International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

Recruitment Assisting Platform

Rahul kumar[1], Shubham kumar jha[2], and Bhat Geetalaxmi Jairam[3]

Department of Information Science and Engineering
The National Institute of Engineering, Mysuru-570008

Abstract—This paper presents an idea of segregating resumes into different classes or groups. As we know there are many different classes or groups in every small or big scale organization. These classes or groups must be utilized to enhance the Resume based search Engine, since many application or resume arrives to recruiters. This creates a recommender system which will prompt the recruiter with the skills related to the filters provided to enhance the search. Association based mining is used in which recruiter just needs to provide the post or position in the organization and the tool should provide the recommended candidates.

Keywords—Recruitment Assisting Platform; Job-resume Match-ing; Candidate's Screening; structured profiles;Topic Modelling; Latent Dirichlet Allocation.

1. Introduction

In recent years, as world is modernizing a rapid development can seen in field of modern Information and Communication Technologies(ICTs)[3]. Now a days, Communication between working professionals or between organizations or between organizations and professionals is basically done through some online means or portal. This modernization has resulted in an increasing number of people turning to the web for job seeking and career development. A lot of companies use online job portals to hire rightful candidate. These online job portals have started to receive enormous number of job application from job seekers which have different working background, having different work experiences and skills. That's why recruitment is considered among most challenging functions for job portals to find and hire the right talent from a wide and heterogeneous range of candidates as it is difficult to manually process and analyze uploaded unstructured document. To address this challenge, many companies have shifted to exploiting e-recruiting platforms.[7] Automating the process of analyzing the applicant profiles to determine the ones that fit the posi-tion's specifications could lead to an increased efficiency[3]. These platforms reduce the cost, time and effort required for manually processing and screening applicant resumes[7]. The use of online recruitment has led to 44% of cost savings and reduced the time to fill a vacancy from 70 to 37 days[3].

Till now several techniques/approaches have been employed by online recruitment systems. Some of these techniques are Analytic Hierarchy Process , Boolean Retrieval, Semantics based techniques, models based on Relevance Feedback and Natural Language Processing (NLP) and Machine leaning based approaches. Although these tech-niques have good matching results,but they are still having following problems. First, the use of automated keyword-based techniques to match an ever-increasing number of resumes (usually in the form of unstructured text) to job posts is unsatisfying since it ignores the semantic aspects of the concepts encoded in the processed documents. Second, semantics and knowledge-based techniques have drawbacks associated with the limited domain coverage and semantic knowledge incompleteness problems highlighted in. Third, suffer from the discrepancies associated with inconsistent CV formats, structure and contextual information[5]. Ma-chine learning techniques try to overcome lots of problem by utilizing machine learning techniques to first classify job posts and resumes under their relevant occupational cate-gories. Although these techniques have proven to be more efficient (i.e. have low run time complexity), they suffer from high error rates and low classification accuracy[7].

To overcome the above mentioned limitations, we propose a Recruitment Assisting Platform that will segregate the resumes into different classes (job categories) to enhance the Resume based Search Engine. We present a prototype of novel implementation of recommendation algorithm that applies topic modelling technique - Latent Dirichlet Allocation (LDA) on jobs data. Topic modelling is the process of discovering hidden patterns that reflects the underlying topics in documents. The main idea of topic models is that it considers entire corpus as a collection of documents, each document as a mixture of topics and each topic as a collection or distribution of terms. It is based on probabilistic distribution of words and topics in documents[4].

We summarize the major contributions of our work as follows:

Segregation of the resumes into different occupa-

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

tional categories.
Employing a recommender system which will prompt the recruiter with skills related to filters provided to enhance the search.
Exploiting Association based mining in which re-cruiter just needs to provide the post in the or-ganization and the tool will provide recommended candidates.

The content of rest of the paper is as follows. In section 2, we introduce the work related. Section 3 describes an overview of the proposed system's architecture. In section 4, we provide the detailed view of our proposed system.Section 5 contains Experimental result and measure the efficiency of the proposed system . In section 6, we discuss the conclusions and outline future work.

2. Related Work

Within this overview,we use many techniques and we are here aiming to describe advancements done in on;line Recruitment through the use of semantic processing tech-niques. Despite of many surveys we have touched to some extent one or more visions of semantic based processing, there is no study offering an overall idea about the bene-fits of semantic based matching when designing, creating and exploiting advanced systems for automation of on;line recruitment processes.

2.1. Boolean Retrieval

Retrieval in existing e-recruitment system is on exact match between applicants stored profiles and enquirers re-quest. These profiles are captured through online forms whose fields are tailored by recruiters and hence, applicants sometimes do not have privilege to present details of their worth that are not captured by the tailored fields thereby, leading to their disqualification. This paper presents a 3-level system which models serialized documents of the applicants worth and they are then analyzed using document retrieval and natural language processing techniques for a human like assessment. Its presentation tier was developed using java server pages and second level functionalities using web service technology. The data tier models resumes that have been put into tokens and tagged using Brill Algo-rithm with my follow-up. Within the second level, indexing was achieved using an inverted index whose terms are noun phrases extracted from resumes that have been tokenized and tagged using Brill Algorithm. Exact matching might result in too few or too many results. It is hard to translate any query into Boolean expressions.Boolean retrieval is more like data retrieval rather than information retrieval.

2.2. Analytic hierarchy process

Analytical Hierarchy Process (AHP) is a mathematical tool of problem solving. The AHP method has been created after understanding the structure of a problem and the real hindrance that managers face while solving it. The AHP method looks at the problem in three parts. The first part is the issue that needs to be resolved, the second part are the alternate solutions that are available to solve the problem. The third and the most important part as far as the AHP method is concerned is the criteria used to evaluate the alternative solutions. The AHP method understands that although there are several criteria, the magnitude of each criterion may not be equal. For instance if you have to choose between two restaurants, the taste and the waiting time are two factors, however both of them may not have equal importance in your perception. The taste may be far more important than the waiting time and so on. Therefore if you assign weight of 2 to taste and 1 to waiting time, you are more likely to arrive at a restaurant that will best meet your need.

The AHP method has its own issues. The method in-volves higher level mathematics. It is based on the concept of eigen vectors. It is for this reason that performing the calculations pertaining to AHP on an Excel sheet are an ordeal. However, of late software tools have been developed that can perform the calculations. The manager therefore just need to be aware of the AHP process , the calculations are automated.

2.3. Semantic based techniques

Due to the huge amount of content and document repositories stored on the web, the problem of relevant search increases. The ability to access and retrieve relevant information remains a difficult task. The lack of efficient indexing method is still a major problem to information retrieval system. This is mainly because web is only under-standable by human and the information cannot be processed by machine. The traditional retrieval systems have limited abilities to exploit the conceptualizations involved in user needs and content meanings due to inability to describe the relation between search terms. The search engines are keyword based which have not bridge the gap of vocabulary mismatch problem in retrieval system.

2.4. Natural language processing

Natural Language Processing, or NLP for short, is broadly defined as the automatic manipulation of natural language, like speech and text, by software. In simple terms, NLP represents the automatic handling of natural human language like speech or text, and although the concept itself is fascinating, the real value behind this technology comes from the use cases. Natural language processing (NLP) is a field of artificial intelligence in which computers analyze, understand, and derive meaning from human language in a smart and useful way. By utilizing NLP, developers can organize and structure knowledge to perform tasks such as automatic summarization, translation, named entity recog-nition, relationship extraction, sentiment analysis, speech recognition, and topic segmentation. NLP is characterized as a difficult problem in computer science. Human language is

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

rarely precise, or plainly spoken. To understand human lan-guage is to understand not only the words, but the concepts and how they're linked together to create mining. Despite language being one of the easiest things for the human mind to learn, the ambiguity of language is what makes natural language processing a difficult problem for computers to master.

One of the biggest limitation is machine translation (MT). Even google translation cannot guarantee a good translation without any modifications. The alignment and language modelling have been a challenging issue for re-searchers to improve MT. The main disadvantage of "latest NLP technologies" is their dependency on huge computing power. Artificial neural networks are also not capable in matching the efficiency of the brain when it comes to process terabytes of data. As a consequence, deep learning based Natural language processing tools are reduced to analyse samples of the Big Text, which, in the case of email surveillance for example, is just not enough.

3. Overview

In this section, we present an overview of the proposed system's architecture. Here, we have implemented a Recruit-ment Assisting Platform that automates categorial segrega-tion and gives skill set required for a given category. As shown in Figure 1, the proposed system comprises several modules that are organized as follows. First, after getting resumes under different category Pre-processing module is used to tokenize and filter out unnecessary words. Next module creates dictionary and corpus taking output of previ-ous module. The third module is all about building suitable topics. The last module of the proposed system will use text categorization techniques to filter out resume into different categories.

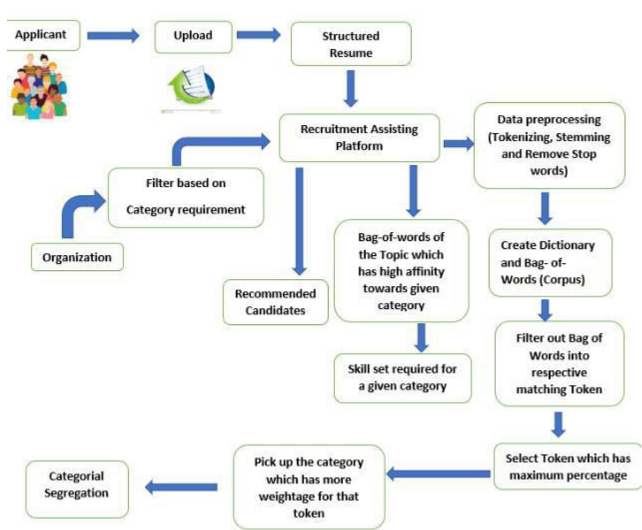

Fig 1. System Architecture

4. Details of Proposed System

Before proceeding to present the details of the methods and techniques used in the proposed system, we introduce – in the context of our work – the terms "Topic Modelling", "Latent Dirichlet Allocation".

Definition 1: Topic Modelling:

In machine learning, topic models are generative models, which provide a probabilistic framework. Topic modelling methods are generally used for automatically organizing, understanding, searching and segregating. The "topics" sig-nifies the concealed, to be estimated, variable relations that link words in a vocabulary and their occurrence in docu-ments. A document is seen as a collection of topics. Topic models find the hidden themes through out the collection and annotate the documents according to those themes. Each word is seen as a part of one of those topics. Finally, A document having collection of topics is generated and it provides a new way to explore the data on the perspec-tive of topics. Topic modelling methods are generally used for automatically organizing, understanding, searching, and summarizing large electronic archives.[6].

Definition 2: Latent Dirichlet Allocation:

Latent Dirichlet allocation (LDA) is a hierarchical Bayesian generative model that allows sets of observations to be explained by unobserved groups that explain why some parts of the data are similar. LDA has made a big impact in the fields of language processing and statistical machine learning and has quickly become one of the most popular probabilistic text modelling techniques in machine learning.

In text pre-processing, punctuation and stop words are excluded(such as, "are", "they", or "to", which contain little topical content). Therefore, each document is regarded as a collection of corpus-wide topics. A topic is consider as a distribution over a set of vocabulary. These topics are generated from the collection of documents. For example, the animals topic has word "cat", "tail" with high probability and the computer topic has word "data", "network" with high probability. Then, a collection of documents has prob-ability distribution over topics, where each word is picked-up from one of those topics. With this document probability distribution over each topic, we will know how much each topic is involved in a document, meaning which topics a document is mainly pointing to[6].

Figure 2 presents the LDA model, which is the funda-mental diagram of it. Each document in the corpus is gener-ated by picking a multinomial topic distribution Dirichlet() for document m. Each word of the document is assigned a topic from the topic distribution. Given the topic, a word is drawn from the multinomial word distribution kDirichlet() from the dictionary for that topic k. W is the topic for nth word in document m and is the specific word[2].

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

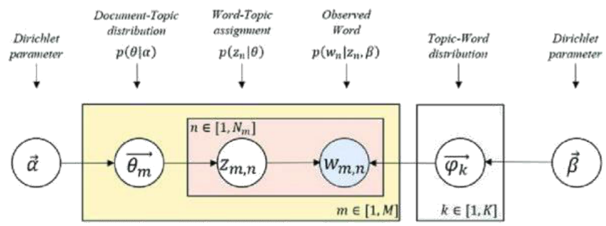

Fig 2: A graphical model for LDA

The method used in our model for approximating the LDA is the online learning algorithm, described by Hoffman et al. which is a variation on the expectation maximization approach. The learning algorithm is streamed and runs in constant memory with regard to the number of documents and can make use of a distributed system, which allows it to be implemented on a much larger corpus. We apply LDA to the job function/job role to find topic level distribution and extract topics out of these job roles on the basis of the user mentioned job roles in his profile. Basically, LDA assigns probability along with most appropriate topic chosen from the job role topic[2].

4.1. Pre-processing

The pre-processing steps consist of loading and preparing the resumes for processing, an essential step for a good analytical result. The first step is to load the resumes into the python environment.The next step is to clean the resumes by removing or altering non-value-adding words.

```
values =
    list("abcdefghijklmnopqrstuvwxyzABCDEFGHI
 JKLMNOPQRSTUVWXYZ0123456789 ")
def remover(my_string = ""):
        for item in my_string:
            if item not in values:
                my_string =
                    my_string.replace(item, "")
        return my_string
```

All words are converted to lower case, and punctuation and white-spaces are removed. Special characters, URLs, and emails are removed, as they often do not contribute to identification of topics. Stop words, misread words and other non-semantic contributing words are removed. Some examples of stop words are “how”, “but”, “make” are some examples of stop words. These words add no value to the topicality of a topic.

```
stop_words = set(stopwords.words('english'))
stop_words.extend(['from', 'subject', 're',
    'edu', 'use'])
for key,value in dict.items():
temp = word_tokenize(value)
token = [w for w in temp if not w in stop_words]

dict[key] = token
```

The loading of resumes into python environment in some instances can cause words to be misread, which must either be rectified or removed. Words are lemmatized means words in third person are changed to first person and verbs in past and future tenses are changed into present. Words are stemmed to their root form for easier comparison.

Function to perform lemmatize and stem preprocessing steps on the data set -

```
 def lemmatize_stemming(text):
    return
        stemmer.stem(WordNetLemmatizer().lemmatize( text,
    pos='v'))

def preprocess(text):
    result = []
    for token in
          gensim.utils.simple_preprocess(text):
        if token not in
              gensim.parsing.preprocessing.STOPWORDS
              and len(token) > 3:
             result.append(lemmatize_stemming(token))
    return result
```

Lastly, many words only occur in a single resume, and these should be removed to make computations easier, as less frequent words will likely provide little benefit in grouping resumes into topics.

4.2. Dictionary and Bag-of-words(Corpus)

Now map the words i.e. output of pre-processing having similar references into unique Topics.

In order to do so, Gensim is used. Gensim is billed as a Natural Language Processing package that does ‘Topic Modeling’. It requires the words (aka tokens) be converted to unique ids. In order to achieve that, Gensim helps to create a Dictionary object that maps each word to a unique id. This is achieved using id2word function of Gensim.

```
#Create a dictionary
  processed = [ v for v in dict.values() ] dictionary-lda =
corpora.Dictionary(process)
```

A ‘token’ typically means a ‘word’. A ‘document’ can typically refer as a collection ‘sentence’ or ‘paragraph’ and a ‘corpus’ is typically a ‘collection of documents as a bag of words. That is, for each document, a corpus contains each word’s id and its frequency count in that document. As a result, information of the order of words is lost. The Corpus is created using doc2bow.. It converts document (a list of words) into the bag-of-words format = list of (token_id, token_count) 2-tuples. Each word is assumed to be a tokenized and normalized string (either Unicode or utf8-encoded). The produced corpus will have a mapping of [word_id, word_frequency].

```
#Creating corpus
corpus = \texttt{[dictionary\_lda.doc2bow( listoftokens)} for
listoftokens in processed]
```

For example, (0, 1) implies, word id 0 occurs once in the first document and [1,4] implies, word id 1 occurs four times.

We have filtered out tokens that appear in :

1) less than no_below documents (absolute number) or
2) more than no_above documents (fraction of total corpus size, not absolute number).
3) after (1) and (2), keep only the first keep_n most frequent tokens (or keep all if None).

no_below is an int that represents a threshold filtering out number of occurrences of the tokens among documents above certain number. Here we have use no_below to filter out words appearing less than 3 times.

no_above is represents faction of total corpus size. This means that tokens appearing in more than 50 % of documents will be removed. It is this Dictionary and the bag-of-words (Corpus) that are used as inputs to topic modeling and other models that Gensim specializes in.

4.3. Building topic model

Now, we have everything required to train the LDA model. In addition to the corpus and dictionary, we have to provide the number of topics as well. Here, we are asking LDA to find 20 topics in the data. Apart from that, alpha and eta are hyper parameters that affect sparsity of the topics. According to the Gensim docs, both have defaults value as 1.0/num topics prior. chunksize gives number of documents to be used in each training chunk . update_every determines how often the model parameters should be updated. passes is the total number of training passes.

```
# Build LDA model
lda_model = gensim.models.ldamodel.LdaModel(
corpus=corpus,id2word=dictionary_lda, num_topics=20, random_state=100,
      update_every=1,   chunksize=1257,   passes=4,
  alpha= [0.01]*num_topics,
      eta=[0.01]*len(dictionary_lda.keys(), per_word_topics=True)
```

With all words combined in all the resumes we have created 20 bunch of words where each denotes a unique TOPIC.

5. Experimental Results

This section describes the experiments that we have carried out to evaluate the techniques of the proposed system. In order to evaluate the efficiency and the effectiveness of the proposed system, we have acquired data from two different datasets - Resumes Dataset (Kaggle), Resumes Entities for NER (Kaggle). The Dataset that we managed to acquire had total of 1388 resumes applied for 46 unique job posts.

The collected resumes are structured documents having the following segments (job title, job description, required skills, years of experience, required education qualifications and additional desired requirements).

	Category	Resume
0	HR	b'John H. Smith, P.H.R.\n800-991-5187 | PO Box...
1	HR	b'Name Surname\nAddress\nMobile No/Email\nPERS...
2	HR	b'Anthony Brown\nHR Assistant\nAREAS OF EXPERT...
3	HR	b'www.downloadmela.com\nSatheesh\nEMAIL ID:\nC...
4	HR	b"HUMAN RESOURCES DIRECTOR\n\xef\x82\xb7Expert...
...	...	...
164	Testing	Computer Skills: â□¢ Proficient in MS office (...
165	Testing	â□□ Willingness to accept the challenges. â□□ ...
166	Testing	PERSONAL SKILLS â□¢ Quick learner, â□¢ Eagerne...
167	Testing	COMPUTER SKILLS & SOFTWARE KNOWLEDGE MS-Power ...
168	Testing	Skill Set OS Windows XP/7/8/8.1/10 Database MY...

1388 rows × 2 columns

The experiments of our system's prototype show that the classification process for the resumes and job posts took approx 1 hours on average on a PC with dual-core CPU (2.1GHz) and (4GB) RAM.

5.1. Identifying the optimal number of topics

To approximate the optimal number of topics, two things were considered. The perplexity was calculated for different amounts of topics, and secondly the need for specificity was considered.

At the extremes, choosing one topic would indicate one topic covering all resumes, which will provide a very coarse view of the resumes. On the other hand, if the number of topics is equal to the number of categories, then a very precise topic description will be achieved, although the topics will lose practical use as the overview of topics will be too complex. Therefore, a low number of topics was preferred as a general overview was required. Identifying what is a low number of topics will differ depending on the corpus of papers, but visualizing the perplexity can often provide the necessary aid for the decision.

Perplexity is a statistical measure of how well a probabil-ity model predicts a sample. The perplexity was calculated over five folds, where each fold would identify 80 percent of the papers for training the model and leave out the remaining 20 percent for testing purposes. Using multiple folds reduces the variability of the model, ensuring higher reliability and reducing the risk of over-fitting[1]. The following number of topics were selected: 5, 10, 15, 20, 25, 30, 35, 40, 45,50. Figure 3 shows perplexity at certain number of Topics.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 04 Issue: 05 | May -2020 ISSN: 2582-3930

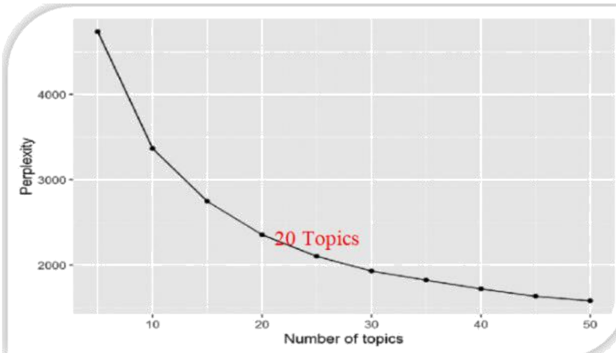

Fig 3. Validation of topic modelling

The goal here is to find the lowest number of topics, which at the same time have a low perplexity. In this case, the slope of the fitted line starts to gradually decline at twenty topics, which is why the selected number of topics is twenty[1].

5.2. Topic Modelling

As the number of topics is chosen, the next step is to run the LDA method on the entire set of resumes. Each topic is a combination of many keywords and each keyword contributes a certain amount of weight to the topic.

Topic 0	0.007*"banking" + 0.006*"bank" + 0.006*"accounting" + 0.006*"Finance" + 0. 005*"tax" + 0.004*"loan" + 0.004*"million" + 0.003*"credit" + 0.003*"finance" + 0.003*"transactions"
Topic 2	0.007*"Less" + 0.005*"NET" + 0.005*"MVC" + 0.005*"Visual" + 0.005*"JavaScript" + 0.005*"Layer" + 0.004*"Framework" + 0.004*"Studio" + 0.004*"Designing" + 0.004*"Net"
Topic 3	0.015*"HR" + 0.008*"Human" + 0.006*"Resources" + 0.005*"recruitment" + 0.0 03*"candidates" + 0.003*"Resource" + 0.003*"managers" + 0.003*"Employee" + 0.003*"benefits" + 0.003*"Recruitment"
Topic 10	0.013*"J" + 0.008*"LM" + 0.008*"Medical" + 0.008*"Association" + 0.007*"Dawson" + 0.006*"Care" + 0.006*"Burke" + 0.006*"Nutrition" + 0.006*"Sports" + 0.006*"Hospital"
Topic 12	0.007*"Flight" + 0.006*"Preparation" + 0.006*"flight" + 0.005*"Attendant" + 0.004*"Pilot" + 0.004*"Air" + 0.004*"letter" + 0.004*"Specific" + 0.003 *"Space" + 0.003*"Force"
Topic 13	0.028*"Law" + 0.011*"Legal" + 0.011*"legal" + 0.008*"Mississippi" + 0.007 *"law" + 0.006*"Bar" + 0.006*"Association" + 0.006*"Conference" + 0.006*"Economics" + 0.004*"Court"
Topic 19	0.004*"Art" + 0.003*"Designer" + 0.003*"Fashion" + 0.003*"Gallery" + 0.002 *"Adobe" + 0.002*"media" + 0.002*"Museum" + 0.002*"content" + 0.002*"managed" + 0.002*"xe2x80xa2"

Table 1. Some topics with top 10 keywords and their weightage

Topic 0 is a represented as 0.007*"banking" + 0.006*"bank" + 0.006*"accounting" + 0.006*"Finance" + 0. 005*"tax" + 0.004*"loan" + 0.004*"million" + 0.003*"credit" + 0.003*"finance" + 0.003*"transactions". It means the top 10 keywords that contribute to this topic are: 'banking', 'bank', 'accounting'. . . ... and the weight of 'banking' on topic 0 is 0.006. The weights reflect how important a keyword is to that topic. Once we have the Topics ready, we can also tell which category has highest affiliation to which topic which will help in recommending the keywords when a specific category is given to us.

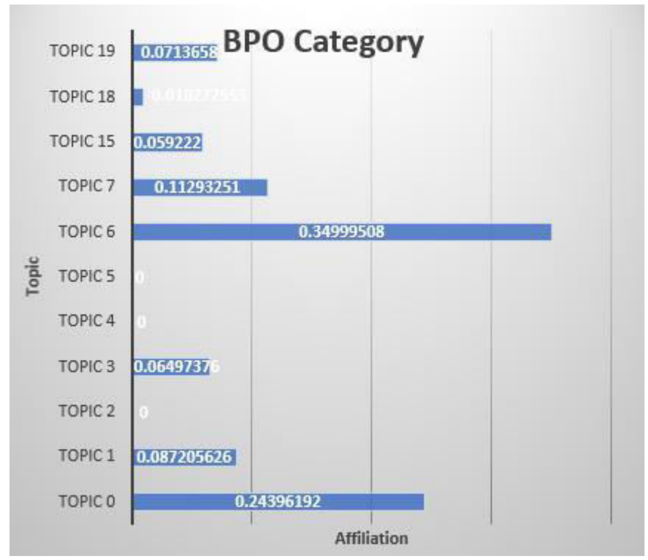

Fig 4. Affiliation of some selected topics for BPO category

The graph in figure 4 clearly shows topic 6 has greater affiliation for BPO category.

5.3. Blending LDA with SVM

We are using supervised learning to determine the category of the resume. In this we have tested many algorithms and compared their efficiency and time taken to train the model. In supervised learning we train our model with sets of labeled input and its corresponding output. We then finally test the output of our model with any random input. In supervised learning, we have compared three main algorithms i.e. Random Forest (RF), Support Vector Machine (SVM) and Linear regression (LR). Figure 5 compares three algorithms based on their efficiency and time taken.

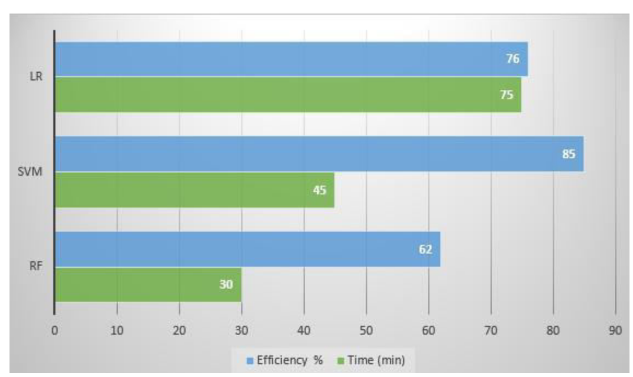

Fig 5. A plot showing Efficiency% and time taken in minutes by three algorithms to run our model.

Based on studies, it is clear that LDA shows good performance than the other topic models and SVM yields better accuracy than the other classifiers. So, we have blended both to enhance performance level of our model. Support Vector Machine is a vector space-based learning method. It is a text categorization technique based on the idea of decision plane that separates a set of objects having multiple class boundaries. The process is segregating the two classes with a hyperplane and selecting a hyperplane with high margin. After training our model with 80% of our dataset, we tested our model with remaining dataset and we got different precision for various categories.

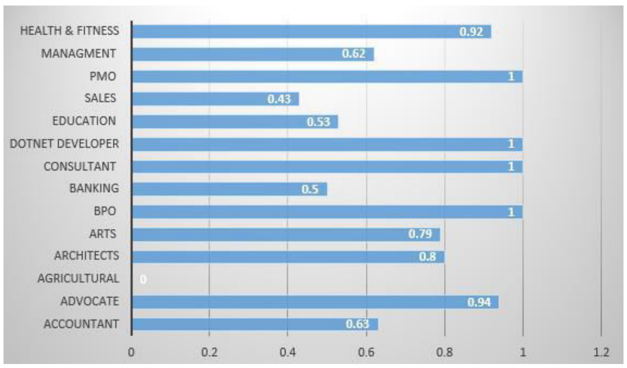

Fig 6. A plot showing some of the categories with their precision.

6. Conclusion and Future Work

In this generation, where the rate of employment has increased, it is a trouble some task for recruiters to recruit new candidates. They have to go through the resume of each and every applicant and have to interview the applicant for hours. With the high precision rate of our model we help recruiters save their time on resume analysis and deciding which profile is suitable for the candidate. The recruiter will also get the skill set that is required for a profile. So if a candidate falls under a profile for which recruiter is having little knowledge, he can use this feature. At last we are helping the recruiter by providing him with the best resume for any given profile that is present with us. The profile must match the profile that is available within the system.So with these features implemented we are helping recruiters save their precious time that is usually wasted on shortlisting the resumes.

For our future work we will work on overcoming the limitations we have faced in our model. We will work on -

The model classifies resumes purely on the content of the resume. We will work checking the authen-ticity of the candidate.
The model has a limited number of profiles (categories). We will try to cover as many categories as possible by taking a larger dataset and a faster algorithm for processing.

References

[1] Claus Boye Asmussen and Charles Møller. Smart literature review: a practical topic modelling approach to exploratory literature review. Journal of Big Data, 6(1):93, 2019.

[2] Shivam Bansal, Aman Srivastava, and Anuja Arora. Topic modeling driven content based jobs recommendation engine for recruitment industry. Procedia computer science, 122:865–872, 2017.

[3] Evanthia Faliagka, Kostas Ramantas, Athanasios Tsakalidis, and Gian-nis Tzimas. Application of machine learning algorithms to an online recruitment system. In Proc. International Conference on Internet and Web Applications and Services. Citeseer, 2012.

[4] Hamed Jelodar, Yongli Wang, Chi Yuan, Xia Feng, Xiahui Jiang, Yanchao Li, and Liang Zhao. Latent dirichlet allocation (lda) and topic modeling: models, applications, a survey. Multimedia Tools and Applications, 78(11):15169–15211, 2019.

[5] Aseel B Kmail, Mohammed Maree, Mohammed Belkhatir, and Saadat M Alhashmi. An automatic online recruitment system based on exploiting multiple semantic resources and concept-relatedness mea-sures. In 2015 IEEE 27th International Conference on Tools with Artificial Intelligence (ICTAI), pages 620–627. IEEE, 2015.

[6] Zhou Tong and Haiyi Zhang. A text mining research based on lda topic modelling. In Proceedings of the Sixth International Confer-ence on Computer Science, Engineering and Information Technology (CCSEIT), pages 21–22, 2016.

[7] Abeer Zaroor, Mohammed Maree, and Muath Sabha. Jrc: A job post and resume classification system for online recruitment. In 2017 IEEE 29th International Conference on Tools with Artificial Intelligence (ICTAI), pages 780–787. IEEE, 2017.