Journal of Science and Technology
ISSN: 2362-1443 (Print) 3059-9490 (Online)
2024, Vol. 4, No. 1, 35-40

Classification of Grievances in Hello Sarkar using Supervised Machine Learning

Binaya Subedi
Nepal College of Information Technology
Pokhara University
binaya.202904@ncit.edu.np

Basanta Joshi
Institute of Engineering,
Tribhuvan University
basanta@ioe.edu.np

Article History:
Received: 15 March 2024
Revised: 8 May 2024
Accepted: 17 June 2024

Keywords— Support Vector Machine, Naïve Bayes, NLTK, Grievances

***Abstract*— Classification of Grievances automatically as per some pre-defined category is an automated task that can be carried out with the aid of some confidence gained after learning the content from a training set of grievance dataset. This research mainly aims to evaluate widely used supervised machine learning approaches like Naïve Bayes and SVM. For this experiment, the dataset of 4 categories with about 3,000 datasets collected from the official site of 'Hello Sarkar' portal which has been used by the Government of Nepal, Office of the Prime Minister and Council of Ministers for handling the public grievances was used. As per the recent research outcomes, none of the researchers had spoken about which machine learning algorithm outperforms the rest of the other and has always suggested that the outcome solely relies on the type of dataset used. So, in this research, above -mentioned machine learning was used and the accuracy of those applied algorithms was compared. As public grievances are either in Nepali, English or in Roman form, the research procedure mainly focused on how well the above-mentioned algorithms perform on those grievances in Nepali language dataset. Moreover, as per the suggestions from various researches conducted previously, the enhancement in accuracy was targeted out by cleaning the dataset by adding the stop words in NLTK corpus and removing extra useless symbols during preprocessing state for refining dataset before the implementation of machine training algorithms.**

I. Introduction

The Office of the Prime Minister and Council of Ministers is currently running a section named 'Hello Sarkar'. This section is dedicated to addressing and resolving public grievances, which is a crucial aspect of effective governance. 'Hello Sarkar' [23] section of Government of Nepal, Office of the Prime Minister and Council of Ministers is using grievance management system called 'Gunaso Portal' where public can file complaint or grievances regarding the problems faced by them due to government officials and offices such as delayed in the work, misbehave, problem faced by them while getting services etc. Manual classification of grievance is a lengthy process and hence costs a lot of time for the analysis of the category of posted grievances and to which agency shall it be forwarded to. Mission critical grievances which seek immediate response might get delayed to get processed due to the bottleneck in the manual classification and analysis of grievances. Hence automatic classification of grievances with the aid of the trained dataset will help to overcome the above-mentioned problem as it has been done in the automatic news classification [1].

Automatic Grievances Classification has two major issues.

- The grievances registered by the public are not in appropriate format as that of news in [1].
- Grievances are not in appropriate format as per the length as in [1]. Some grievances are too long and some are too short. So, classification of such diversified grievances is also a challenging task. How can automatic classification of grievances be performed effectively while addressing the challenges of diverse language formats and varied length of grievances, with a focus on Nepali-language grievances?

The major objectives of this research are – (1) Preparation of the dataset for this research purpose that contains Nepali Grievances and cleaning the dataset in the preprocessing stage of the research so as to enhance the accuracy of the classification; and (2) Classification of grievances in Hello Sarkar using Support Vector Machines and Naïve Bayes along with the comparison of the accuracy obtained.

II. Related Work

Machine learning and artificial intelligence doesn't have a very long history. In 1943 Water and MCCulloch created a neural network based mathematical algorithm. Research regarding Neural network really was stagnated right after the machine learning approach of research was proposed by Papert and Minsky in 1969. During the course of their study, they found the most fundamental but important aspect of perceptron about its incapability of processing exclusive-or

circuits. The next important aspect about the finding of that research is that the computers at that time were not equipped with enough processor capabilities so as to handle task required by such neural network. In 1975, the pioneer multilayered neural network was developed which is also regarded as the first unsupervised neural network. The term Deep learning was coined only in 1986. First application of Support Vector Machine was in 1998 by Joachims in 1998. SVM was regarded as the classifier that uses the hyper-plane concept to maximize distance between supporting hyper-planes.

In 2018 Ashok Kumar Pant and Tej Bahadur Shahi from Central Department for Computer Science and IT, in Tribhuvan University, located in Kathmandu carried out research in news classification of Nepali language using Support Vector Machine, Naïve Bayes and Neural Networks. They used TF-IDF as feature selection method which assumes independence in between the words in text corpus due to which the accuracy was comparatively lower.

Moreover, in 2020 Oyesh Mann Singh, Sandesh Timilsina, Bal Krishna Bal and Anupam Joshi had done research in Abusive Sentiment Detection as Aspect Based in Social Media Texts of Nepali Language.

In 2021, a research paper was published in Journal of Artificial Intelligence with the title "Sentiment Analysis of Nepali COVID19 Tweets Using NB, SVM and LSTM". Which was carried out by Milan Tripathi. Due to the complex structure of Nepali Language and aspect-based classification approach, the result was not that satisfactory. LSTM outperforms other algorithms on entire sentences whereas NB performs best on aspect-based classification approaches.

Not much work has been done in Nepali text classification and sentiment analysis because of the insufficient dataset. In 2021, T. B. Shahi, A. Basnet ,C. Sitaula and A. Mainali, had published their research in the Journal of Computational Intelligence and Neuroscience, 2021. They used "Deep Learning-Based Methods for sentiment Analysis on Nepali COVID-19-Related Tweets." They used eight traditional machine learning algorithms in their work like Naïve Bayes ,SVM, Logistic regression, XGBoost, and K-nearest neighbors (K-NN) and concluded that research has a limitation as their method totally ignored most crucial aspect of sequential approach for tokens, their methods exploit fastText as (" ft") and embeddings based on probability ("ds" and "da") for text classification. They also suggested that the performance could have increased if they had used other embeddings like Word2Vec and GloVe.

In 2016, Dadgar, Araghi, M.S. and Farahani [24] had conducted research regarding the news classification using TF-IDF vectorization and Support Vector Machine. From the research, it is clearly seen that the term frequency and inverse document frequency i.e. TF-IDF is that approach of vectorization that is used to vectorized the text obtained after the preprocessing and hence count the term frequency of the text being used before commencing the pre- processed text to the machine learning algorithm.

In [1,2,3,24], standard matrices like precision, recall, F1 score has been used for the evaluation of the performance of the algorithm. These standard matrices are the key indicators to show how well the algorithm performs regarding the supplied dataset.

From the above discussion, it can be assumed that performance of the algorithm solely relies on the type of dataset being used and none of the researchers had claimed that the specific machine learning algorithm is better than the other. The output and results are surprisingly different with the variation of dataset as well as the attributes of the text being used in the dataset to classify the text.

III. Methodology

In our research, supervised based machine learning approaches; Naïve Bayes as well as Support Vector Machine was used to classify the grievances of 'Hello Sarkar'. Following procedures was followed sequentially to carry out the research.

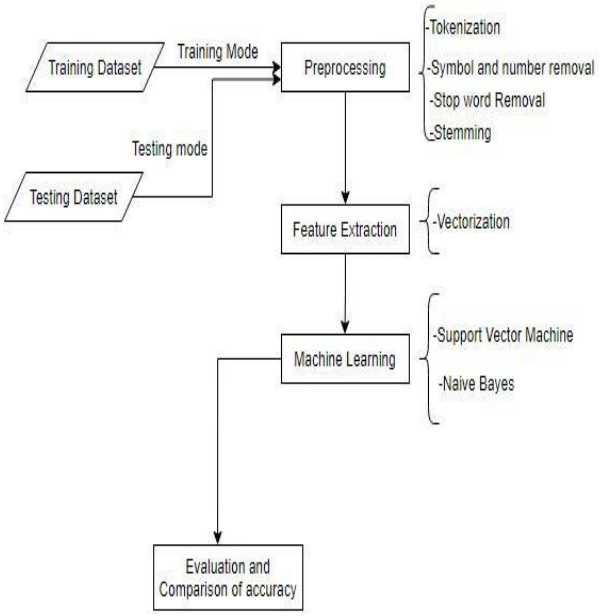

Fig. 1. Grievances Classification System Pipeline.

A. *Dataset Preparation*

As we all know about the fact that Nepali language solely belongs to that common script, Devanagarik, which was invented by ancestors in around 11th century. Literally, it consists of 12 number of vowels "अ, आ, इ, ई, उ, ऊ, ए, ऐ, ओ, औ, अं, अः", and 36 consonant "क, ख, ग, घ, ङ, च, छ, ज, झ, ण, त, थ, द, ध, न, प, फ, ब, भ, म, य, र, ल, व, श, ष, स, ह, क्ष, त्र, ज्ञ" and 10 numbers "१,२,३,४,५,६,७,८,९". The registered grievances and manual level of the data till date has been used in our research purpose. The dataset contains Nepali text format which is then cleaned in the preprocessing stage of the research and hence it has been tried to enhance the accuracy of the algorithm during the classification of grievances. Eleven classes of labeled dataset with almost ten thousand three hundred datasets has been developed which will be beneficial for the community so as to carry out the similar projects in near future.

B. *Preprocessing*

After successful preparation of the dataset, preprocessing of text data has been carried out. It includes the reduction of noise in the data and is helpful in further training and evaluation procedure. In this stage, special and useless characters have been removed that doesn't have a significant

role in text classification. After that, stop words have been removed. In this step, the stop word criteria has been modified that is usually used from NLTK corpus and ultimately it has been tried to enhance the accuracy in the text classification as in [1]. Finally, word streaming step has been carried out. Finally, a new Stopword library was developed and used in the experiment so as to enhance the accuracy of the classification.

C. Tokenization of the text

In this step, the whole text has been broken down into the sentences. The sentences are ended with full stop, question marks, exclamation, vertical bars etc. which has been analyzed properly. Finally, the space in the sentences has been used to break down the sentences into words.

1) Removal of special symbols and characters.

In this step, the special symbols and characters was removed. Symbols like,), (, *, &, ^, %, !, @, #, $, <, >, ;, :, , " and numbers like 1,2,3,4,5,6,7,8,9, १,२,३,४,५,६,७,८,९,० that has low significance in classification has been removed.

2) Removal of stop word

Those words with higher frequency in the text and has not much to influence has been removed in this step to classify the text with the aim to achieve higher accuracy. Here the state-of-the- art method in [1] used the 255 stop-words like "भयो, हुन्छ, थियो, हो, म, आँफु, हुनेछ, हामी, त्यो, त्यसैले" etc. Here, the stop-word criteria has been modified as per the obtained dataset and such words with little significance and have low influencing nature in the text has been removed to obtain accuracy more than in [1].

3) Word Stemming

As we all know the fact during the stemming stage, it helps to reduce the words into their respective stem. Stem of the words reflects the actual meaning of the text so segmentation of the word into its stem has been done in order to reduce the dimension of vocabulary in some manner. [19] Example:

"एउटा स्वास्थ्य सेवाका लागि स्वास्थ्य चौकीको एकदमै आपातकालीन भएकोले उपलब्ध गराउनु हुन विनम्र अनुरोध गर्दछौ। "

['एउटा' 'स्वास्थ्य' 'सेवा' 'लागि' 'स्वास्थ्य' 'चौकी' 'एकदम' 'आपतकाल' 'हु' 'उपलब्ध' 'गराउ' 'नम्र' 'अनुरोध' 'गर']

D. Feature Extraction

In this stage, various feature vector construction has been carried out. Relevant features to classify the text has been sorted and analyzed properly and finally TF-IDF vectorization has been used to vectorize the Nepali text. This has been calculated as,

$$W_{t,d,D} = tf_{t,d} * idf_{t,D} \quad (1)$$

Where,

$tf_{t,d} = f_{t,d} / \max [f_{i,d} : i \in d]$

$idf_{t,D} = \log [N/1 + |\{d \in D: t \in d\}|]$

tf is the Term Frequency,

idf is the Inverse document frequency,

$f_{t,d}$ is number of the terms that the document d contains,

$\max [f_{i,d} : i \in d]$ is the maximally occurring terms t which document d contains,

N is regarded as the number of the document that is contained in corpus D,

And finally, $|\{d \in D: t \in d\}|$ gives the total number of documents where we can observe term t .

There is the use of Supervised based Machine learning Algorithms; "Naïve Bayes and Support Vector Machine" for text classification.

After the successful features extraction within the dataset from the text, then supervised based machine learning technique; "Naïve Bayes and Support Vector Machine" has been used as in [1].

1) Naïve Bayes Classifier

It is also considered as; "Bayes' Rule or Bayes' law" that depends on the conditional probability. It is given as:

$$P(A|B) = [P(B|A) \ P(A)] / \ P(B)] \quad (2)$$

Where,

P (A|B) is regarded as the conditional posterior probability (i.e., hypothesis A's probability on the observed i.e. event B.) and P (B|A) is considered as the Probability of Likelihood. (i.e., evidence's probability stated that the probability of that hypothesis).

Following are various Naive Bayes' based algorithms which were used for the classification purpose.

a) Gaussian Naive Bayes

It is a probabilistic algorithm that makes the assumption that the data follows a normal distribution. [21]

b) Multinomial Naive Bayes

It is considered as on of the variation of the Bayesian algorithm which is used for discrete data..[21]

c) Bernoulli Naive Bayes

It is a variation of the Multinomial Naive Bayes algorithm. It is used for binary/boolean features.[21]

d) Complement Naive Bayes

It is an extension of the standard Naive Bayes algorithm.[21]

2) Support Vector Machine Classifier

It is regarded as the hyperplane-based classifier which is discriminative in nature that finds the hyperplane to distinguish the multidimensional data into separate classes. So, these can be considered as the binary classifier. Linear kernels and Radial Basis Function (RBF) was used for the purpose of decision making in SVM.

a) Linear Kernel

The kernel which is linear in fashion, also known as the linear SVM, is regarded as the function which is ultimately used by the support vector machine for those kind of data that are linearly separable. The function of the kernel in this regard is given as K(x,y) = x .y , here x as well as y are the input vectors.[22]

The linear kernel function maps the input data into the original input space, where it can be separated by a linear hyperplane.

b) Radial Basis Function (RBF) Kernel

The 'Radial Basis Function (RBF) 'kernel is also known as the Gaussian based kernel which is regarded as the kernel function used for non-linearly separable data. Ultimately it maps those input data in the higher dimensional space in order to make it linearly separable.

The Radial Basis Function kernel is known by the equation K (x,y) = exp(-|| x – y||^2 / 22 σ^2). Here, x and y are regarded as the input vectors and σ is considered as the width that the radial function has.

c) Polynomial Kernel

The polynomial kernel, also known as the polynomial SVM, is one of the kernel functions that is used in SVM for classifying the data that are non-linear in fashion. It transforms those of data of input space into higher dimensional space in order to make it the linearly separable form.

The kernel function can be defined by the function K (x,y) = (x . y + c)^ d, where parameters x as well as y are input vectors, c is basically the constant, and d is the degree of the polynomial. The function for the kernel ultimately maps the input into the higher- dimensional space, where ultimately the data can be generated by the polynomial hyperplane.

It is important to carefully tune these parameters to achieve optimal performance on the dataset.

IV. CONCEPTUAL MODEL AND MODEL VERIFICATION

Every steps listed in the Fig 3 conceptual model diagram was used to pre-process, vectorization and Classification of the grievances into their respective classes. In every experiment, 70 % of the data was used as training dataset and remaining dataset was used as testing dataset.

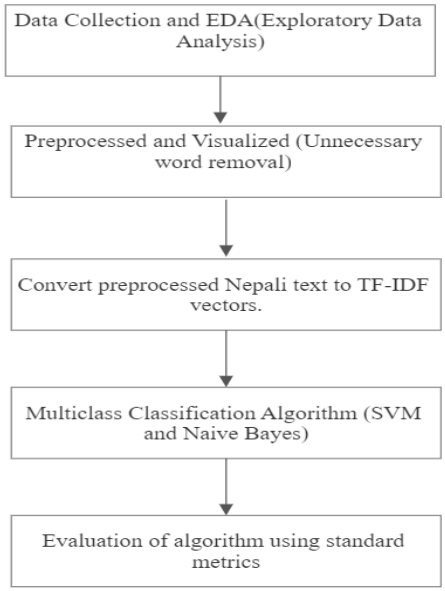

Fig. 2. Conceptual model of grievance classification.

TABLE I. TOTAL DATASET PREPARED FROM 'HELLO SARKAR' PORTAL (TOTAL COUNT:10335)

Grievance Categories (गुनासो वर्ग)	Total Grievances (गुनासोहरु)
लागु पदार्थ सम्बन्धी	159
प्राकृतिक श्रोत साधन सम्बन्धी	233
भ्रष्टाचार सम्बन्धी	73
कर्मचारी सम्वन्धी	1379
अर्थ सम्बन्धी	856
सोधपुछ, सुझाव, प्रशंसा सम्बन्धी	1899
सूचना तथा संचार सम्बन्धी	642
स्वास्थ्यसँग सम्बन्धी	899
वेबसाइट सम्बन्धी	2863
खानेपानी सम्बन्धी	770
शान्ति सुरक्षा सम्बन्धी	562

TABLE II. TABLE 1 OUTPUT OF FIRST EXPERIMENT WITH THE USE OF 11 CLASSES AND 10335 DATASET LISTED ABOVE.

ALGORITHMS	ACCURACY	F1 SCORE
SVM - LINEAR KERNEL	0.63	0.62
SVM - RBF KERNEL	0.60	0.59
SVM - POLY KERNEL	0.44	0.39
NAIVE BAYES-GAUSSIAN	0.35	0.37
NAIVE BAYES-MULTINOMIAL	0.44	0.36
NAIVE BAYES-BERNOULLI	0.38	0.29
NAIVE BAYES-COMPLEMENT	0.55	0.54

Finally, in second experiment, total of four classes and 2898 dataset was used to keenly analyze the performance of the algorithm and fine tuning with the application of Nepali Stop words developed in our research which can be listed as:

'यो','म','यी','त्यो','ती','प्रति','हरेक','प्रतेक','मान नीय','अलिकति','थोरै','सबै','केहि','एक','दुई','तिन','प्रथम','रही','आज','द्रितीय','तृतीय','को','सब','कुन','जी','श्री','वैसा','कति','जो', 'जसरी', 'जुन', 'र', 'तथा', 'तर', 'आदि', 'किन्तु', 'इत्यादि', 'परन्तु', 'ले', 'बाट', 'लाई', 'ले', 'द्वारा', 'वारि', 'पारि', 'अघाडी', 'कुमारी', 'चित्रकार', 'राम', 'उन', 'उनका', 'उनकी', 'पर्दिप', 'उनको', 'उस', 'एक', 'एवं', 'एस', 'कई', 'कहा', 'का', 'काफ़ी', 'कि','की', 'कुल', 'के', 'को', 'कोई', 'घर', 'जब', 'जहाँ','जा','जो','तब','तिस','थिए','हरि','थे','होला','द्वारा','न','ने','पर','पूरा','पे','यदि','यहाँ','यही','या','वर्ग','संग','हो','जेसे','दवारा','यहां','जे','यहि','मोबि','अदि','छ','हेरीच्छ','जहां''अघ', 'अद', 'अल', 'आद', 'इत', 'उनक', 'एव', 'कत', 'कह', 'जसर', 'जह',

'तथ', 'दव', 'नन', 'परन', 'यद', 'यह', 'रक', 'रत', 'रथम', 'रह', 'वर', 'हर' 'भनि', 'गरी'

TABLE III. DATASET FOR SECOND AND THIRD EXPERIMENT

गुनासो वर्ग	सङ्ख्या
अर्थ सम्बन्धी	251
शान्ति सुरक्षा सम्बन्धी	161
सूचना तथा सञ्चार सम्बन्धी	191
स्वास्थ्य सम्बन्धी	267

Finally, the Output of the second experiment is as follows:

TABLE IV. OUTPUT OF SECOND EXPERIMENT

ALGORITHMS	ACCURACY	F1 SCORE
SVM - LINEAR KERNEL	0.61	0.61
SVM - RBF KERNEL	0.59	0.58
SVM - POLY KERNEL	0.52	0.49
NAÏVE BAYES- GAUSSIAN	0.40	0.39
NAIVE BAYES- MULTINOMIAL	0.56	0.52
NAIVE BAYES- BERNOULLI	0.56	0.56
NAIVE BAYES- COMPLEMENT	0.61	0.61

After successful cleaning of both Nepali and English text using NLTK corpus and rigorously tuning the performance parameters the accuracy of third experiment increases which is as follows:

Table 2 Output of third experiment

ALGORITHMS	ACCURACY	F1 SCORE
SVM - LINEAR KERNEL	0.77	0.77
SVM - RBF KERNEL	0.76	0.76
SVM - POLY KERNEL	0.60	0.59
NAIVE BAYES-GAUSSIAN	0.55	0.55
NAIVE BAYES-MULTINOMIAL	0.70	0.70
NAÏVE BAYES-BERNOULLI	0.68	0.67
NAIVE BAYES-COMPLEMENT	0.76	0.76

The graphical representation of the final output is as follows:

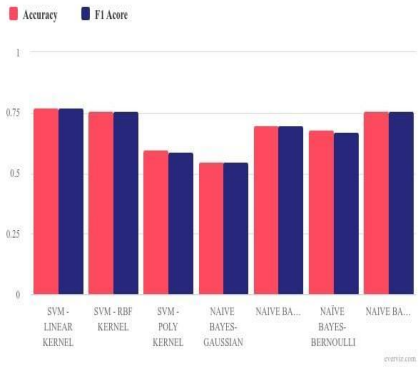

Fig. 3. Bar Graph of the accuracy and F1 Score of third experiment

V. CONCLUSION AND RECOMMENDATION

This study aimed to build a machine learning model for Nepali grievance classification abstracted from Nepal's government portal (Office of prime minister or council of minister) using a multi-class classification algorithm. The dataset that was considered during this study contained a varied number of documents written in Nepali, which were divided into several classes. The classification algorithms tested were support vector machines (SVMs) and Naive Bayes.

The results obtained from the study showed that SVM algorithm performed the best among the tested algorithms, with the highest accuracy This demonstrates the effectiveness of the SVM algorithm in handling multi-class classification tasks, particularly in this context of text classification in Nepali Language.

Overall, this research of grievance classification highlights potentiality of machine learning techniques in addressing real-world problems, such as text classification, and the importance of carefully selecting the appropriate algorithms and tuning their parameters to achieve the best results.

The accuracy can be enhanced if the language format used to post the grievance was within the appropriate Nepali Language. For example, the dataset contains words like बिध्यालय, विद्यालय, विध्धयालय, बिदालय etc. for which the correct word is विद्यालय. There are a lot of words that can be found in the text which are grammatically incorrect. Hence, somehow if we are able to conduct research to correct the grammar of the dataset prepared, there is no doubt that the accuracy ultimately increases. Further research could be carried out to classify the grievances using some neural network-based classification approach using the prepared dataset. Furthermore, some other classification algorithm can be explored so as to compare the accuracy obtained in this research.

REFERENCE

[1] Shahi, T.B. and Pant, A.K., 2018, February. Nepali news classification using naïve bayes, support vector machines and neural networks. In 2018 International Conference on Communication Information and Computing Technology (ICCICT) (pp. 1-5). IEEE.

[2] Tripathi, M., 2021. Sentiment analysis of nepali covid19 tweets using nb svm and lstm. Journal of Artificial Intelligence, 3(03), pp.151-168.

[3] Sitaula, C., Basnet, A., Mainali, A. and Shahi, T.B., 2021. Deep learning-based methods for sentiment analysis on Nepali covid-19-related tweets. Computational Intelligence and Neuroscience, 2021.

[4] W. W. Cohen et al., "Learning rules that classify e-mail," i n AAAI spring symposium on machine learning in information access, vol. 18. California, 1996, p. 25.

[5] I. Stuart, S.-H. Cha, and C. Tappert, "A neural network classifier for junk e-mail," Document Analysis Systems VI, pp. 442–450, 2004.

[6] X. Carreras and L. Marquez, "Boosting trees for anti-spam e mail filtering," arXiv preprint cs/0109015, 2001.

[7] A. K. Tegegnie, A. N. Tarekegn, and T. A. Alemu, "A compara tive study of flat and hierarchical classification for amharic news text using svm," Culture, vol. 2007, p. 1, 2010.

[8] S. Lai, L. Xu, K. Liu, and J. Zhao, "Recurrent convolutional neural networks for text classification." In AAAI, vol. 333, 2015, pp.2267-2273.

[9] A. Joulin, E. Grave, P. Bojanowski, and T. Mikolov, "Bag of tricks for efficient text classification," arXiv preprint arXiv:1607.01759, 2016.

[10] K. Kowsari, D. E. Brown, M. Heidarysafa, K. J. Meimandi, M. S. Gerber, and L. E. Barnes, "Hdltex: Hierarchical deep learnin g for text classification," arXiv preprint arXiv:1709.08267, 2017.

[11] T. B. Shahi, T. N. Dhamala, and B. Balami, "Support vector machines based part of speech tagging for nepali text," International Jo urnal of Computer Applications, vol. 70, no. 24, 2013.

[12] A. Paul, B. S. Purkayastha, and S. Sarkar, "Hidden markov model based part of speech tagging for nepali language," in Advanced Computing and Communication (ISACC), 2015 International Symposium on. IEEE, 2015, pp. 149–156

[13] B. K. Bal and P. Shrestha, "A morphological analyzer and a st emmer for Nepali," PAN Localization, Working Papers, vol. 2007, pp. 324–31, 2004.

[14] S. B. Bam and T. B. Shahi, "Named entity recognition for nepali text using support vector machines," Intelligent Information Management, vol. 6, no. 02, p. 21, 2014.

[15] A. Dey, A. Paul, and B. S. Purkayastha, "Named entity recognition for nepali language: A semi hybrid approach," International Journal of Engineering and Innovative Technology (IJEIT) Volume, vol. 3, pp. 21–25, 2014.

[16] Sun, C., Qiu, X., Xu, Y. and Huang, X., 2019, October. How to fine-tune bert for text classification?. In China national conference on Chinese computational linguistics (pp. 194-206). Springer, Cham.

[17] González-Carvajal, S. and Garrido-Merchán, E.C., 2020. Comparing BERT against traditional machine learning text classification. arXiv preprint arXiv:2005.13012.

[18] Rao, A. and Spasojevic, N., 2016. Actionable and political text classification using word embeddings and LSTM. arXiv preprint arXiv:1607.02501.

[19] B. K. Bal and P. Shrestha, "A morphological analyzer and a st emmer for Nepali," PAN Localization, Working Papers, vol. 2007, pp. 324–31, 2004.

[20] Shahi, T.B. and Pant, A.K., 2018, February. Nepali news classification using naïve bayes, support vector machines and neural networks. In 2018 International Conference on Communication Information and Computing Technology (ICCICT) (pp. 1-5). IEEE.

[21] Sunil Ray, "Naive Bayes Classifier Explained With Practical Problems" downloaded from https://www.analyticsvidhya.com/ blog/2017/09 /naive-bayes-explained (Accessed 2017)

[22] Patle, A. and Chouhan, D.S., 2013, January. SVM kernel functions for classification. In 2013 International Conference on Advances in Technology and Engineering (ICATE) (pp. 1-9). IEEE.

[23] https://gunaso.opmcm.gov.np/home.

[24] Dadgar, S.M.H., Araghi, M.S. and Farahani, M.M., 2016, March. A novel text mining approach based on TF-IDF and Support Vector Machine for news classification. In 2016 IEEE International Conference on Engineering and Technology (ICETECH) (pp. 112-116