



iJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 8 Issue: IV Month of publication: April 2020

DOI:

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Finding Hidden Jobs via Twitter

Veena Patil¹, Abhijeet Rander², Madhuri Kalbhor³, Pragati Mane⁴, Prof. Suvarna Satkar⁵

^{1, 2, 3, 4}GHRCEM, Department of Computer Engineering

Abstract: In the job classification field, precise classification of jobs to profession categories is important for harmonizing job seekers with appropriate jobs. The automatic text job post classification system that utilizes machine learning is an example of such a job title classification system. Machine learning based job related classification techniques for text and related entities have been well researched and successfully adopted by many industries. The individuals who are looking for a job or any kind of career opportunities are mostly attracted digitally by online methods such as social media. This popular online method is known as Digital recruitment method. This system is a machine learning-based semi-supervised job title classification system uses SVM. The method contains the collection of classification and filtration techniques to overcome the challenges of designing a powerful and flexible classification system for a large taxonomy of job categories. The architecture of system consists of various filtration and classification algorithm. We are presenting experimental results on real world live data which is twitter feeds.

Keywords: Cloud, SVM, Machine Learning, Twitter, SVM.

I. INTRODUCTION

Data is next Oil, and Data Processing is next engine. The processing of data must match the rate of data generation. In the job classification domain, accurate classification of jobs to occupation categories is important. Here the project idea is pre-processing of raw job data from Social Networking Sites. To apply the NLP APIs for Data Text Classification. To process the live Job related Data from Twitter and Extract the JOB feeds from all over the world in real-time and to prepare a trend analysis for the various job types and categories.

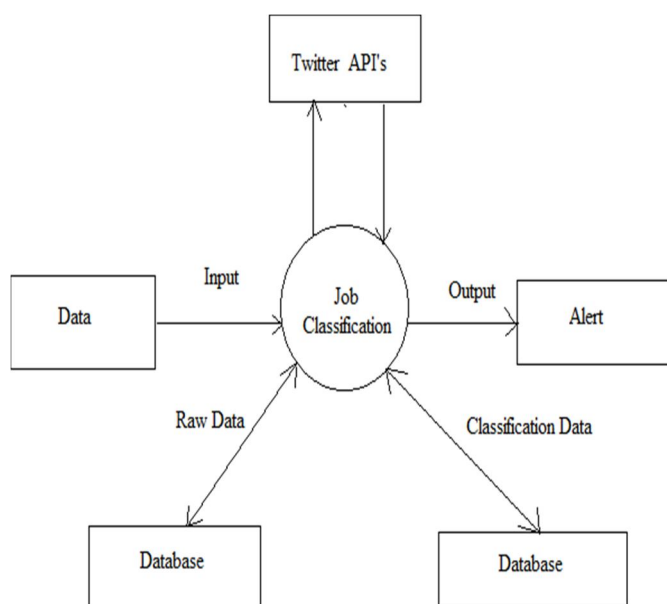


Fig.1. Flowchart of the system

II. SUPPORT VECTOR MACHINE(SVM)

Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms which are used for classification as well as regression. SVM has the ability to handle multiple continuous and categorical variables. The goal of SVM is to find a maximum marginal hyperplane (MMH) by dividing the datasets into different classes. To minimize the error, SVM generate the hyperplane in an iterative manner. Basically, SVM is a representation of different classes in a hyperplane in multidimensional space. Hyperplane is a space which is divided between a set of objects having different classes and margin is the gap between two lines. Support vectors are the data points that are closest to the hyperplane.

III. SYSTEM ARCHITECTURE

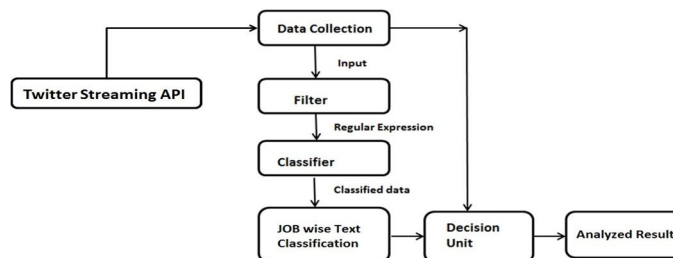


Fig. 2 System Architecture

To overcome the limitations of the existing system we are going to make a framework which will work using machine learning algorithms. This framework will allow user to find jobs of their interest. To find such jobs, all they need is to provide the key word which deflects user's interest and the framework will filter and fetch the data based on the keyword.

- 1) *Input:* Live Data Feed Filter related data (likewise, URL, Special Characters, Emotions, and Rewets). All other unnecessary data will be re- moved and the data is divided appropriate key value pair. Output: Filtered data.
- 2) *Input:* Filtered data Gather the filtered data from data store. Apply NLP using machine learning API's for individual data items from data store. Store the final summary into data store. Output: Analyzed and classified data.
- 3) *Job trend Summarization Input:* Analyzed and classified data

Technology wise Categorical Data is extracted or fetched for each job related data or for the technology data,. The data for every live feed is categorized. Persist the data into data store. Output: Trend Characterization for each job Category.

The important task in the data pre-processing stage is to select the useful and appropriate data. The initial step in this task is to remove the irrelevant, unnecessary and noisy data. The higher value of 'time index' is given to the most recently accessed data where as at the bottom with lowest value, the least recently accessed data were placed. Due to the time dependence characteristics of Web usage data, the analysis and classification of data becomes the critical step to obtain more accurate and appropriate analysis result.

IV. GOAL AND OBJECTIVE

- 1) *Goal:* Goal of the system is to efficiently and effectively process the Job data for extracting the pattern from it and to find hidden job's for job seekers.
- 2) *Objective:* The objective of the system is to preprocess the raw job data from twitter. Along with this, the system uses NLP APIs.

V. APPLICATION

The proposed system could be used for mining large data sets by applying filtration algorithms. It can be used to fetch live data from social media and can store in the database. It can be used for processing the data which is stored in the cloud. The proposed system is well suited to find hidden jobs as per the user's personal interest

VI. LITERATURE SURVEY

"Mining Twitter Data for a More Responsive Software Engineering Process" by Grant Williams and Ana's Mahmoud display that, the twitter has created an unprecedented opportunity for software developers to monitor the reaction or assumption of a huge population of end-users of their software. However, classification of useful tweets automatically is a minor task. The challenge in classification of the appropriate and useful data is the huge scale of the data available, unique format of the data, and high percentage of spammed data. To overcome these challenges, a procedure is introduces that influence Twitter as a main source of technical feedback. The main objective is to generate a more impactful, user friendly and active software engineering process. The analysis is conducted by fetching a dataset of tweets collected from the Twitter feeds. "A Survey on Sentiment Analysis on Twitter Data Using Different Techniques" by Bholane Savita Dattu, Prof. Deepali V. Gore explains about the Sentiment analysis which has many applications in various domains like political domain, sociology and real time event detection like earthquakes. Previously a research was carried out to determine and track public opinions. But with the advancement in research, today we can use it for interpreting the reasons of the sentiment change in public opinion, mining and summarizing products reviews. Here different algorithms are used to perform the above tasks like Naive Bayes classifier, Support Vector Machine (SVM) algorithm and so on.

VII. CONCLUSION

The unstructured data is analyzed, categorized and classified to job search entity based upon specific keywords, experience and area of interest. Twitter can feed tweets which are available free of cost to both the user and the profession. Job Trend Analysis can be done for various Job categories on Social Media Data feed. Various deep machine learning API techniques helps to improve the performance in job classification domain. Cloud technology is used for data processing as there is vast generalization occurs. Cloud stores data which can be accessible at any place anywhere. Real time data processing can be done. Machine learning can improve the classification. Various job categorizations can be taken places which is beneficial for job seekers.

REFERENCES

- [1] T. Joachims, Transductive inference using support vector machine (SVM) for classification of various texts, in proceedings of ICML 1999, pp. 200-209, 2016.
- [2] M. E. Ruize and P. Srinivasan, Categorization of hierarchical text using neural networks, Information Retrieval, vol. 5(1), pp. 87-118, 2016.
- [3] Guo, H. Wang, D. Bell, Y. Bi and K. Greer, Automatic text categorization using kNN algorithm, Soft Computing, vol. 10(5), pp. 423-430, 2015.
- [4] S. Fong, J. Liang, R. Wong, and M. Ghanavati, A novel feature selection by clustering coefficients of variations, in Proc. 9th Int. Conf. Digital Inf. Manag., Sep. 29, 2014, pp. 205213
- [5] A. Murdopo, Learning of distributed decision tree for mining big data streams, Masters of Science thesis, European Master Distrib. Comput., Jul. 2013
- [6] S. Fong, X. S. Yang, and S. Deb, Searching of Swarm for feature selection in classification, in Proc. 2nd Int. Conf. Big Data Sci. Eng., Dec. 2013, pp. 902909.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)