



IJIRCCCE

e-ISSN: 2320-9801 | p-ISSN: 2320-9798



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

Volume 12, Issue 4, April 2024

ISSN INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA

Impact Factor: 8.379



9940 572 462



6381 907 438



ijircce@gmail.com



www.ijircce.com

Detecting Fraudulent Employment Listings with NLP-Enhanced Multilayer Perceptron

G Reshma, P Keerthi, P Sai Mounika, A Kalisha, S Rakesh

Assistant professor, Dept. of ECE, Gates Institute of Technology, Gooty, India

UG Student, Dept. of ECE, Gates Institute of Technology, Gooty, India

UG Student, Dept. of ECE, Gates Institute of Technology, Gooty, India

UG Student, Dept. of ECE, Gates Institute of Technology, Gooty, India

UG Student, Dept. of ECE, Gates Institute of Technology, Gooty, India

ABSTRACT: In modern time, the development in the field of industry and technology has opened a huge opportunity for new and diverse jobs for the job seekers. With the help of the advertisements of these job offers, job seekers find out their options depending on their time, qualification, experience, suitability etc. Recruitment process is now influenced by the power of internet and social media. Since the successful completion of a recruitment process is dependent on its advertisement, the impact of social media over this is tremendous. Social media and advertisements in electronic media have created newer and newer opportunity to share job details. Instead of this, rapid growth of opportunity to share job ads has increased the percentage of fraud job postings which causes harassment to the job seekers. So, people lack in showing interest to new job postings due to preserve security and consistency of their personal, academic, and professional information. Thus, the true motive of valid job postings through social and electronic media faces an extremely hard challenge to attain people's belief and reliability. Technologies are around us to make our life easy and developed but not to create unsecured environment for professional life. If jobs ads can be filtered properly predicting false job ads, this will be a great advancement for recruiting new employees. Therefore, this project proposed to use different data mining techniques and classification algorithm like K-nearest neighbour, decision tree, support vector machine, naive bayes classifier, random forest classifier, and multi-layer perceptron to predict a job Advertisement if it is real or fraudulent. We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. Deep neural network as a classifier, performs great for this classification task. We have used three dense layers for this deep neural network classifier. The trained classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job ad.

KEYWORDS: fake job prediction, machine learning, neural networks, deep neural networks.

I. INTRODUCTION

In modern time, the development in the field of industry and technology has opened a huge opportunity for new and diverse jobs for the job seekers. With the help of the advertisements of these job offers, job seekers find out their options depending on their time, qualification, experience, suitability etc. Recruitment process is now influenced by the power of internet and social media. Since the successful completion of a recruitment process is dependent on its advertisement, the impact of social media over this is tremendous[1]. Social media and advertisements in electronic media have created newer and newer opportunity to share job details. Instead of this, rapid growth of opportunity to share job posts has increased the percentage of fraud job postings which causes harassment to the job seekers. So, people lackin showing interest to new job postings due to preserve security and consistency of their personal, academic and professional information. Thus,the true motive of valid job postings through social and electronic media faces an extremely hard challenge to attain people's belief and reliability. Technologies are around us to make our life easy and developed but not to create unsecured environment for professional life. If jobs posts can be filtered properly predicting false job posts, this will be a great advancement for recruiting new employees.

Fake job posts create inconsistency for the job seeker to find their preferable jobs causing a huge waste of their time. An automated system to predict false job post opens a new window to face difficulties in the field of Human Resource Management[2].

1.1 Fake Job Posting Job Scam Online job advertisements which are fake and mostly willing to steal personal and professional information of job seekers instead of giving right jobs to them is known as job scam. Sometimes fraudulent people try to gather money illegally from job seekers. A recent survey by Action Fraud from UK has shown that more than 67% people are at great risk who look for jobs through online advertisements but unaware of fake job posts or job scam. In UK, almost 700000 job seekers complained to lose over \$500000 being a victim of job scam. The report showed almost 300% increase over the last two years in UK. Students, fresh graduates are being mostly targeted by the frauds as they usually try to get a secured job for which they are willing to pay extra money. Cybercrime avoidance or protection techniques fail to decrease this offence since frauds change their way of job scam very frequently[3].

1.2 Common types of Job Scam Fraudsters who want to gain other people's personal information like insurance details, bank details, income tax details, date of birth, national id create fake job advertisements. Advance fee scams occur when frauds ask for money showing reasons like admin charges, information security checking cost, management cost etc. Sometimes fraudsters act them-selves as employers and ask people about passport details, bank statements, driving license etc. as pre-employment check. Illegal money mulling scams occur when they convince students to pay money into their accounts and then transfer it back[4]. This 'cash in hand' technique causes to work cash in hand without paying any tax. Scammers usually create fake company websites, clone bank websites, clone official looking documents etc. to trap job seekers. Most of the job scammers try to trap people through email rather than face to face communication. They usually target social media like LinkedIn to prove themselves as recruitment agencies or headhunters. They usually try to represent their company profile or websites to the job seeker as realistic as possible. Whatever the type of job scam they use, the always target the job seeker to fall in their trap, collecting information and making benefit either earning money or any other things.

Employment scam is one of the serious issues in recent times addressed in the domain of Online Recruitment Frauds (ORF). In recent days, many companies prefer to post their vacancies online so that these can be accessed easily and timely by the job-seekers. However, this intention may be one type of scam by the fraud people because they offer employment to job-seekers in terms of taking money from them. Fraudulent job advertisements can be posted against a reputed company for violating their credibility. These fraudulent job post detection draws a good attention for obtaining an automated tool for identifying fake jobs and reporting them to people for avoiding application for such jobs. For this purpose, machine learning approach is applied which employs several classification algorithms for recognizing fake posts. In this case, a classification tool isolates fake job posts from a larger set of job advertisements and alerts the user. To address the problem of identifying scams on job posting, supervised learning algorithm as classification techniques are considered initially. A classifier maps input variable to target classes by considering training data. Classifiers addressed in the paper for identifying fake job posts from the others are described briefly. These classifiers-based prediction may be broadly categorized into -Single Classifier based Prediction and Ensemble Classifiers based Prediction

II. LITERATURE SURVEY

Habiba et. al [6] proposed to use different data mining techniques and classification algorithm like KNN, decision tree, support vector machine, naïve bayes classifier, random forest classifier, multilayer perceptron and deep neural network to predict a job post if it is real or fraudulent. We have experimented on Employment Scam Aegean Dataset (EMSCAD) containing 18000 samples. Deep neural network as a classifier, performs great for this classification task. We have used three dense layers for this deep neural network classifier. The trained classifier shows approximately 98% classification accuracy (DNN) to predict a fraudulent job post.

Amaar et. al [7] used six machine learning models to analyze whether these job ads are fraudulent or legitimate. Then, we compared all models with both BoW and TF-IDF features to analyze the classifier's overall performance. One of the challenges in this study is our used dataset. The ratio of real and fake job posts samples is unequal, which caused the model over-fitting on majority class data. To overcome this limitation, we used the adaptive synthetic sampling approach (ADASYN), which help to balance the ratio between target classes by generating the number of samples for minority class artificially. We performed two experiments, one with the balanced dataset and the other with the imbalanced data. Through experimental analysis, ETC achieved 99.9% accuracy by using ADASYN as over-sampling and TF-IDF as feature extraction. Further, this study also performs an in-depth comparative analysis of our proposed approach with state-of-the-art deep learning models and other re-sampling techniques.

Mehboob et. al [8] handles the recruitment fraud/scam detection problem. Several important features of organization, job description and type of compensation are proposed and an effective recruitment fraud detection model is constructed using extreme gradient boosting method. It develops an algorithm that extracts required features from job ads and is tested using three examples. The features are further considered for two-step feature selection strategy.

The findings show that features of the type of organization are most effective as a stand-alone model. The hybrid composition of selected 13 features demonstrated 97.94% accuracy and outperformed three state-of-the-art baselines. Moreover, the study finds that the most effective indicators are “salary_range,” “company_profile,” “organization_type,” “required_education” and “has_multiple_jobs.” The findings highlight the number of research implications and provide new insights for detecting online recruitment fraud

III. PROPOSED METHOD

3.1 EMSCAD Dataset The Employment Scam Aegean Dataset (EMSCAD) is a publicly available dataset containing 17,880 real-life job ads that aims at providing a clear picture of the Employment Scam problem to the research community and can act as a valuable testbed for scientists working on the field. To train the system, this project used EMSCAD dataset, where first row represents dataset column names and remaining rows contains dataset values such as Company profile, job description, salary etc. In dataset last column contains ‘fraudulent’ values as ‘f’ for Fake and ‘t’ for “True” jobs.

3.2 Data Preprocessing in Machine learning Data pre-processing is a process of preparing the raw data and making it suitable for a machine learning model. It is the first and crucial step while creating a machine learning model. When creating a machine learning project, it is not always a case that we come across the clean and formatted data.

3.3 TF-IDF Feature extraction TF-IDF which stands for Term Frequency –Inverse Document Frequency. It is one of the most important techniques used for information retrieval to represent how important a specific word or phrase is to a given document. Let’s take an example, we have a string or Bag of Words (BOW) and we have to extract information from it, then we can use this approach.

3.4 Multilayer perceptron (MLP) MLP is one of the most frequently used neural network architectures in MDSS, and it belongs to the class of supervised neural networks. The multilayer perceptron consists of a network of nodes (processing elements) arranged in layers. A typical MLP network consists of three or more layers of processing nodes: an input layer that receives external inputs, one or more hidden layers, and an output layer which produces the classification results.

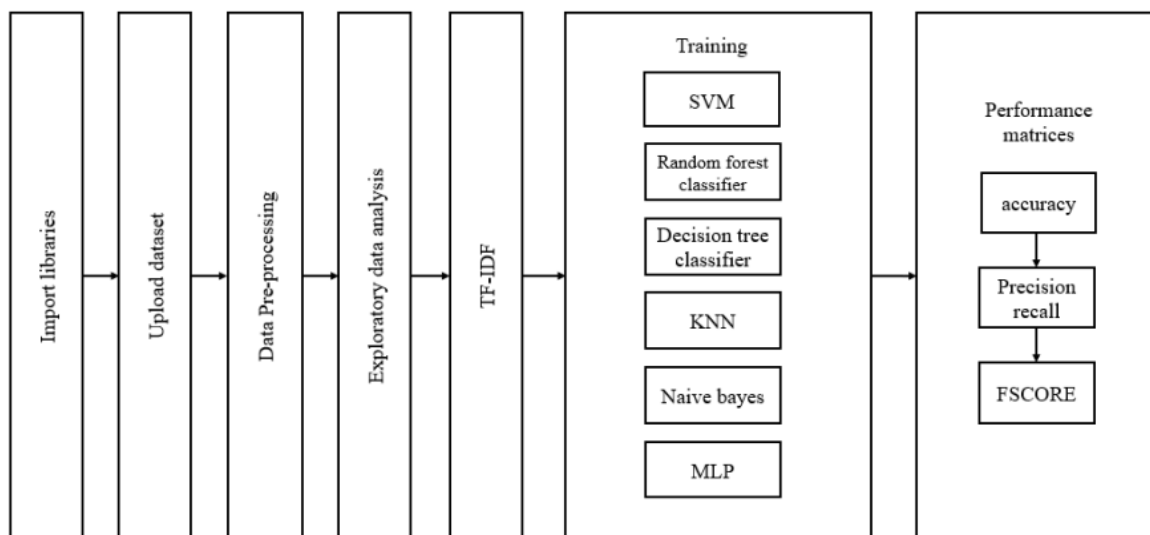


Fig. 1: Block diagram of proposed system.

IV. RESULT AND ANALYSIS

Proposed project is designed using python-GUI concepts. Tkinter library is used to prepare the GUI for easy use. Below is the proposed method GUI interface,



Fig.2 GUI for Proposed Model

On the above screen, to upload the dataset click on the ‘EMSCAD Dataset’ button and to get to the below screen, Select and upload the ‘EMSCAD dataset, in the above screen, and to load the dataset then click on the ‘Open’ button. To get the below screen,

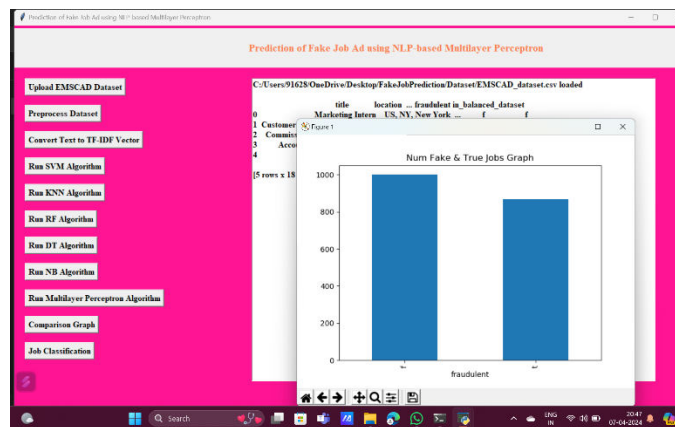


Fig. 3 Number of fake and true jobs graph from Dataset

Select and upload the ‘EMSCAD dataset, in the above screen, and to load the dataset then click on the ‘Open’ button. We can see the number of fake and true jobs graph, It indicates in above graph that approximately 1000 jobs are fake and 850 approx jobs are true or genuine. These jobs are loaded to model for performance analysis which will get first train and then will go for testing.

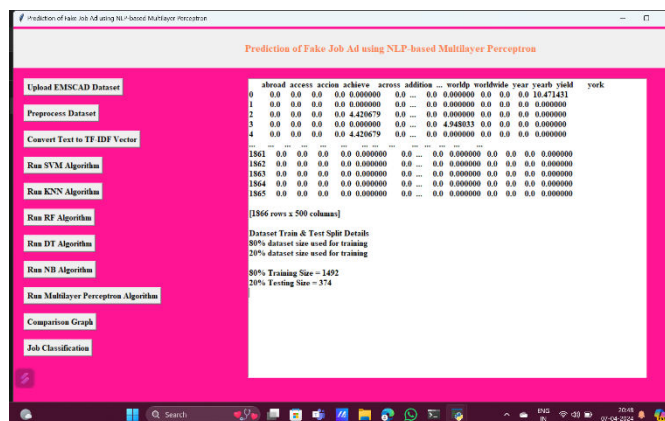


Fig. 4 Pre-process Dataset

Here displaying few records from dataset and in above dataset we can see some values are non numeric and machine learning will not accept non numeric values so we need to preprocess those values to assign integer id to each unique non numeric values.

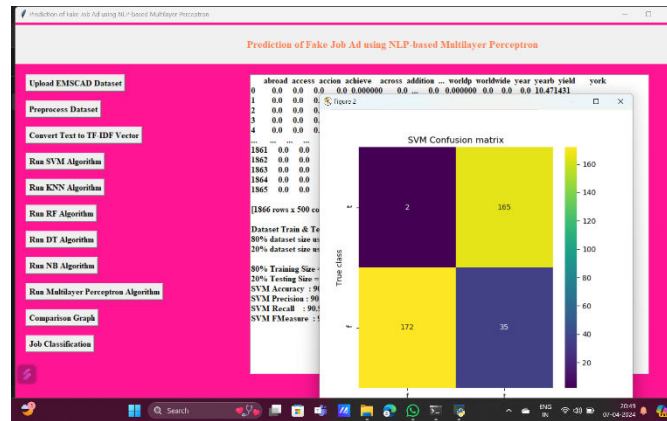


Fig.5 Confusion Matrix for Support Vector Machine (SVM)

Click on the run svm button and in the above screen we can get the confusion matrix for support vector machine, Similarly Click on the run KNN, RF, DT, NB, MLP(Multilayer perceptron).In the below screens we can get the confusion matrixes of each algorithm. It shows the difference between predicted and actual results. Also it show the percentages of Accuracy, Precision, Recall, FMeasure.

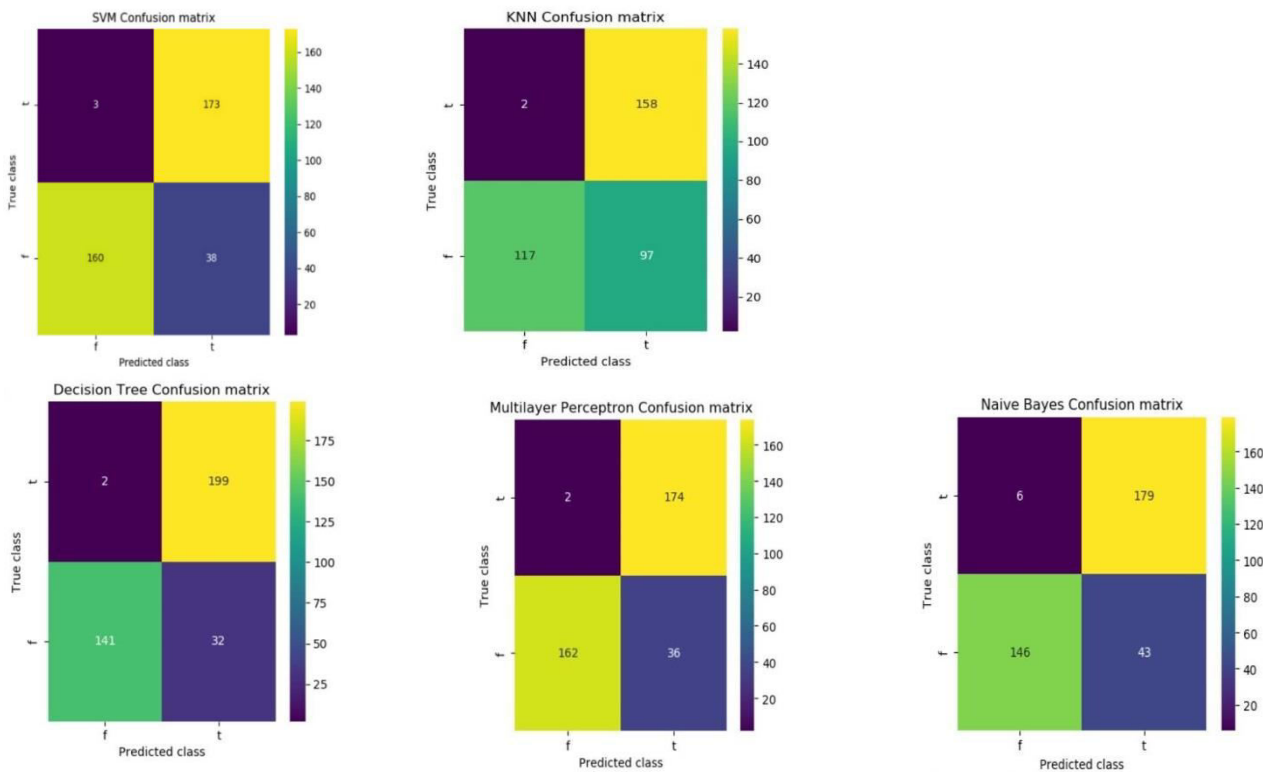


Fig.6 Confusion matrixes for SVM, KNN, RF, DT, NB, Multilayer Perceptron Algorithms

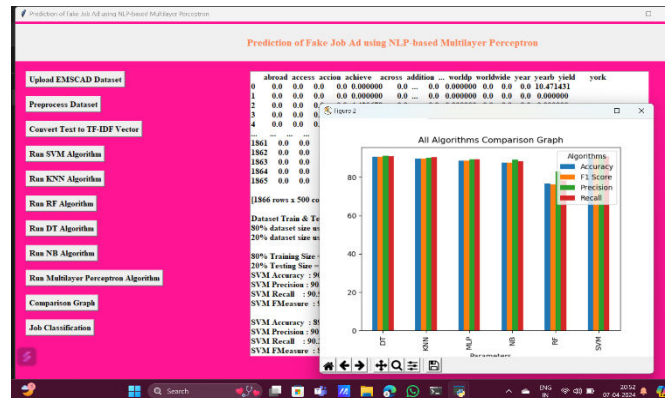


Fig. 7 Comparison Graph

In above screen we can the comparison graph for all algorithms and below screen we can see the job description. It is observed from above graph that MLP classifier performs better compare to the all others. So for further job prediction

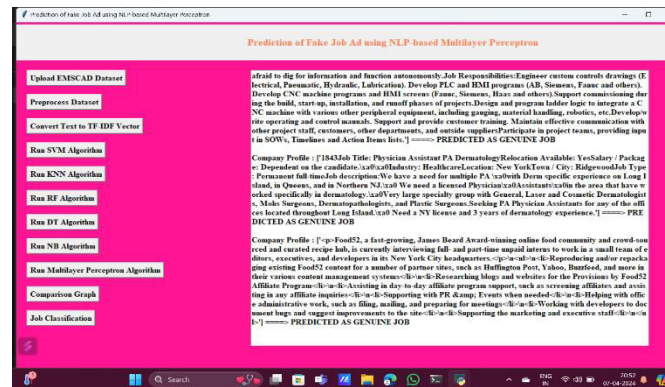


Fig. 8 Test Jobs Predicted as Genuine or Fake

MLP classifier is used for prediction of test jobs are fake or genuine. We added few job requirements in text file and applied MLP classifier for prediction

V.CONCLUSION

Job scam detection has become a great concern all over the world at present. In this project, we have analyzed the impacts of job scam which can be a very prosperous area in research filed creating a lot of challenges to detect fraudulent job posts. We have experimented with EMSCAD dataset which contains real life fake job posts. In this paper, we have experimented both machine learning algorithms SVM, KNN, Naive Bayes, Random Forest and a neural network concept called MLP. This work showed the evaluation of machine learning and MLP-based classifiers.

REFERENCES

[1]S. Vidros, C. Koliass, G. Kambourakis, and L. Akoglu, "Automatic Detection of Online Recruitment Frauds: Characteristics, Methods, and a Public Dataset", Future Internet 2017, 9, 6; doi:10.3390/fi9010006.

[2]B. Alghamdi, F. Alharby, "An Intelligent Model for Online Recruitment Fraud Detection", Journal of Information Security, 2019, Vol 10, pp. 155-176, <https://doi.org/10.4236/jis.2019.103009>.

[3]Tin Van Huynh1, Kiet Van Nguyen, Ngan Luu-Thuy Nguyen1, and Anh Gia-Tuan Nguyen, "Job Prediction: From Deep Neural Network Models to Applications", RIVF International Conference on Computing and Communication Technologies (RIVF), 2020.

[4]Jiawei Zhang, Bowen Dong, Philip S. Yu, "FAKEDETECTOR: Effective Fake News Detection with Deep Diffusive Neural Network", IEEE 36th International Conference on Data Engineering (ICDE), 2020.

[5]B. Alghamdi and F. Alharby, "An Intelligent Model for Online Recruitment Fraud Detection," J. Inf. Secur., vol. 10, no. 03, pp. 155–176, 2019, doi: 10.4236/jis.2019.103009



- [6]S. U. Habiba, M. K. Islam and F. Tasnim, "A Comparative Study on Fake Job Post Prediction Using Different Data Mining Techniques," 2021 2nd International Conference on Robotics, Electrical and Signal Processing Techniques (ICREST), 2021, pp. 543-546, doi: 10.1109/ICREST51555.2021.9331230.
- [7]Amaar, A., Aljedaani, W., Rustam, F. et al. Detection of Fake Job Postings by Utilizing Machine Learning and Natural Language Processing Approaches. *Neural Process Lett* 54, 2219–2247 (2022). <https://doi.org/10.1007/s11063-021-10727-z>
- [8]Mehboob, A., Malik, M.S.I. Smart Fraud Detection Framework for Job Recruitments. *Arab J Sci Eng* 46, 3067–3078 (2021). <https://doi.org/10.1007/s13369-020-04998-2>
- [9]D. Ranparia, S. Kumari and A. Sahani, "Fake Job Prediction using Sequential Network," 2020 IEEE 15th International Conference on Industrial and Information Systems (ICIIS), 2020, pp. 339-343, doi: 10.1109/ICIIS51140.2020.9342738.
- [10]Sudhakar, M., Kaliyamurthi, K.P. (2023). Efficient Prediction of Fake News Using Novel Ensemble Technique Based on Machine Learning Algorithm. In: Kaiser, M.S., Xie, J., Rathore, V.S. (eds) *Information and Communication Technology for Competitive Strategies (ICTCS 2021)*. LectureNotes in Networks and Systems, vol 401. Springer, Singapore. https://doi.org/10.1007/978-981-19-0098-3_1



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

 9940 572 462  6381 907 438  ijircce@gmail.com



www.ijircce.com

Scan to save the contact details