A DECISION METHOD USING NEURAL NETWORKS TOCLASSIFIED JOB DESCRIPTION

[1]SIRAPOP KANOKCHAN, [2]WUDHICHAI ASSAWINCHAICHOTE

[1,2]Department of Electronic and Telecommunication Engineering, King Mongkut's University of Technology Thonburi, Bangkok 10140, Thailand
E-mail: [1]sirapop.k@kmutt.ac.th, [2]wudhichai.asa@kmutt.ac.th

Abstract - The transformation in Industry 4.0 the organization requires more people with IT skills. On the other hand, within the organization personnel are lacking IT skills. Therefore, there must be a solution to the problem such as providing IT skills training or outsourcing IT recruiting. However, IT jobs are complex and have multiple times in the update which makes it a hard decision to classify IT job descriptions. This paper proposes a decision method to use a model of Convolutional Neural Network (CNN) and Natural Language Processing (NLP) techniques to classify IT job descriptions to find insights data of IT skills in many job titles of IT. The IT job description data used for analysis are collected from the online recruitment sites.

Keyword - CNN,MPS, JD Classification, Key Skills.

I. INTRODUCTION

The organization requires more people with IT skills in Industry 4.0 [1]. Many organizations have a problem with lacking IT skills. The solution to the problem can be done by providing IT skills training or outsourcing IT recruiting. However, it is a hard decision to classify IT job descriptions. Because IT jobs are complex and have multiple times in the update. The analysis of IT skills will benefit the recruitment of human resources Personnel training organizational structure design and improving the organization's strategy [2]. However, the need for IT personnel is not only factored into the skill of the personnel. On the other hand, there is also a decision from the discretion of the organization's management as well. They often use discretion in formulating strategies in conjunction with technology to help decision-making [3].

Nowadays, Technology of Convolutional Neural Network (CNN) and Natural Language Processing (NLP) techniques have been used in data analysis to classify text [4] for application in job description work such as job predictions from candidate job description data were analyzed using deep learning technologies including CNN, GRU, and LSTM with the application [5]. IT positions based on similar skills using Auto-Encoder [6] and related skills analysis based on job descriptions using Word2vec techniques [7] etc.

This paper proposes a comparison of the performance of Neural Network Models and Natural Language Processing techniques for use in classification IT job descriptions. Based on 10,000 job description resources on online recruitment websites.

Which can be applied to solve the problem for classification IT job description data in each position. The organization can use in the decision to send the personnel to train for increasing the IT skills. However, data analysis of job descriptions can use in the case when the organization recruits IT Outsourcing to solve the problems to respond to organizational policies in Industry 4.0.

II. RELATED WORK

Hackman and Oldham's proposes a Characteristics Theory [8], a study of job characteristics that affect job performance divided into five factors: Skill Variety, Task Identity, Task Significance, Autonomy, and Feedback. These five factors will affect your mental state. which will influence the level of motivation job satisfaction and the efficiency of the worker's work Thus, the Motivation Potential Score (MPS)(1) equation is created:

$$\text{MPS} = \frac{\begin{matrix}\text{Skill variety} + \\ \text{Task Identity} + \\ \text{Task Significance}\end{matrix}}{3} \times \text{Autonomy} \times \text{Feedback} \quad (1)$$

It can be seen from the equation that if we can add Skill Variety and Task Identity for each job position, it will increase the MPS value of each job position. This has a positive effect on the performance of personnel in the organization. This research uses this concept to create a decision-making method using artificial neural networks for job classification. to find outstanding skills and skills of each IT position to help increase training potential and effective management of IT personnel within the organization.

Word2vec was developed by Mikolov et al in 2013 and in 2019 is used to improve strategies for recruiting people with the right skills[7]. By converting skills into vectors as well. Preprocessing Data and Skip-gram.

Later in 2020, jobs were compared for similar levels of expertise across companies. for use in recruiting talent Job seekers for positions [6], calibrations, and salary forecasts for IT personnel and engineers, offering data-driven enterprise solutions. A graph comparing job titles is created from the job change record. to evaluate real-life datasets using the Word2Vec model combined with Skip-Gram. encode-decode In addition, for sentence analysis, higher MRR results showed the relationship of job title to the questionnaire.

There has been research on how Text CNN is used to optimize text classification [5], especially in conjunction with natural language processing techniques. The neural network has been used in conjunction with pre-trained word embeddings Glove and FastText. In addition, a combination of techniques for text classification for enhanced performance is Bi-RNN, Bi-LSTM model. Bi-LSTM-CRF and Bi-LSTM-CNN models. CNN-LSTM-Glove to be used to classify the nature of the job and in-depth skills analysis of 25 IT job titles.

III. DATA SET

For this research, we will use the Dataset[9] IT job descriptions compiled from 10,000 online job sites, divided into 25 IT job titles, displayed in graphs between job titles. and the amount of job description data, which can be summarized in Fig 1.

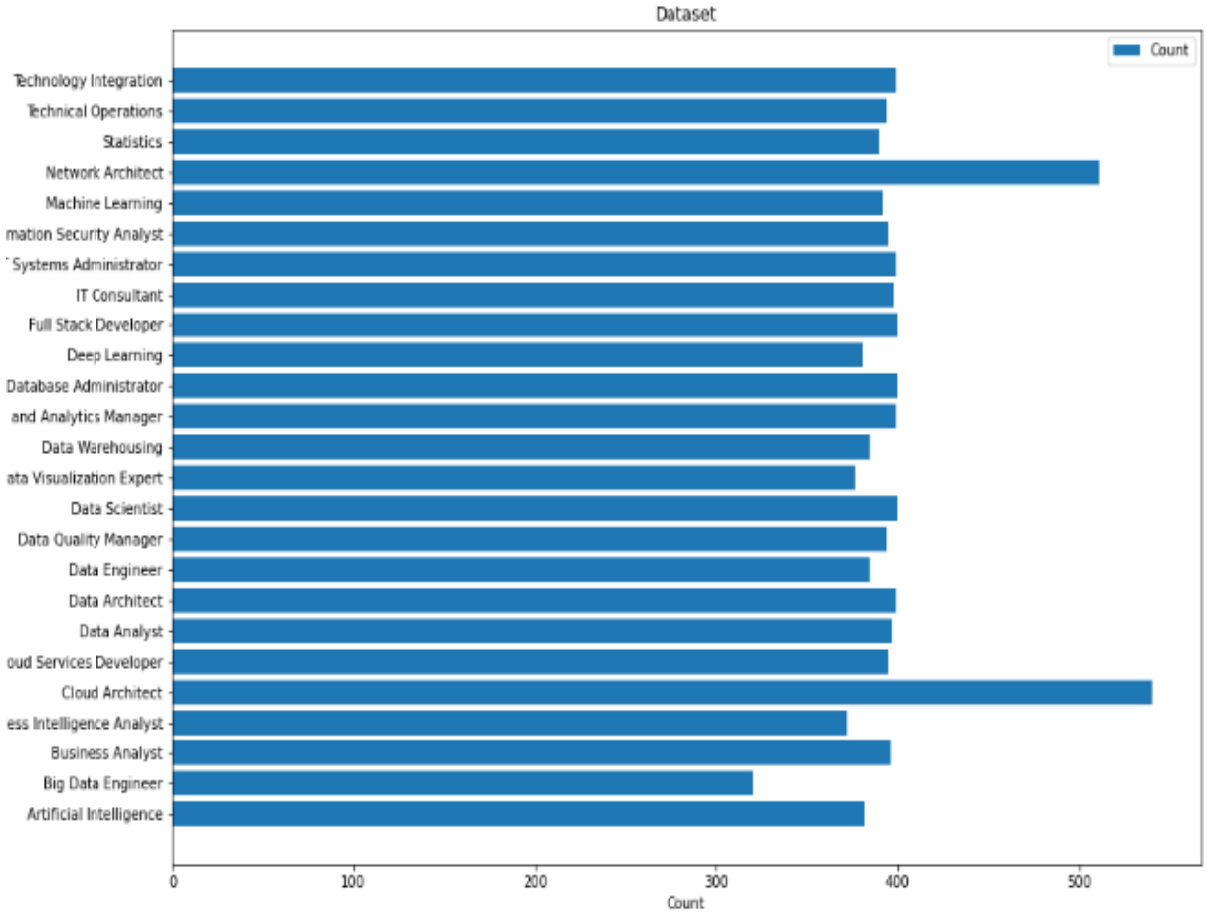


Figure 1: Dataset of IT JD

IV. METHODOLOGY

This research presents neural network-based modeling combined with natural language processing techniques to categorize IT job description datasets. to extract features and skills of each job position This will help analyze to increase the efficiency of decision making to recruit IT personnel from outside or used to consider sending personnel from within to train IT skills in related positions. By finding the best model to solve the problem of personnel classification according to the nature of the job. In which this research chooses Python language to develop together with Jupyter Notebook and Google Colab programs as shown in Fig 2.

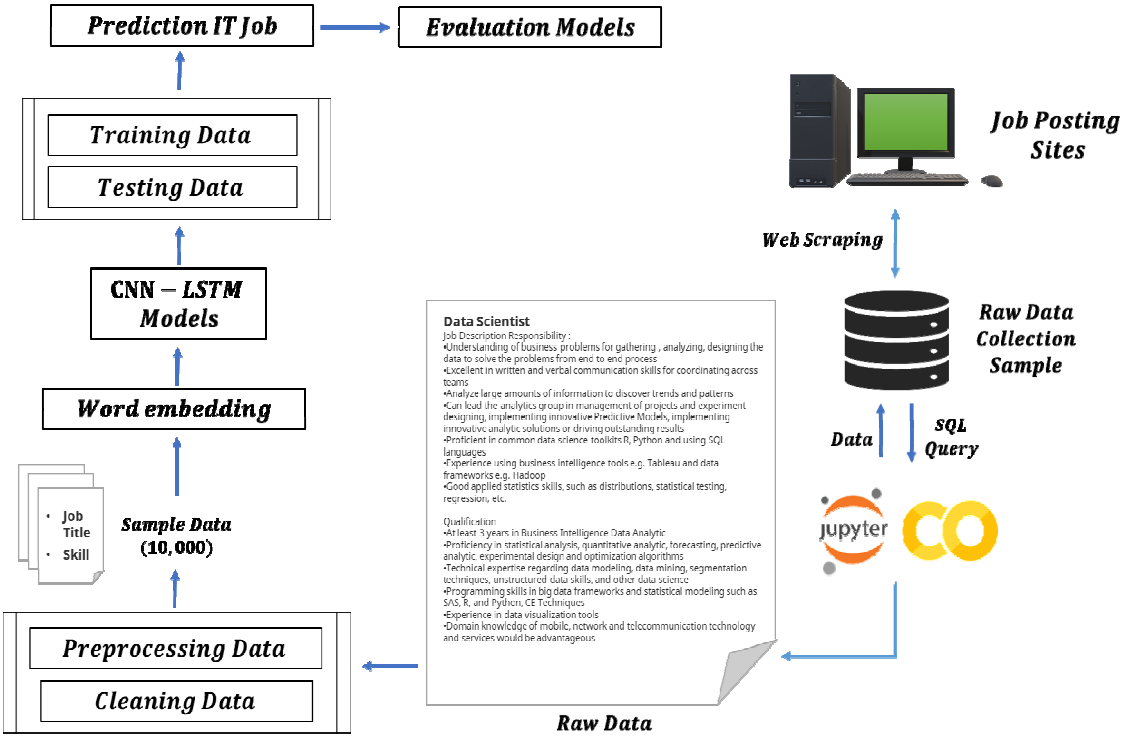


Figure 2: Overview of Process to Classified JD with CNN

A. Word Embedding

Start classifying IT job description data by feeding the dataset as a CSV file into the program for analysis. Then it will go into the data preparation stage by doing randomized job placement data experiments. and 10 job description data with the Sample ( ) command to see the nature of the text within the data. You'll notice that our raw data also contains word clutter in the sentence. Therefore, it is necessary to clean the data (Data Cleaning) before creating a model. Because punctuation must be filtered. Convert all letters to lowercase For example, “look” and “looking” are very different for encoding as most of the text is not pre-processed. Our text is converted into numbers (Encoding) that our model can understand. and can learn In the next step, we will use the .groupby and .count ( ) commands to view the amount of IT job description information for each job position. There are a total of 25 positions. After that, we will divide the DataFrame of all positions into 2sets: Training Set (Training Set) in the ratio of 80 percent or 8,000 items and Test Set (Test Set) in the ratio of 20 percent or 2,000. The list to prepare for training the model (Training Model) and testing the model (Testing Model) starts with the release of the Sklearn library train_test_split. and then use the command train_test_split which set the values test_size = 0.2 and random_state = 8. Next, convert the data that is extracted from the DataFrame of every job position into Series format, separated between Job Title and Job Description. (Description) Next, we will create a token of words from the job description sentence. (Description)" to move on to the next step of encoding and training model, starting with the declaration of the Keras library called Tokenizer, and then assigning a library of 1000 word datasets with the Tokenizer command (vocab_size = 1000) then use the function call .fit_on_texts to train the dataset In the Train process, all punctuation marks are removed and replaced with spaces. And special characters are ! " # $ % & ( ) * + , - . / : ; < = > ? @ [\ \ ] ^ _ ` { | } ~ \t \n out of all, then reduce the characters. From uppercase to lowercase all Each word is then created as a IndexVector, where index 0 is reserved not assigned to any word, and the most frequent word is stored in index 1 of num_words. The top 9 most common words are words that fall into the category of Stop words that are often found in sentences such as and , to , the , of , in , a , with , for and or etc. In this research, we will not cut Stop. words removed due to degradation of model performance and ignoring word context after we have tokens for all job description words We will convert the job description token in letter form to a numeric index indicating each term. It starts by declaring a Keras library called sequence and converting the dataset to an index number with the .texts_to_sequences command. After getting the index numbers, we convert the job description index numbers into a 2D, scalable array. Before importing the CNN model for further training, it starts by declaring a Keras library called sequence, and then converts the job description index to a 2D array with .pad_sequences . Set sequences_length . equal to 1200 and padding = 'post' to arrange an Array in a 2D format, aligned to the left, and replace the blank with 0 to get Training Set and Test Set. Can't understand it in the form of numbers first with the job encoder, starting from the declaration of the Sklearn library called LabelBinarizer, and then encoder the job position with the command LabelBinarizer ( ) by calling a function .fit_transform to convert Job Title Labels into binary format for Train Setand for Test Set. When converting Job Title Labels to numeric format, the computer will be able to understand the Job Title Labels value which number is.

what job position We now have a data set ready for the various Train Models, which can be summarized in Fig3.

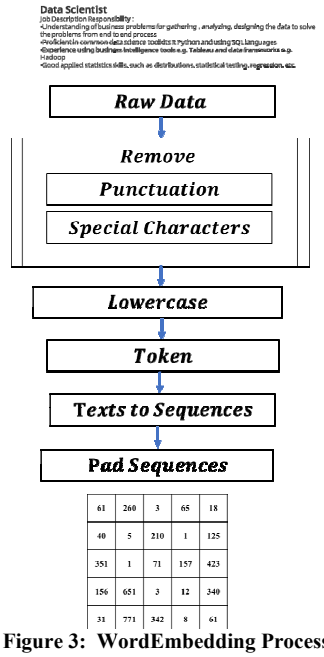


Figure 3: WordEmbedding Process

B. Text CNN Model

To model a CNN to classify text is similar to analyzing an image that can be represented as an array of pixel values. Similarly, we can represent text as an array of vectors. where each term is mapped to a specific vector in the vector space that contains all the terms. In this step, we will bring Data, job descriptions, IT jobs, and Labels for IT jobs. that have already been prepared for the train to be trained in the model But first, we need to model the neural network. In this research, we will start by applying Text CNN as an algorithm to train the deep learning model. Text CNN is divided into 3 layers: Convolution Layer, Pooling Layer, and Fully Connected Layer. The workflow starts with configuring a filter or kernel that helps to extract the object recognition attributes. Usually, a filter or kernel will extract one feature of interest. So we also need multiple filters. to find a combination of multiple spatial features Subsequently, the data passes through the Pooling Layer, which is responsible for reducing the spatial size of the feature. and reduce the computational power required to process the data. Finally, the final output will be flattened. and fed into the neural network in the Fully Connected Layer for classification using the softmax classification techniqueaccording to the process in Fig4.

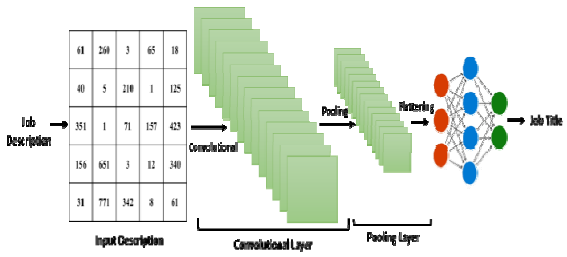


Figure 4: Overview of CNN in Job Classification Model

C. CNN-LSTM Model

LSTM is a type of Recurrent Neural Network (RNN) proposed by Sepp Hochreiter and JuergenSchmidhuber. It was created to solve the problem of data that the sequence is too long. It contains memory to save the previous hidden state. It can also tell when to write, forget (delete) or allow to read. By considering the details of IT skills the architecture of this CNN-LSTM model is shown in Figure 5.

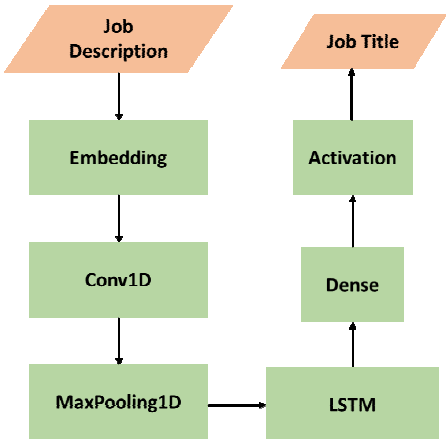


Figure 5: CNN–LSTM Architecture

Initialize the word layer setup by setting the input to a 1200x200 matrix. Each job description contains 2000 words, and each word is represented by 64dimensional word embedding. Before Glove training and FastText, later on Conv1D layer, in this architecture we use 1D spatial drop out with dropout rate of0.2.

V. EXPERIMENTS

In this research We have set up a procedural experiment as follows:

Job Title	Top 3 Skill		
	1	2	3
Artificial Intelligence	AI	Python	Java
Big Data Engineer	Python	SQL	Java
Business Analyst	SQL	Sap	Scrum
Business Intelligence Analyst	SQL	Python	Sap
Cloud Architect	Azure	Aws	DevOps
Cloud Services Developer	Java	JavaScipt	Salesforce
Data Analyst	SQL	Python	Laws
Data Architect	SQL	Python	Azure
Data Engineer	SQL	Kafka	Python
Data Quality Manager	Laws	SQL	Salesforce
Data Scientist	Python	SQL	Java
Data Visualization Expert	SQL	Python	Java
Data Warehousing	SQL	Sap	Laws
Data and Analytics Manager	SQL	Python	Laws
Database Administrator	SQL	Oracle	Linux
Deep Learning	Python	Java	html
Full Stack Developer	JavaScript	Java	React
IT Consultant	Laws	Salesforce	Financial
IT Systems Administrator	Linux	DNS	Unix
Information Security Analyst	Linux	Laws	SQL
Machine Learning	Python	AI	Java
Network Architect	Aws	Linux	Java
Statistics	Laws	SQL	Python
Technical Operations	Laws	SQL	html
Technology Integration	AI	Python	Laws

Table I. SUMMARY RESULTS OF THE EXPERIMENTMODELS

1) Set a random number of IT skills vocabulary from our dictionary to 2000 words.
2) The ratio of the division of training sets and the test set is divided into 80% and 20% ratios.
3) This research examined the IT skills of each of the top 3 jobs by using a majority vote in a sentence. The results were shown in Table I.
4) Neural network model and the evaluation of the model used to classify the IT job description data of this research is in Table II.

Models	Precision	Recall	F1-score	Accuracy
FNN	0.63	0.63	0.62	0.63
TextCNN	0.66	0.65	0.65	0.65
TextCNN – FastText (300)	0.68	0.69	0.69	0.69
TextCNN – Glove (300)	0.65	0.65	0.64	0.65
CNN-LSTM+ FastText (300)	0.71	0.70	0.70	0.70
CNN-LSTM– Glove (300)	**0.71**	**0.71**	**0.71**	**0.72**

Table II . SUMMARY RESULTS OF THE EXPERIMENTMODELS

VI. APPLICATION

The proposed solution to classify job description data using a neural network was designed using CNN-LSTM in combination with Glove, which is efficient compared to other models used in the experiment. It is, therefore, useful to categorize IT job description datasets. to extract features and skills of each job position This will help analyze to increase the efficiency of decision making to recruit IT personnel from outside or used to consider sending personnel from within to train IT skills in related positions.

VII. CONCLUSION

Combining the CNN-LSTM neural network model with Glove will improve the efficiency of job description data classification. In this research, we have tried to find in-depth information about IT skills in each position. which can be used as a guideline for sending personnel to train to increase IT skills pool It also helps to classify information about IT job descriptions in each position for use in HR management.

REFERENCES

[1] Görmüş, A., "Future of Work with the Industry 4.0", The 6th International Congress on Social Sciences (INCSOS 2019), November 15, 2019, Turkey, pp. 317-323.

[2] Pinzone, M., Fantini, P., Perini, S., Garavaglia, S., Taisch, M. and Miragliotta, G., "Jobs and Skills in Industry 4.0: An Exploratory Research", IFIP International Conference on Advances in Production Management Systems (APMS), September 3–7, 2017, Hamburg, Germany, pp. 282-288.

[3] Waschull, S., Bokhorst, J. and Wortmann, H., "Impact of Technology on Work: Technical Functionalities that Give Rise to New Job Designs in Industry 4.0", IFIP International Conference on Advances in Production Management Systems (APMS), September 3-7, 2017, Hamburg, Germany, pp. 274-281.

[4] Minaee, S., Kalchbrenner, N., Cambria, E., Nikzad, N., Chenaghlu, M. and Gao, J., Deep Learning Based Text Classification: A Comprehensive Review, Apr 6, 2020.

[5] Huynh, T., Nguyen, K., Nguyen, N. and Nguyen, G., "Job Prediction: From Deep Neural Network Models to Applications", 2020 RIVF International Conference on Computing and Communication Technologies (RIVF), October 14-15, 2020, Ho Chi Minh City, Vietnam.

[6] Zhang, D., Zhu, H., Liu, Y. and Liu, J., "Job2Vec: Job Title Benchmarking with Collective Multi-View Representation Learning", The 28th ACM International Conference on Information and Knowledge Management, November 3, 2019, Beijing, China, pp. 2763-2771.

[7] Le, V., Vo, M. and An, D., Skill2vec: Machine Learning Approaches for Determining the Relevant Skill from Job Description, Jul 31, 2017.

[8] Ali,S., Said, N., Yunus, N., Kader, S., Latif, D. and Munap, R.,Hackman and Oldham's Job Characteristics Model to JobSatisfaction, International Conference on Innovation, Management and Technology Research, Malaysia, Sepb22 – 23, 2013.

[9] Papachristou, G., Job-Recommendation-Engine [Online],Available: https://github.com/ggeop/Job-Recommendation-Engine.

★★★