ISSN 2278-3091
Volume 9, No.5, September - October 2020
International Journal of Advanced Trends in Computer Science and Engineering
Available Online at http://www.warse.org/IJATCSE/static/pdf/file/ijatcse87952020.pdf
https://doi.org/10.30534/ijatcse/2020/87952020

Dimensionality Reduction for Classification of Filipino Text Documents based on Improved Bayesian Vectorization Technique

Hajah T. Sueno[1], Bobby D. Gerardo[2], Ruji P. Medina[3]
[1]Technological Institute of the Philippines-Quezon City, Philippines, qhsueno@tip.edu.ph
[2]West Visayas State University, Philippines, bgerardo@wvsu.edu.ph
[3] Technological Institute of the Philippines-Quezon City, Philippines, ruji.medina@tip.edu.ph

ABSTRACT

Dimensionality reduction of feature vector size plays a vital role in enhancing the text processing capabilities to reduce the size of the feature vector used in the mining tasks to achieve a higher classification accuracy. While dimensionality reduction for text classification is becoming a great area of research in most languages, Filipino documents have received little or no attention from researchers. Thus, this paper addresses the issue of dimensionality reduction in representing relevant data from Filipino texts using an improved Bayesian vectorization technique. To validate the effectiveness of improved Bayesian vectorization, the model was compared to the Term Frequency and Inverse Document Frequency (TF-IDF) method. The outcomes are presented using standard measures such as precision, recall, f-score, and accuracy. The results revealed that the improved Bayesian vectorization has significantly better results having 98% classification accuracy compared to 76% classification accuracy of the TF-IDF vectorization technique.

Key words : Dimensionality Reduction, Bayesian Vectorization, Filipino Text Documents, OPM Songs, Lyrics, Text Classification.

1. INTRODUCTION

With the growing availability of online text documents and the rapid growth of the World Wide Web, the task of classifying text documents is turning into interesting research [1]. There are many different types of information that may be extracted from the Internet as a source of data for different machine learning (ML) analysis including natural language processing. From the classification perspective, it is crucial to retain only those attributes which maximize the effectiveness of the classification. But, for text documents to be used in text classification, it needs to be processed and transformed from the text version to a document vector, making it much easier to manage and reduce the dimensionality of features [2]. There may be a lot of consistency in the way data is collected. These data may contain large quantities of unwanted information, making the analysis process very difficult. [3]. Dimension reduction is one of the important processes in text mining and enhances the performance of classification by reducing dimensions so that text mining procedures process text documents with a reduced number of features [4]. One ofthe first steps that were taken to solve this problem was to find a way to vectorize words [5][6]. Transforming textual data into meaningful vectors is a way of communicating with the machines to perform several machine learning tasks and mathematically resolving problems.

While dimensionality reduction for text classification is becoming a great area of research in most languages apart from English such as Chinese and Arabic, Filipino documents have received little or no attention from researchers [7]. Thus, this study aims to investigate the impact of the improvements made to Bayesian vectorization in our previous study [8] in transforming the Filipino text document of Original Pilipino Music (OPM) song lyrics into the numerical format of vector space probability distribution. Specifically, this research aims to gather Filipino song lyrics and annotate them based on their category, build a Filipino text Bayesian vectorizer and classifier, and evaluate classification performance with the use of the f-score, precision, recall, and accuracy metrics. To validate the performance of the Bayesian vectorization technique results are also compared over TF-IDF vectorization on the preprocessing of textual data for the SVM classifier.

The rest of the paper is organized as follows. Section II is a review of related literature. Section III discusses the methodology. Section IV presents the results and discussions. The paper ends with conclusions and further scope of work in Section V.

2. REVIEW OF RELATED LITERATURE

Several works study the impact of dimensionality reduction applied to document datasets to show significant achievement in dimension reduction by eliminating unnecessary and redundant features in high-dimensional data to enhance the classification accuracy [9][10][11][12]. Decreasing the

dimensionality is an effective way to downsize the data [13]. It is a method that tries to predetermine a set of high dimensional vectors into a space of lower dimensionality while retaining metrics among them [14].

In [5], dimensionality reduction techniques have been employed in the processing of high-dimensional data for a variety of feature extraction algorithms for unsupervised learning, supervised learning, and linear and non-linear applications. Patil and Sane [15] conducted a relative study for effective classification after data reduction. They examined reduction approaches in brief with the results of an accuracy correlation after dimension reduction. They used fuzzy rough methodology and the outcomes showed that fuzzy rough feature selection enhances the accuracy of artificial neural system classifiers. In [16][17], a survey and a comparative study of the dimensionality reduction techniques were introduced for the classification of text documents. It focuses on the filter approach to achieve a reduction in dimensionality and techniques to improve classification accuracy and save on feature size. In [18], it compares the output of Document Frequency, TF-IDF, Term Frequency Variance as feature selection methods and Latent Semantic Analysis (LSA), Random Projection (RP) and Independent Component Analysis (ICA) as dimensionality reduction methods; on the one hand, it tests each method.

Previous works introduced the use of vectorization to reduce the dimensionality of the dataset and improve the accuracy of classification tasks. In the case of TF-IDF vectorization, some others have introduced the use of the TF-IDF vectorization model to capture the semantic similarity of news articles to identify unifiable groups by using the k-means algorithm to cluster the vectorized news articles to produce the best results in terms of cluster purity[6]. As seen, [19] conducted feature reduction by the TF-IDF method, and then based on the mutual information method, adding relative word frequency factor and combining the weight of feature items, the mining process is adaptively improved, which makes the frequency information of feature items effectively used. Bation et al. [7] proposed an automatic document classifier of Tagalog news articles by stemming each document, representing it with TF-IDF values, and using it to train an SVM classifier.

Recent work [20] explored an approach to Russian text vectorization based on using State Rubricator of Scientific and Technical Information (SRSTI) categories as vector space dimensions. We can know from this study that in addition to the keywords selection process, vector calculation and comparison algorithms are employed. In [21][22] the Bayes formula was used to vectorize as opposed to classifying a document according to a probability distribution reflecting the probable categories that the document may belong to. The experiments showed that the proposed approach of the Naive Bayes vectorizer and SVM classifier has improved classification accuracy compared to the pure Naïve Bayes classification approach [23]. The transformation of data by the Bayesian vectorization technique, which reduces the dimensionality of data has contributed to a highly efficient great classification accuracy.

In our previous research[24], we introduced an improved dimension reduction technique that enhances accuracy by using Bayesian vectorization and a Laplace smoothing method. The study transformed each of the text documents in the dataset into the format of probability distribution in the vector space using the Bayesian vectorizer then feed this probability distribution to Support Vector Machine (SVM) for training and classification purposes. The results revealed that the improvement has significantly better classification accuracy compared to the TF-IDF vectorization technique.

3. METHODOLOGY

3.1. Dataset

A sample of 325 OPM song lyrics was collected from different websites such as Lyrics (www.lyrics.com), Genius (www.genius.com), and Musixmatch (www.musixmatch.com). The determination of category based on the following – Love songs, Christmas songs, Friendship songs, Worship songs, and Nationalism songs are manually annotated to determine the labels to be used for supervised learning.

For both the Bayesian vectorizer and the SVM classifier the analysis uses the same training dataset.The Bayesian vectorizer uses the vectorized training data supplied by the SVM classifier.

3.2. Preprocessing of Filipino Text Documents

The OPM song lyrics were taken from user-generated websites. Such submissions are not reviewed and can contain mistakes or errors.Therefore the lyrics such as word spelling were checked for correctness. It omits parts such as intro, refrain, chorus, and other parts.Instructions like chorus 4x are replaced with exactlyrics.We used Python tools and libraries to perform preprocessing of lyrics. These include the Natural Language Toolkit (NLTK), and Scikit-learn machine learning library. These provide a lot of options for conducting dataset experiments and performing analysis of text, audio or image, but here we are concerned with the song's text part.

3.3. Splitting Dataset

Data were divided into two in this process, those were data from training and testing. The Bayesian vectorizer provided training data that were to be used to build the classifier. Using Bayesian vectorization as well as TF-IDF vectorization, this splitting data used in this work was 70% training data and 30% testing data. This was done to determine the model's performance effect of the training data.

3.4. Vectorization

3.4.1. TF-IDF Vectorization

This research studied on two approaches to translating lyrics into a representation of the vectors. First is the use of TF-IDF. The TF-IDF score increases proportionally by the amount of a specific word that appears in a given document (term frequency) and is counteracted by the count (inverse document frequency) of the total number of documents in the corpus. The *tf-idf*matrix as shown in equation (1) converts all documents into rows, with all the terms represented as column vectors in the documents. The *tf* and *idf*product is used to calculate your *tf-idf* score [25][26].

$$tfidf\,(t,d,D) = tf(t,d) \times idf\,(t,D) \quad (1)$$

where ***t*** denotes the terms; ***d*** denotes each document; ***D*** denotes the collection of documents.

In equation (2), Term Frequency (**tf**) is calculated using:

$$tf(t,d) = \frac{(Number\ of\ times\ the\ term\ t\ appears\ in\ a\ document)}{(Total\ number\ of\ terms\ in\ the\ document,\ d)} \quad (2)$$

while Inverse Document Frequency (**idf**) is calculated using the equation (3)

$$idf(t,D) = \frac{log_e(Total\ number\ of\ documents,\ D}{Number\ of\ documents\ with\ term\ t\ in\ it).} \quad (3)$$

The calculated *tf-idf* values are then used as features for SVM to build classification models that identify the right category of each document.

3.4.2. Improved Bayesian Vectorization

The second approach is using the improved Bayesian vectorization technique as presented in our previous study[27]. The Bayesian vectorization technique is carried out to transform each of the text documents in the dataset into the format of probability distribution in the vector space, by using the Bayesian formula. This probability distribution is then fed to the SVMs classifier for training and classifying purposes. Laplace smoothing was applied to decrease the dimensionality and to improve the accuracy of the classification. Figure 1 shows the improved model applied for text vectorization.

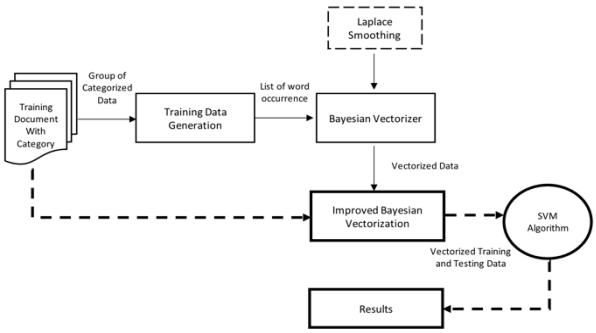

Figure 1: Improved Bayesian Vectorization Model

The prior probability of every category can then be computed using the equation (4).

$$Pr\,(C) = \frac{Total\ Number\ of\ Document\ in\ Category}{Total\_Number\_Of\_Document\_in\_Training\_Dataset} \quad (4)$$

To calculate the likelihood of a particular category for a particular word, the equation (5) below is used.

$$Pr\,(X|\,C) = \frac{Occurrence\ of\ Word\ In\ the\ Category}{Total\_Number\_of\_All\_Words\_in\_Category} \quad (5)$$

To avoid zero probability, smoothing technique is used using the equation (6).

$$Pr\,(X|\,C) = \frac{Occurrence\ of\ Word\ In\ the\ Category\ +1}{Total\_Number\_of\_All\_Words\_in\_Category + |V| + 1}$$

where |V| is the total unique words in the training set (6)

The overall probability for a document to be annotated to a particular category is calculated using the equation (7).

$$Pr\,(C|\,X) = Pr\,(W_1|\,C) * Pr\,(W_2|C) * Pr\,(W_3|C) * Pr\,(W_n|C) * Pr\,(C) \quad (7)$$

To avoid this underflow error, a mathematical *log* is applied using the equation (8).

$$Pr\,(C\,|\,X) = log\,(Pr\,(W_1|C) + log\,(Pr\,(W_2|C) + log\,(Pr\,(W_3|C) + log\,(Pr\,(W_n|C) + log\,(Pr\,(C)) \quad (8)$$

3.5. Classification

3.5.1. TF-IDF – SVM Classification

For TF-IDF, we use LIBSVM [28] machine learning toolkit to train the classification models, where cross-validation is used to tune the parameters. For each article in the training data set, we calculate the TF-IDF values of all terms selected by mutual information as a vector of real numbers, which is an input data instance for the SVM package. In the training phase, attribute scaling, kernel, and cross-validation are used to build the SVM model [4]. In the testing phase, we calculate the TF-IDF values of all selected terms in the testing article, use the same scaling factors for training to scale the TF-IDF values, and then classify the testing article using the SVM model learned in the training phase.

3.5.2. Improved Bayesian – SVM Classification

For Bayesian vectorization, the model was trained using a set of well-categorized vectorized training data supplied by the Bayesian vectorizer. The vectorized training data was split into a 70% training set and a 30% testing set or classification by performing the SVM. In the training phase, a 10 fold cross-validation was applied to limit the problems of

overfitting and underfitting. The rest of the classification tasks is performed using the linear kernel function with the implementation of parameter C that is set to 1.

4. SIMULATION RESULTS AND DISCUSSIONS

A comparison is performed to evaluate the SVM classification result using the TF-IDF vectorization technique and the improved Bayesian vectorization model to determine if the improved Bayesian vectorization model results in better classification accuracy as compared to the TF-IDF vectorization in SVM classifier.

Figure 2 illustrates the vectorized training data produced for classification by the Bayesian vectorizer that will be fed in SVM.

The predicted results of a classifier are presented in the confusion matrix that shows the number of correctly and incorrectly predicted and actual classifications.

	Dimensions				
Document Vectors	Love	Friendship	Nationalism	Christmas	Worship
D1	-2580.349993	-2516.334786	-2474.428634	-2262.621802	-2442.679011
D2	-1828.103199	-1949.473296	-1995.040641	-1973.467746	-2047.910439
D3	-1727.352492	-1692.9876	-1544.21037	-1672.625958	-1681.369276
D4	-957.3867199	-1003.832631	-1005.444345	-980.8188683	-944.3536479
D5	-446.3532096	-464.2874812	-422.9897976	-456.7388647	-439.7727406
D6	-361.3669245	-350.4719835	-340.8241934	-339.2494678	-316.397041
D7	-562.1941614	-512.9861315	-536.8724575	-539.4080285	-552.0452298
D8	-553.2400566	-637.1449015	-669.7632184	-637.2073184	-651.1753393
D9	-541.1405689	-497.2426322	-552.9954499	-532.3144759	-561.4067353
D10	-553.2400566	-637.1449015	-669.7632184	-637.2073184	-651.1753393
D11	-551.1535418	-547.1836843	-544.6356917	-551.2129395	-496.3401497
D12	-549.4614543	-587.3143457	-597.8084644	-529.0216214	-580.0977775
D13	-541.1405689	-497.2426322	-552.9954499	-532.3144759	-561.4067353
D14	-540.8106269	-532.1132048	-504.9112197	-538.0454528	-519.7914224

Figure 2: Sample Vectorized Training Data Generated by the Bayesian Vectorizer

As shown in Figure 3, the TF-IDF-SVM classifier made a total of 98 predictions. Out of 98, the classifier predicted 17 love songs, 5 Friendship songs, 11 Nationalism songs, 15 Christmas songs, and 26 Worship songs.

	Predicted Values				
Actual Values	Love	Friendship	Nationalism	Christmas	Worship
Love	17	0	0	0	2
Friendship	7	5	0	0	2
Nationalism	0	1	11	0	5
Christmas	3	0	0	15	1
Worship	3	0	0	0	26

Figure 3: Confusion Matrix for TF-IDF Vectorization and SVM Classification

The improved Bayesian-SVM classification correctly classified 17 Love songs, 11 Friendship songs, 24 Nationalism songs, 24 Christmas songs, and 21 Worship songs as shown in Figure 4.

	Predicted Values				
Actual Values	Love	Friendship	Nationalism	Christmas	Worship
Love	17	0	0	0	1
Friendship	0	11	0	0	0
Nationalism	0	0	24	0	0
Christmas	0	0	0	24	0
Worship	0	0	0	0	21

Figure 4: Confusion Matrix for Improved Bayesian Vectorization and SVM Classification

Method	Precision	Recall	F1-score	Accuracy
TF-IDF-SVM Classifier	0.57	0.89	0.69	0.76
Enhanced Bayesian-SVM Classifier	1.00	0.94	0.97	0.98

Figure 5: Comparison of the TF-IDF-SVM classifier and the Enhanced Bayesian-SVM classifier

The improved model achieved an average accuracy of 98% as shown in Figure 5, which is significantly higher than the classification method of TF-IDF-SVM which is 76%. The improved model also yields the highest values for Precision, Recall, and F1-score.

5. CONCLUSION

In this study, an improved Bayesian vectorization in transforming Filipino text for reducing the dimensionality of the used feature vectors for text document classification. It can be seen that, while TF-IDF-SVM yielded the lowest accuracy of 76%, the better Bayesian-SVM yielded 98%.

Although high-performance measures wereachieved in using the improved Bayesian vectorization model in building a machine learning classifier, itwould be better to use a larger dataset with a moreeven distribution for each class. Future Pilipino researches can also explore more on other possibilities of using Filipino documents for machine learning.

ACKNOWLEDGEMENT

This work was supported by the Commission of Higher Education K-12 Scholarship Program. We also thank the Technological Institute of the Philippines, Quezon City (TIP-QC) for providing us with helpful feedback and suggestions.

REFERENCES

[1] M. I. Abdulhussain and J. Q. Gan, “An experimental investigation on PCA based on cosine similarity and correlation for text feature dimensionality reduction,” *2015 7th Comput. Sci. Electron. Eng. Conf. CEEC 2015 - Conf. Proc.*, pp. 1–4, 2015, doi: 10.1109/CEEC.2015.7332689.

[2] M. K. Elhadad, K. Badran, and G. I. Salama, "A novel approach for ontology-based dimensionality reduction for web text document classification," *Proc. - 16th IEEE/ACIS Int. Conf. Comput. Inf. Sci. ICIS 2017*, pp. 373–378, 2017, doi: 10.1109/ICIS.2017.7960021.

[3] K. K. Bharti and P. K. Singh, "Hybrid dimension reduction by integrating feature selection with feature extraction method for text clustering," *Expert Syst. Appl.*, vol. 42, no. 6, pp. 3105–3114, 2015, doi: 10.1016/j.eswa.2014.11.038.

[4] A. I. Kadhim, Y. N. Cheah, and N. H. Ahamed, "Text Document Preprocessing and Dimension Reduction Techniques for Text Document Clustering," *Proc. - 2014 4th Int. Conf. Artif. Intell. with Appl. Eng. Technol. ICAIET 2014*, pp. 69–73, 2015, doi: 10.1109/ICAIET.2014.21.

[5] G. . Ramadevi and K. Usharani, "Study on Dimensionality Reduction Techniques and Applications," *Publ. Probl. Appl. Eng. Res.*, vol. 04, no. 1, pp. 134–140, 2013.

[6] A. K. Singh and M. Shashi, "Vectorization of Text Documents for Identifying Unifiable News Articles," *Int. J. Adv. Comput. Sci. Appl.*, vol. 10, no. 7, pp. 305–310, 2019, doi: 10.14569/ijacsa.2019.0100742.

[7] A. D. C. Bation, E. Q. Manguilimotan, and A. J. O. Vicente, "Automatic categorization of Tagalog documents using support vector machines," *PACLIC 2017 - Proc. 31st Pacific Asia Conf. Lang. Inf. Comput.*, no. Paclic 31, pp. 346–353, 2019.

[8] H. T. Sueno, B. D. Gerardo, and R. P. Medina, "Transforming Text Documents Into Numerical Format Using Enhanced Bayesian Vectorization For Multi-Domain Document Classification," in *International Conference on Science, Technology, and Management (ICSTM)*, 2019, doi: IRES.27112019.11807.

[9] G. Orellana, B. Arias, M. Orellana, V. Saquicela, F. Baculima, and N. Piedra, "A study on the impact of pre-processing techniques in Spanish and english text classification over short and large text documents," *Proc. - 3rd Int. Conf. Inf. Syst. Comput. Sci. INCISCOS 2018*, vol. 2018-Decem, pp. 277–283, 2018, doi: 10.1109/INCISCOS.2018.00047.

[10] A. K. Uysal and S. Gunal, "The impact of preprocessing on text classification," *Inf. Process. Manag.*, vol. 50, no. 1, pp. 104–112, 2014, doi: 10.1016/j.ipm.2013.08.006.

[11] R. Sergienko, M. Shan, and A. Schmitt, "A Comparative Study of Text Preprocessing Techniques for Natural Language Call Routing," vol. 427, pp. 23–37, 2017, doi: 10.1007/978-981-10-2585-3.

[12] V. R. Sayoc, T. K. Dolores, M. C. Lim, L. Sophia, and S. Miguel, "Nature Inspired Dimensional Reduction Technique for Fast and Invariant Visual Feature Extraction," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 8, no. 3, pp. 195–200, 2019, doi: https://doi.org/10.30534/ijatcse/2019/57832019.

[13] T. Sabbah, M. Ayyash, and M. Ashraf, "Hybrid Support Vector Machine based Feature Selection Method for Text Classification," vol. 15, no. 3, pp. 599–609, 2018.

[14] R. A. Aziz *et al.*, "Two Stages Song Subject Classification on Indonesian Song Based on Lyrics, Genre & Artist," *2016 Int. Conf. Inf. Technol. InCITe 2016 - Next Gener. IT Summit Theme - Internet Things Connect your Worlds*, no. 05, pp. 21–24, 2018, doi: 10.1109/SIET.2018.8693201.

[15] L. J. P. Van Der Maaten, E. O. Postma, and H. J. Van Den Herik, "Dimensionality Reduction: A Comparative Review," *J. Mach. Learn. Res.*, vol. 10, pp. 1–41, 2009, doi: 10.1080/13506280444000102.

[16] S. Kaur, "A Survey on Dimension Reduction Techniques for Classification of Multidimensional Data," vol. 2, no. 12, pp. 31–37, 2016.

[17] H. Xie, J. Li, Q. Zhang, and Y. Wang, "Comparison among dimensionality reduction techniques based on Random Projection for cancer classification," *Comput. Biol. Chem.*, vol. 65, pp. 165–172, 2016, doi: 10.1016/j.compbiolchem.2016.09.010.

[18] B. Tang, M. I. Heywood, and M. Shepherd, "Comparing and Combining Dimension Reduction Techniques for Efficient Text Clustering," *Siam*, pp. 1–10, 2005.

[19] H. Wang, Y. Yuan, J. Kou, X. Zhang, C. Wang, and J. Duan, "Research on Feature Mining Algorithm Based on Product Reviews," *Proc. 2019 IEEE Int. Conf. Artif. Intell. Comput. Appl. ICAICA 2019*, pp. 205–210, 2019, doi: 10.1109/ICAICA.2019.8873491.

[20] Y. Solomonova and M. Khlopotov, "Russian Text Vectorization: An Approach Based on SRSTI Classifier," in *International Conference on Digital Transformation and Global Society*, 2020, pp. 754–764, doi: https://doi.org/10.1007/978-3-030-37858-5_64.

[21] D. Isa, L. Lam Hong, V. P. Kallimani, and R. Rajkumar, "Text Document Pre-Processing Using the Bayes Formula for Classification Based on the Vector Space Model," *Comput. Inf. Sci.*, vol. 1, no. 4, pp. 79–90, 2008, doi: 10.5539/cis.v1n4p79.

[22] L. H. Lee, R. Rajkumar, and D. Isa, "Automatic folder allocation system using Bayesian-support vector machines hybrid classification approach," *Appl. Intell.*, vol. 36, no. 2, pp. 295–307, 2012, doi: 10.1007/s10489-010-0261-0.

[23] L. H. Lee, C. H. Wan, R. Rajkumar, and D. Isa, "An enhanced Support Vector Machine classification framework by using Euclidean distance function for text document categorization," *Appl. Intell.*, vol. 37, no. 1, pp. 80–99, 2012, doi: 10.1007/s10489-011-0314-z.

[24] H. T. Sueno, B. D. Gerardo, and R. P. Medina, "Multi-class document classification using support

vector machine (SVM) based on improved naïve bayes vectorization technique," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 3, pp. 3937–3944, 2020, doi: 10.30534/ijatcse/2020/216932020.

[25] R. N. Waykole and A. D. Thakare, "A Review of Feature Extraction Methods for Text Classification," *Int. J. Adv. Eng. Res.*, no. Vdv, pp. 351–354, 2018.

[26] Z. Qu, X. Song, S. Zheng, X. Wang, X. Song, and Z. Li, "Improved Bayes Method Based on TF-IDF Feature and Grade Factor Feature for Chinese Information Classification," *Proc. - 2018 IEEE Int. Conf. Big Data Smart Comput. BigComp 2018*, pp. 677–680, 2018, doi: 10.1109/BigComp.2018.00124.

[27] H. T. Sueno, B. D. Gerardo, and R. P. Medina, "Converting text to numerical representationusingmodified bayesian vectorization technique for multi-class classification," *Int. J. Adv. Trends Comput. Sci. Eng.*, vol. 9, no. 4, pp. 5618–5623, 2020, doi: 10.30534/ijatcse/2020/211942020.

[28] C. C. Chang and C. J. Lin, "LIBSVM: A Library for support vector machines," *ACM Trans. Intell. Syst. Technol.*, vol. 2, no. 3, pp. 1–39, 2011, doi: 10.1145/1961189.1961199.