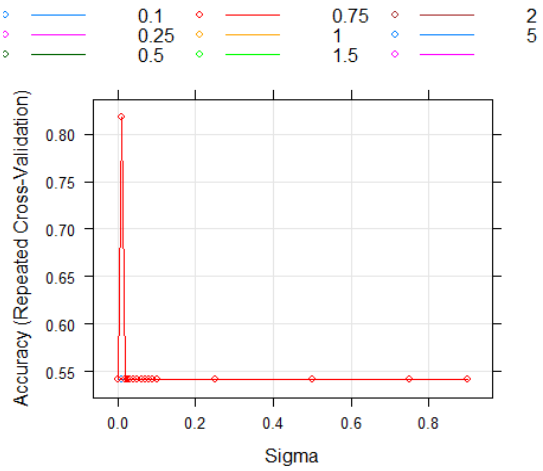

Figure 10. SVM radial grid accuracy

Table 6. Prediction accuracy of SVM radial grid

prediction	0	2	4
0	7	0	0
2	3	16	5
4	0	0	0

Table 7. Prediction accuracy of svm_linear_grid

Accuracy	Kappa
0.8419	0.5108

Table 8. Comparative analysis of SVM models

Pred	0	2	4
Sensitivity			
SVM radial kernel	1.0000	1.0000	0.13462
SVM linear grid	0.7000	1.0000	0.20000
SVM Radial Grid	0.7000	1.0000	0.0000
Specificity			
SVM radial kernel	1.0000	0.6538	0.99595
SVM linear grid	1.0000	0.5333	1.00000
SVM Radial Grid	1.0000	0.4667	1.0000
Pos Pred Value			
SVM radial kernel	1.0000	0.7887	0.87500
SVM linear grid	1.0000	0.6957	1.00000
SVM Radial Grid	1.0000	0.6667	NaN
Neg Pred Value			
SVM radial kernel	1.0000	0.9884	0.84536
SVM linear grid	0.8750	1.0000	0.86667
SVM Radial Grid	0.8750	1.0000	0.8387
Prevalence			
SVM radial kernel	0.2609	0.5652	0.17391
SVM linear grid	0.3226	0.5161	0.16129
SVM Radial Grid	0.3226	0.5161	0.1613
Detection Rate			
SVM radial kernel	0.2609	0.5619	0.02341
SVM linear grid	0.2258	0.5161	0.03226
SVM Radial Grid	0.2258	0.5161	0.0000
Detection Prevalence			
SVM radial kernel	0.2609	0.7124	0.02676
SVM linear grid	0.2258	0.7419	0.03226
SVM Radial Grid	0.2258	0.7742	0.0000
Balanced Accuracy			
SVM radial kernel	1.0000	0.8240	0.56528
SVM linear grid	0.8500	0.7667	0.60000
SVM Radial Grid	0.8500	0.7333	0.5000

True positive rate is test sensitivity and true negative rate is test specificity. Sensitivity measures the correctly identified positives. Specificity measures the correctly identified true negatives. The performance analysis of the algorithms can be estimated by using two measures sensitivity S and positive predictive value (PPV). They are very much useful for estimation of performance [22-25]. Sensitivity S is calculated and measured by;

$$S = TP/(TP + FN), \tag{12}$$

Specificity S is measured as;

$$SP = TN/(TN + FP) \tag{13}$$

– TP is a number of really and truly identified relevant features recognised by an algorithm;
– FN is a number of attributes not identified by an algorithm
– FP is a number of not relevant attributes that are not correctly recognised
– Positive predictive value PPV is measured and calculated as

$$PPV = TP/(TP + FP) \tag{14}$$

Negative predictive value NPV is calculated and measures measured by;

$$NPV = TN/(TN + FP) \tag{15}$$

The SVM classifier with SVM kernel, SVM linear grid, SVM radial grid were performed for Twitter data set and performance was analyzed with various tuning parameters. SVM linear grid achieves better performance than the other. The basic SVM model without parameter tuning achieves minimum level accuracy. The overall performance is presented in Table 9. From the experimental results, it is observed that, the performance of SVM linear grid model shown better results which is shown in Figure 11.

Table 9. Comparative analysis of accuracy of SVM models

Model	Accuracy	Kappa
SVM	0.15	0.12
SVM radial kernel	0.8497393	0.7063484
SVM linear grid	0.8936508	0.7973095
SVM Radial Grid	0.8419	0.5108

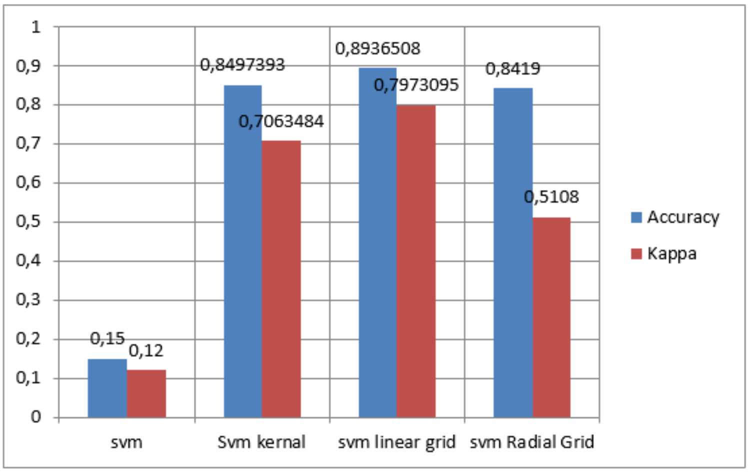

Figure 11. Accuracy of SVM models

The performance measures were selected to compare the SVM classifier with various parameter tuning model, the highest accuracy of 89.36% of SVM linear grid, 84.97% of SVM kernel, 84.19 of SVM radial grid was achieved. In SVM, the values of c and gamma were selected using 5-fold cross validation and 10-fold cross validation.

5. CONCLUSION

The experimental results of SVM classifier attains satisfied results in the classification when compare with previous researches. There are many performance measures of SVM model is available. In this work, the parameter tuning of various SVM model was proposed. SVM model Cross Validated with 10 fold and looping through 3 times with of mean sample sizes 69, 67 and 68 out performs than the other models. The accuracy achieved was satisfied because Twitter posts are emotional one. The people post their emotions when they are not feeling comfort. Neutral data plays more important role, because the people could not take any decision. The concentration towards the neutral people will reflect either positive or negative. This idea is suit for election when the people are not ready to vote for any particular option. The future work will focus on implementing different algorithms with different dataset.

ACKNOWLEDGEMENTS

This work was supported by the Soonchunhyang University Research Fund.

REFERENCES

[1] V. Pandey and C. Iyer, "Sentiment analysis of micro blogs," *Encyclopedia of Social Network Analysis and Mining*, pp. 1-17, 2009.
[2] A. Ortigosa, J. M. Martín, R. M. Carro, "Sentiment analysis in Facebook and its application to e-learning," *Computers in Human Behavior*, vol. 31, pp. 527-541, 2014.
[3] E. Boiy and M. F. Moens, "A machine learning approach to sentiment analysis in multilingual Web texts," *Information retrieval*, vol. 12, pp. 526-558, 2009.
[4] Nurulhuda Zainuddin, Ali Selamat, Roliana Ibrahim, "Twitter Feature Selection and Classification Using Support Vector Machine for Aspect-Based Sentiment Analysis," *International Conference on Industrial, Engineering and Other Applications of Applied Intelligent Systems*, vol. 9799, pp. 269-279, 2016.
[5] Abinash Tripathy, Ankit Agrawal, Santanu, Kumar Rath, "Classification of Sentimental Reviews Using Machine Learning Techniques," *Procedia Computer Science*, vol. 57, pp. 821-829, 2015.
[6] Ahmed Abbasi, Arab Salem, Hsinchun Chen, "Sentiment analysis in multiple languages: Feature selection for opinion classification in Web forums," *ACM Transactions on Information Systems (TOIS)*, vol. 26, no. 3, 2008.
[7] Nadia F. F., da Silva, Eduardo R. Hruschka, Estevam R. Hruschka Jr., "Tweet sentiment analysis with classifier ensembles," *Decision Support Systems*, vol. 66, pp. 170-179, 2014.
[8] Walaa Medhat, Ahmed Hassan, Hoda Korashy "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014.
[9] L. K. Ramasamy, S. Kadry, S. Lim, "Selection of optimal hyper-parameter values of support vector machine for sentiment analysis tasks using nature-inspired optimization methods," *Bulletin of Electrical Engineering and Informatics (BEEI)*, vol. 10, no. 1, pp. 290-298, 2020.
[10] Bifet and E. Frank, "Sentiment Knowledge Discovery in Twitter Streaming Data," in *Proceedings of the 13th International Conference on Discovery Science*, vol. 6332, 2010, pp. 1-15.
[11] Shiyang Liao, Junbo Wang, Ruiyun Yu, Koichi Sato, Zixue Cheng, "CNN for situations understanding based on sentiment analysis of twitter data," *Procedia Computer Science*, vol. 111, pp. 376-381, 2017.
[12] Walaa Medhat, Ahmed Hassan, Hoda Korashy "Sentiment analysis algorithms and applications: A survey," *Ain Shams Engineering Journal*, vol. 5, no. 4, pp. 1093-1113, 2014.
[13] Songbo Tan, and Jin Zhang, "An empirical study of sentiment analysis for chinese documents," *Expert Systems with Applications*, vol. 34, no. 4, pp. 2622-2629, 2008.
[14] A. Go, R. Bhayani, L. Huang, "Twitter sentiment classification using distant supervision," *ACM digital library, COLING '10 Proceedings of the 23rd International Conference on Computational Linguistics*, 2010, pp. 36-44.
[15] A Madi, O. K. Zein, S Kadry, "On the improvement of cyclomatic complexity metric" *International Journal of Software Engineering and Its Applications*, vol. 7, no. 2, pp. 67-82, 2013.
[16] Abhishek Shreevats and Manaswi Gustav, "Classifying latent user attributes in twitter," *SMUC '10 Proceedings of the 2nd international workshop on Search and mining user-generated contents*, 2010, pp. 37-44.
[17] Eiji Aramaki, "Twitter catches the flu: detecting influenza epidemics using Twitter," *EMNLP '11 Proceedings of the Conference on Empirical Methods in Natural Language Processing*, 2011, pp. 1568-1576.
[18] M. N. Meqdad, R. Al-Akam, S Kadry, "New prediction method for data spreading in social networks based on machine learning algorithm," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 18, no. 6, pp. 3331-3338, 2020.
[19] E. Haddi, X. Liu, Y. Shi, "The role of text pre-processing in sentiment analysis," *procedia computer science*, vol. 17, pp. 26-32, 2013.
[20] R. Moraes, J. O. F. Valiati, W. P. G. O. Neto, "Document-level sentiment classification: An empirical comparison between SVM and Ann," *Expert Systems with Applications*, vol. 40, no. 2, pp. 621-633, 2013.
[21] X. Li, Z. Deng, Z. Chen and Q. Fei, "Analysis and Simplification of Three-Dimensional Space Vector PWM for Three-Phase Four-Leg Inverters," in *IEEE Transactions on Industrial Electronics*, vol. 58, no. 2, pp. 450-464, 2011.
[22] Mohammed Zuhair Al-Taie, Seifedine Kadry, Joel Pinho Lucas, "Online data preprocessing: a case study approach," *International Journal of Electrical and Computer Engineering (IJECE)*, vol. 9, no. 4, pp. 2620-2626, 2019.
[23] Kaushik Sekaran, P. Chandana, J. Rethna Virgil Jeny, Maytham N. Meqdad, S. Kadry, "Design of optimal search engine using text summarization through artificial intelligence techniques," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 18, no. 3, pp. 1268-1274, 2020.
[24] H. Kadhim Tayyeh, and A. Sabah Ahmed Al-Jumaili, "Classifying confidential data using SVM for efficient cloud query processing," *TELKOMNIKA Telecommunication Computing Electronics and Control*, vol. 17, no. 6, pp. 3155-3160, 2019.
[25] Seifedine Kadry and Rafic Younès, "Etude Probabiliste d'un Systeme Mecanique a Parametres Incertains par une Technique Basee sur la Methode de transformation," *Conference: Canadian Congress of Applied Mechanics*, Montréal, Canada, 2005.