JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

# Enriching Personality Classification through Analysis of CV using NLP and Deep Learning

**Sharmila Shankar Patil[1], Mr.S.S.Redekar[2]**
[1]M.Tech Student, Department of Computer Science & Engineering, Ashokrao Mane Group of Institution,Vathar,Maharastra,India.
[2]Asst.Professor, Department of Computer Science & Engineering, Ashokrao Mane Group of Institution,Vathar, Maharastra,India.

*Abstract* – **Increased urbanization has resulted in a significant rise in work opportunities. As a result, a broad organization has evolved, necessitating the employment of a diversified workforce. Organizations advertise job openings in order to recruit talented and helpful employees. Most of the time, the firm receives an excessive number of job applications due to high demand and an insufficient quantity of available positions, as well as rising unemployment. Companies have a tremendous incentive to hire the person with the most personality. Choosing from the massive quantity of resumes or curriculum vitae received is a challenging task. As a result, the development of an efficient and reliable automated personality rating system is necessary. For the objective of a CV content extraction system, a few ways have been investigated. A number approaches have been examined in order to achieve the goal of a CV content extraction system. The bulk of these approaches have faults that must be addressed. As a result, the relevant research on the CV data extraction methodology is performed to achieve the presented approach that utilizes Bag of Words, TF-IDF and Deep Belief Networks along with Decision Making and Protocol Mapping. The approach has been elaborated to achieve the improved analysis of the CVs which has been quantified using effective experimentation that has resulted in highly satisfactory results.**

***Keywords***: *Curriculum Vitae, Natural Language Processing, Bag of Words, Term Frequency, Inverse Document Frequency, Deep Belief Network, DecisionMaking and Protocol Mapping.*

## I. INTRODUCTION

As a result of rising connectivity and large-scale developments in the healthcare industry, consumers' livelihoods and increasing life expectancy have grown, causing an overall population growth. Because these people need to work to make ends meet, the large increase in population translates to a significant increase in employment. The growing population is one of the reasons why most big firms receive an increasing number of applications for a given job post. Several of these large corporations obtained applications through a number of means, notably email, postal, and broadband services, which comprise the bulk of programs accessible today.

The internet recruitment infrastructure has evolved rapidly over the last several decades, resulting in a large flood of applications for a given job position inside the company. With such a high proportion, it is challenging to identify the right individual for the job since manually sifting and classifying these data gets difficult after a time. As a result, in order to achieve the most productive and cost-effective use of recruitment time, an innovative approach for automatic application separation is necessary. A résumé or curriculum vitae (CV) summarizes a candidate's experience and qualifications for a job opportunity [1]. Resumes are available in a number of formats, obviously it depends upon the way the user wishes to present the content.

Thin internet technology has evolved into a massive repository of knowledge accessible with a mere fingertip touch. The internet is not just a repository of information; it also provides for a wide variety of benefits that have recently been utilized to make life simpler for the general population. Not only does the internet platform assist the general people, but it also benefits a huge spectrum of corporations and businesses that utilize it to increase employee cooperation and accessibility. The bulk of the services offered by this platform are extensively employed through both big and small companies, providing a significant boost and incentive for worldwide technological advancements.

The online community has been critical to numerous of the world's achievements by providing a framework for the development of new ideas and approaches. The network was developed to make it easier for researchers to communicate and exchange materials. The primary goal of the internet platform was to facilitate military collaboration as well as to unite the activity of numerous scholars over long distances. It was mostly useful and changed the course of humanity's history after it was created and promptly publicized for broad usage. Since that time, the online platform has demonstrated to be excellent in disseminating information and expertise to everybody who has internet access.

To highlight specific areas of the resume, different fonts and designs, in addition to color combination, are used. This type of resume diversity typically increases the time it requires employees to determine whether or not it meets the employment criteria. The application of NLP (Natural Language Processing) tools to handle text will be an ideal method for automatic classification [2]. NLP is one of the most cutting-edge areas for segmenting and investigating the core significance of an English sentence.

This research article on personality classification is classified in 5 segments. The segment 1 is provides the introduction to this publication, whereas section 2 describes the related works done on this topic in the past. Section 3 presents the proposed idea of our technique and section 4 discusses the evaluation paradigm along with the results obtained. Finally, Section 5 provides the conclusion and the directions for future enhancement to this research.

## II. Literature Survey

Gunaseelan B [3] states that in this publication different characteristics from the text have indeed been retrieved and the researchers have constructed a classification model to forecast whether one message body in a curriculum vitae is heading or not. According to the findings, XG boost surpasses some other classifications in identifying headers. The authors effectively recovered skillset content from each of the Portable document format and Microsoft files after predicting it using just a fuzzy search strategies focused on the distance measure. Data that was individually marked provided a more accurate picture of the data formation process. As a result, it specialized well on resumes that had not been seen.

Ramraj S [4] has analyzed multiple state-of-the-art techniques employing comparable self-generated information, allowing the researchers to evaluate them objectively. The most difficult challenge was data production since the data provided by the researchers is especially sensitive and not widely obtainable to everyone. Furthermore, because the data was raw and contained more abnormalities, the authors cleansed it up many times. After all of the effort, the findings were encouraging. As a consequence, when contrasted with comparable state-of-the-art algorithms, CNN achieves the best outcomes; it can therefore be inferred that CNN may also be used to classify text instead of simply pictures and get extraordinary outcomes in comparison to certain other text classification methods.

Qiqiang Xu [5] explains that when training on a résumé block collected data composed of genuine resume fragments, the standard classification model will come across the issue of imbalanced classifications. To address this issue, the writers trimmed the number of component classifications. Expanding the number of units will result in additional data collecting and labeling labor, while lowering the amount of blocks will result in over-fitting as well as under issues. The proposed B-BRNN model fits the observed block allocation better. Because of the RNN structure's capability to comprehend the probabilistic model of distinct block locations in the resume, the performance is relatively optimum even when the sample distribution is imbalanced.

Maksym Lupei [6] expresses that a consultant is often directed by a collection of applicant buzzwords as the primary marker to assess the level of appropriateness for a certain position when picking an applicant. The automation of this task, together with the right evaluation of the examined factors, considerably improves the recruiting agency's productivity. The computational experiment demonstrates that the proposed technique and software product may be employed in real-world recruiting systems. The next approach is to expand the amount of inputs that might be associated with a candidate's fitness for a certain post, which should greatly enhance the required output as proven in this study article.

Yunxiang Liu [7] presents a brief text categorization model that aggregates text relationships using the attention mechanism the authors suggest using WAN to acquire the relationships between words, then LAN to discover the local dependencies, and lastly CAN to consolidate these relationships. In the investigation, similarity measure is employed for data contention, which specifically prohibits the algorithm from overfitting. The experimental findings demonstrate that the suggested multi-attention network achieves excellent classification outcomes.

Andrey Kapitanov [8] demonstrates that when completing the problem of text document classification, we can only utilize a limited quantity of input data Input increases nonlinearly, as do resources like as memory and time. There may be times when it is necessary to use proprietary or commercial datasets. This research provides a technique that relies on statistical variations dependent on data set size that would estimate predicted accuracy while saving resources. Despite certain preconceptions, the outcomes were extremely positive. On the new data, a probabilistic evaluation of categorization quality may be conducted by calculating the least quantity of data necessary for a particular accuracy.

R. S. Jagdale [9] states that there has been an increase in the popularity of the automated resume processing systems. These have been critical in the evaluation of the suitable candidates that are required for the recruitment by the employers. The manual processing of the resumes can be detrimental as it can take a large amount of time to achieve the goals of recruitment that can hamper the productivity of the organization considerably. As the number of candidates is increasing consistently, this time constraint also increases dramatically, which leads to a lot of problems which warrant the need for an effective framework for the realization of automatic extraction of resume segments.

Meng-Jin Wu [10] discusses how AI technology has advanced tremendously in current history every firm would want to use deep learning model in effort to expand its earnings. The authors trained the computer using web crawler technologies and Jieba character segmentation in this article. This will assist journalists in reducing their labor utilization. It will assess which segments the readers are interested in on social media. Then it may exactly propose the news to users. The more important the feature, the quicker it is to recognize. However, a piece of news may fall under more than one category. This is another explanation for the decreased accuracy. Furthermore, everyone has multiple interpretations, perspectives, and previous knowledge, which will result in distinct categorization. All of these challenges must be resolved in order to enhance news categorization.

Meng Zhang [11] narrates that this paper investigates the sentiment classification task of online based marketplace comments, recommends an emotion classification method predicated on sentiment words and CNN automated system, and employs e-commerce marketplace data for a particular amount of time to validate the model's accuracy. The dataset's comments have been evaluated as data sources. The results suggest that using sentiment dictionaries to help with recognition and classification of data sets enhances effectiveness and reduces experimental time. Furthermore, the emotion classification algorithm described in this research has a greater classification quality than the SVM model.

Xi Peng [12] explains how this study carried on the experience from the text summarization area and optimized the underlying network topology of RNN and LSTM According to the findings, the scientists suggest that, except if a bidirectional organism is utilized, RNN and LSTM have no evident benefit in text processing. As a result, the addition of a two-way network structure provides critical support for the text's situational sentiment classification. This structure can increase the model's prediction accuracy while also reducing overfitting. RNN and LSTM that have combined the unidirectional mechanism have a faster precision fitting rate and a faster loss drop rate than a basic neural network.

Satyen M. Parikh [13] describes that the sentiment analysis makes use of NLP approach for analyzing sentiments in tweeter data. This task involves several steps of preprocessing, feature extraction and classification. In this paper, various classification algorithms are compared for analyzing sentiments. This work is based on the analysis of twitter data to study the sentiments of users. The feature extraction phase will tokenize the sentences for the further processing. The machine learning algorithm will take input of the training and testing set for the sentiment analysis. N-Gram technique has been used for feature extraction. Thereafter k-nearest neighbor classifier is used for classification of the tweets into positive, negative and neutral classes.

Yangyang Lan [14] expresses that In this paper, two models for text classification are presented, which are semantic features for text classification and Char-CNN-SRCLA. They are different from the previous methods that improve performance by mixed linguistic features. The novel models are able to alter linguistic features and select more semantic features for text classification. Semantic features for text classification first utilizes a stacked structure to alter linguistic features, and then uses a novel attention method. The method is a cross-layer attention capable of refining the altering process. The Char-CNN-SRCLA further introduces morphological features at the lower layer of semantic features for text classification to validate the filtering process.

## III. PROPOSEDSYSTEM

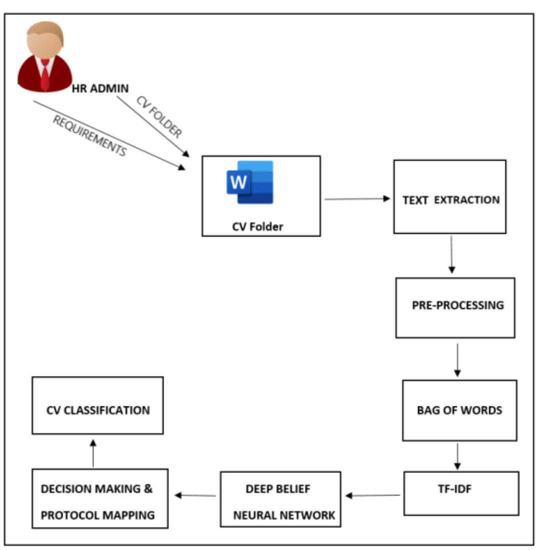


**Figure 1: System Overview**

The presented technique for Classification of the CVs is illustrated in figure 1 given above and the procedure that is executed to implement this system is detailed below.

***Step 1:*** *Data Collection* –This is the first step of the proposed methodology in which the requisite data is provided to the system for the purpose of utilizing the Curriculum Vitae for personality classification. In this step of the procedure a number of CV's are provided as in input from a folder. This folder contains a number of CVs that are supposed to be classified. The classification is done through a given criteria. The criteria for the classification of the CVs are also provided as an input to the system, where the gender and qualification form the basis of classification. The qualifications can be added by accessing the qualifications option from the settings menu, here the user can add the qualifications that are required for the job.

The resume directory is provided as the input and resume contained in the path are extracted and read. The Apache POI is used for the purpose of reading the resume doc files in a line by line manner. All the CVs in the folder given as an input are read through the Apache POI line by line. The extracted contents of the resume are then subsequently stored in a list and transferred to the next step of the procedure.

***Step 2:*** *Pre-processing* – The objectives section in the CVs achieved in the previous step as a feature is taken as an input for the pre-processing step. The objective section from the CV is extracted in a string format and subjected to the preprocessing procedure as detailed below.

*Special Symbol Removal* – Special Symbols are symbols that are used for the purpose of providing a structure to the English language while speaking. These are not as important in our implementation as they do not provide any additional information and are subsequently purged from the string.

*Tokenization* – The special symbol removed string is then given as input for this procedure which performs the tokenization of the string. This process divided the string into smaller tokens for the creation of a well indexed string. This is done to enable the easy conversion of the string into an array list for utilization in the further processing in the proposed methodology.

*Stopword Removal* – The English language has a collection of words that are purely to provide aesthetic to the language and the sentence formation. These words are known as stop words that denote pauses. These are specifically used for joining two sentences together to form an uninterrupted flow of the conversation. These words do not provide any additional meaning to the string and can be removed without any change in the meaning of the string.

For example, the phrase "going to walk" if subjected to the stopword removal procedure of the preprocessing. The stopword in this phrase is "to" which is be removed and the phrase transformed into "going walk". This example depicts that the stopwords removal procedure does not change the meaning of the string.

*Stemming* – In the English language there are a lot of words that are added with a suffix to denote the timing or the tense of the words. This is done to ensure that the conversation is precise and to the point for the individual. This is not necessary in out implementation as it does not change the semantics of the string in any way. Therefore, in the stemming process the words are purged off their suffixes and this makes processing the newly shortened string significantly easier and faster.

For example, "sleeping" will be reduced into "sleep" by the removal of the substring "ing" which is replaced with an empty character in its position. It can be noticed that there is no semantic difference between "sleeping" and "sleep". But stemming can have a significant impact on the time and resources needed for the processing of the string in the methodology further.

***Step 3:*** *Bag of Words* – The bag of words is a collection of predefined words that are selected for the purpose of personality estimation. The words describing the personality of the candidate from the objective and hobbies are utilized in this step.

The extracted words are utilized for the purpose of comparison through the use of the protocol estimation module in the proposed methodology. The protocol estimation extracts the relevant hobbies to achieve the nature of the applicant. The relevant character extracted from the words used in the objective according to the link and stored into a list. The module then compares it with the bag of words and produces an estimation of character according to the candidates. The character of the applicant can be any of the following, Agreeable, conscientious, extrovert and neuroticism.

***Step 4****: Deep Belief Network* –This is perhaps the most important and difficult stage of the process. This is the point where the applicant is evaluated. The TF-IDF list is used as an input in this stage of the approach to evaluate the hidden state and output neurons. The features obtained earlier such as Term Frequency and Inverse Document Frequency were utilized as input for this step of the method. This is advantageous since it allows for accurate and quick candidate evaluation.

This component receives a TF-IDF list of the dataset's required properties as input for assessing the hidden and output layers. Deep Belief Networks operate on three layers: input layers, hidden layers, and output layers. The achieve features are merged with the provided random weights to create the three hidden layers. The bias weights are combined with both the two weights for every feature.

The ReLU activation function is being used to evaluate the five hidden layers as shown in the equation 1 given below. The acquired hidden values are sent to the output layer, which calculates the output error probability.

$$f(x) = max(0, x) \text{---- (1)}$$

The hidden layer values, and also the weights and bias weights, are used to produce the output layer values. These figures, along with the two target variables T1 of 0.01 and T2 of 0.99, are utilized in equation 2 to calculate the error probability rate.

$$Error\ Probability = \sum \frac{1}{2}(\mathrm{T}_0 - O_L)^2 \text{----- (2)}$$

Where,
T = Target Values
$O_L$ = Output Layer Values

The error probability estimate mentioned above is then appended to the end of the row. This is repeated for each quality. The resultant list is ordered in descending order of the error probability rate, with the higher numbers indicating the least amount of error. The dependability of the produced probability, which will be used in the next phase of the classification procedure, is inversely proportional to the error probability percentage. The top 80% of the list is selected and then provided to the Decision Making and Protocol estimation module to complete the candidate evaluation. The technique of hidden layer evaluation is demonstrated in Algorithm 1.

ALGORITHM 1: Hidden Layer Estimation

```
//Input: TF-IDF List TFI_L, Weight set W_S= { }
//Output: Hidden Layer value list H_LV
hiddenLayerEstimation (TFI_L, W_S),index=0
Start
H_LV =Ø {Hidden Layer value]
  for i=0 to size of TFI_L
      ROW= TFI_L [i]
         for j=0 to size of ROW
                X=0
        for k=0 to N [Number of Neurons]
                ATR=ROW[j]
                    X = X + (ATR* W_S[index])
                    index++
            end for
         H_LV= reLUmax(0, X)
      end for
   end for
   return H_LV
Stop
```

***Step 5:*** *Decision Making and Protocol Mapping* – This list of comparisons produces a consolidated list which eventually contains each candidate's name and obtained personality, respectively. Any duplicates attained previously are eliminated in this step of the procedure. This obtained unique list examined thoroughly for the user's list for the obtained personalities. This process ultimately yields the classified candidates according to the requirements of the recruiter with their relevant qualifications and personality that is displayed in an elegant user interface.

## IV RESULT AND DISCUSSIONS

The proposed methodology for Curriculum Vitae Classification is designed using java programming language on a windows based machine through the utilization of the NetBeans IDE. The technique has been deployed on a laptop comprising of a standard configuration such as an Intel Core i5 processor assisted by 1 TB of storage along with 8 GB of RAM.

For the assessment of the accuracy of the system the presented technique is extensively tested through rigorous experimentation as stated below.

### Performance Evaluation through Precision and Recall

Precision and recall are two really useful approaches for understanding how appropriately a certain module in our approach is used. A module's precision defines its relative correctness and offers a wide range of reliability.

The precision metric was calculated using our technique as the ratio of correct CV classifications to total CVs provided. The recall criteria, on the other hand, supplements the accurateness metric and assist in determining the exact effectiveness of the Deep Belief Network module.

This evaluation calculates recall as the proportion of correct CV classifications to total number of inaccurate CV classifications. The following equations quantitatively describe this process.

Precision can be depicted as below

- ✓ TP (True Positive) = The number of accurate CV classifications for the given input CVs
- ✓ FP (False Positive) = The number of CV classifications for the given input CVs
- ✓ FN (False Negative) = The number of accurate CV classifications that are not done for the given input CVs
- ✓ FP (False Positive) = The number of inaccurate CV classifications not done for the given input CVs

So, precision can be defined as

Precision = (TP / (TP + FP)) *100
Recall = (TP / (TP + FN)) *100
F Measure = 2*(Precision*Recall)/ (Precision + Recall)

Table 1 below summarizes the experimental results acquired using the aforementioned formula. Figure 2 shows how these tabular information are combined to produce a visual representation in the form of a line graph.

| No of CVs | No. of accurate CV Classifications (True Positive - A) | No. of inaccurate CV Classifications (True negative - B) | Relevant Classifications not done for input CVs (False Negative-C) | Irrelevant CV Classifications not done (False Positive - D) | Precision | Recall | F Measure | Accuracy |
|---|---|---|---|---|---|---|---|---|
| 10 | 10 | 0 | 0 | 1 | 100 | 100 | 100 | 100 |
| 20 | 17 | 1 | 2 | 3 | 94.4444 | 89.47 | 91.891892 | 86.95652 |
| 30 | 27 | 1 | 2 | 2 | 96.4286 | 93.1 | 94.736842 | 90.625 |
| 40 | 35 | 2 | 3 | 12 | 94.5946 | 92.11 | 93.333333 | 90.38462 |
| 50 | 44 | 4 | 2 | 15 | 91.6667 | 95.65 | 93.617021 | 90.76923 |

Table 1: Precision and Recall Measurement Table

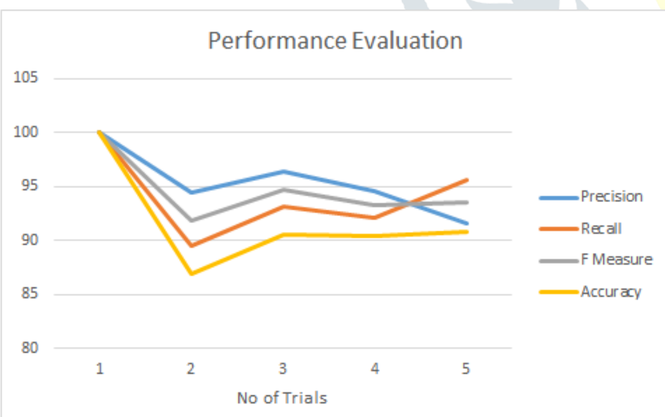


Figure 2: Comparison of Precision, Recall, F-Measure & Accuracy

The graph demonstrates the Deep Belief Network's efficacy in classification of CVs for a given number of input CVs. The approach's outstanding efficiency is demonstrated by precision and recall ratings of 95.42 percent and 94.06 percent, respectively. These figures are quite beneficial and substantial for a first-time realization of such an approach.

The precision, recall, and accuracy scores examined for CV classification revealed the suggested system's effectiveness in great detail. The proposed method was successfully compared to the methods described in [15]. Our approach has a precision of 95.42 percent and an accuracy of 91.74 percent. The correlation of the Naïve Bayes, SVM and Random Forest based CV Classification strategy with the proposed methodology using DBN is shown in table 2 below in a tabular style.

| Performance Metric | Our approach (DBN) | Naïve Bayes[15] | SVM [15] | Random Forest [15] |
|---|---|---|---|---|
| Precision | 95.42 | 52.1 | 45.2 | 44.8 |
| Recall | 94.06 | 45.2 | 59.7 | 68.3 |
| F - Measure | 94.71 | 44.8 | 59.4 | 67.8 |
| Accuracy | 91.74 | 45 | 60 | 70 |

Table 2: Precision, Recall, F-Measure & Accuracy comparison

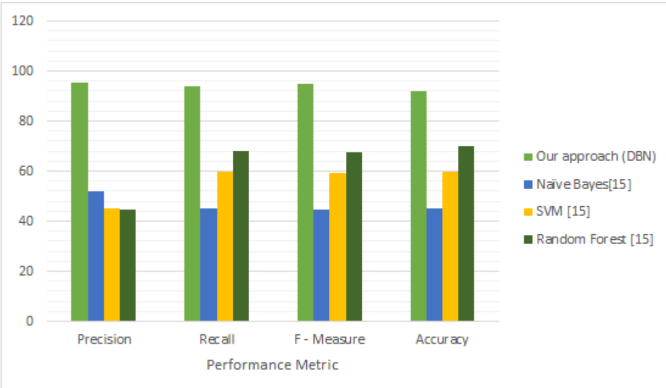


Figure 3: Comparison with Naïve Bayes, SVM and Random Forest based technique depicted in [15]

As seen in Figure 3, the deep learning methodology proposed in this research paper outperforms the Naïve Bayes, SVM and Random Forest based CV classification approaches proposed in [15]. This is owing to the Deep Belief Network that has been implemented to significantly improve the accuracy of CV classification. These findings are extremely satisfactory because the given system achieves the classification accuracy indicated by the performance scores.

## V CONCLUSION AND FUTURE SCOPE

The presented approach for the purpose of achieving the Curriculum Vitae classification through the use of deep learning techniques has been elaborated in this research article. The approach is initiated by the recruiter by registering and then gaining access into the system through the use of the relevant login credentials. Once the user is logged into the system they can update their profile or change any aspects of their profile. The recruiter can then add a list of qualifications that are required for the particular job posting. The recruiter can also add the type of hobbies the person is interested, such as entertainment, exercise, knowledge, music and sports. The recruiter also gets the ability to evaluate the character of the candidate based on the words used in the objective that classifies them into categories such as, agreeableness, conscientiousness, extroversion and neuroticism. Once these categories are initialized, the recruiter can select the qualifications and also the gender and provide the directory of the resumes collected for this purpose. After the system achieves the directory, all the resumes and their contents are read and pre-processed. The pre-processed text is then provided to the bag of words module for the evaluation of the contents. The contents are evaluated and the resumes with the respective qualifications are filtered out and provided for the Term Frequency and inverse document frequency assessment. The values of TF-IDF are provided in a list format to the Deep Belief Network which performs hidden layer and output layer assessments using the ReLU activation function. The achieved error probability values are then provided to the Decision Making and Protocol Mapping module which classifies and extracts the personality of the candidate based on the objective and presents are resultant list to the recruiter. The output has been evaluated for its precision, recall, F Measure and Accuracy and compared with the leading CV classification approaches that resulted in a highly satisfactory outcome.

The future research on this topic can per performed to improve the proposed CV classification approach to be able to execute in a real time scenario for an organization.

## REFERENCES

[1] H. H.A.S.S, H. K.U, J. K.M.J.B, H. K.P.C., A. N. Senarathne and J. Wijekoon, "Smart Human Resource Management System to Maximize Productivity," 2020 International Computer Symposium (ICS), 2020, pp. 479-484, doi: 10.1109/ICS51289.2020.00100.

[2] A. Chaudary, Z. Nasar, M. M. Mubasher and S. W. ul Qounain, "Extraction of Useful Information from Crude Job Descriptions," 2020 IEEE 23rd International Multitopic Conference (INMIC), 2020, pp. 1-4, doi: 10.1109/INMIC50486.2020.9318132.

[3] B. Gunaseelan, S. Mandal and V. Rajagopalan, "Automatic Extraction of Segments from Resumes using Machine Learning," 2020 IEEE 17th India Council International Conference (INDICON), 2020, pp. 1-6, doi: 10.1109/INDICON49873.2020.9342596.

[4] S. Ramraj, V. Sivakumar and K. Ramnath G., "Real-Time Resume Classification System Using LinkedIn Profile Descriptions," 2020 International Conference on Computational Intelligence for Smart Power System and Sustainable Energy (CISPSSE), 2020, pp. 1-4, doi: 10.1109/CISPSSE49931.2020.9212209.

[5] Q. Xu, J. Zhang, Y. Zhu, B. Li, D. Guan and X. Wang, "A Block-Level RNN Model for Resume Block Classification," 2020 IEEE International Conference on Big Data (Big Data), 2020, pp. 5855-5857, doi: 10.1109/BigData50022.2020.9377771.

[6] M. Lupei, A. Mitsa, I. Povkhan and V. Sharkan, "Determining the Eligibility of Candidates for a Vacancy Using Artificial Neural Networks," 2020 IEEE Third International Conference on Data Stream Mining & Processing (DSMP), 2020, pp. 18-22, doi: 10.1109/DSMP47368.2020.9204020.

[7] Y. Liu and Q. Xu, "Short Text Classification Model Based on Multi-Attention," 2020 13th International Symposium on Computational Intelligence and Design (ISCID), 2020, pp. 225-229, doi: 10.1109/ISCID51228.2020.00057.

[8] A. Kapitanov, V. Troyanovskiy, I. Kapitanova, V. Dorogov and E. Dorogova, "Analysis of the Statistical Fluctuations of Text Documents Classification Accuracy," 2020 IEEE Conference of Russian Young Researchers in Electrical and Electronic Engineering (EIConRus), 2020, pp. 1964-1968, doi: 10.1109/EIConRus49466.2020.9039417.

[9] R. S. Jagdale and S. S. Deshmukh, "Sentiment Classification on Twitter and Zomato Dataset Using Supervised Learning Algorithms," 2020 International Conference on Smart Innovations in Design, Environment, Management, Planning and Computing (ICSIDEMPC), 2020, pp. 330-334, doi: 10.1109/ICSIDEMPC49020.2020.9299582.

[10] M. -J. Wu, T. -Y. Fu, Y. -C. Chang and C. -W. Lee, "A Study on Natural Language Processing Classified News," 2020 Indo – Taiwan 2nd International Conference on Computing, Analytics and Networks (Indo-Taiwan ICAN), 2020, pp. 244-247, doi: 10.1109/Indo-TaiwanICAN48429.2020.9181355.

[11] M. Zhang, "E-Commerce Comment Sentiment Classification Based on Deep Learning," 2020 IEEE 5th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA), 2020, pp. 184-187, doi: 10.1109/ICCCBDA49378.2020.9095734.

[12] X. Peng, "A Comparative Study of Neural Network for Text Classification," 2020 IEEE Conference on Telecommunications, Optics and Computer Science (TOCS), 2020, pp. 214-218, doi: 10.1109/TOCS50858.2020.9339702.

[13] S. M. Parikh and M. K. Shah, "Analysis of Various Sentiment Analysis Techniques of NLP," 2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV), 2021, pp. 1323-1331, doi: 10.1109/ICICV50876.2021.9388525.

[14] Y. Lan, Y. Hao, K. Xia, B. Qian and C. Li, "Stacked Residual Recurrent Neural Networks With Cross-Layer Attention for Text Classification," in IEEE Access, vol. 8, pp. 70401-70410, 2020, doi: 10.1109/ACCESS.2020.2987101.

[15] Riya Pal, Shahrukh Shaikh, Swaraj Satpute, Sumedha Bhagwat, “Resume Classification using various Machine Learning Algorithms”, ITM Web Conf. 44 03011 (2022)
DOI: 10.1051/itmconf/20224403011.

*****