iJRASET
International Journal For Research in Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 11 **Issue:** V **Month of publication:** May 2023

DOI: https://doi.org/10.22214/ijraset.2023.52503

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)
ISSN: 2321-9653; IC Value: 45.98; SJ Impact Factor: 7.538
Volume 11 Issue V May 2023- Available at www.ijraset.com

Predicting Career Path using ML

Sujit Magar[1], Siddharth Parulekar[2], Mohammad Kaif Patel[3], Prasad Kandalkar[4], Prof. Dipali Kadam[5]
[1, 2, 3, 4, 5]*Dept. of Computer Engineering Pune Institute of Computer Technology Pune, India*

Abstract: ***When you choose a career, you are choosing an industry of employment where you are likely to gain expertise over time through your education and experience. You will probably work in this field for at least a few years. One of the most important decisions you can make is choosing and focusing on your career path. Finding the career path you want can be difficult, and you can be at a loss as to where to start and find the career path that's right for you. This includes endless Google searches that lead to various websites, giving you a list of different things you can do, such as talking to a career advisor or taking a professional aptitude test. You can enter anything that interests you. Also, curiosity-pumping paragraphs are much more effective and time-saving. Our website uses text analysis to extract keywords from user input and match them to potential career paths. This saves you time and effort compared to spending hours searching Google for job options based on your interests with unsatisfactory results. In this way, you can clear up the confusion that many people have about their careers and future prospects.***

Index Terms: SVM, Na¨ıve Bayes, Logistic Regression

I. INTRODUCTION

Currently many people are puzzled with what career path is best suited for their passion and skill set. Some of the very few methods of going about this is filling a questionnaire, going to a counsellor or time consuming google searches which can be tedious and can end up giving unsatisfactory or inaccurate results. The goal of this project is to use technologies like machine learning and natural language processing allowing a user to quickly and conviniently find the career meant for them.

Machine Learning is the field of study that gives computers the capability to learn without being explicitly programmed.

NLP(Natural Language Processing) is a a technology that allows computing machines to understand, interpret and pro-cess human languages.

Like text analysis, text mining is a technique that involves "automatically" extracting useful information from written sources to generate new, previously undiscovered knowledge by computer.Text mining and text analysis combine learning, statistics, and linguistics to find text patterns and trends in unstructured data.

The current task of displaying suitable career paths based on text entered by the user(consisting of user's interests, expertise and hobbies)makes use of the above mentioned technologies of Machine learning, NLP and text analysis. The informa-tion entered by user in the form of plain human language statements are first interpreted by the machine learning model using NLP. Then by text analysis and text mining useful insights are derived from the input text and based on the information and keywords extracted the matching career paths and associated information are lifted from the career database and displayed. Simply defined, it's a system that provides career recommendations based on information gathered from users.

II. MOTIVATION

When you choose a career, you are choosing an industry of employment where you are likely to gain expertise over time through your education and experience. You will probably work in this field for at least a few years. One of the most important decisions you can make is choosing and focusing on your career path. Finding the career path you want can be difficult, and you can be at a loss as to where to start and find the career path that's right for you. This includes endless Google searches leading to various websites, talking to career advisors, taking professional aptitude tests, focusing on choosing a career path, and much more. Finding the career path you want can be difficult and you can be at a loss as to where to start and find the career path that's right for you. It's an endless Google search that leads you to a variety of websites with lists of things you can do, like talking to a career advisor or taking a professional aptitude test

III. RELATED WORKS

S. Vignesh et al. [1] have created framework that s iden-tify where candidates lack certain skills and help candidates improve those skills. ANNs are well suited for classification problems. KNN claims 90 percent accuracy, higher than any other model such as SVM or RandomForest. They created their own dataset. The obtained results are passed to the ML model using the Flask API.

©IJRASET: All Rights are Reserved

Kartikey. Joshi et al. [2] have created system that uses a very complex system that is difficult to maintain and has privacy risks. To use this system, students must first register on the portal. They have created different test levels for different standards. The system given is only useful for 10th grade students. They developed this system using SVM and decision trees.

G. Ravi Kumar et al. [3] provides a brief overview of text mining models for improving the text mining process. Apply specific patterns and sequences to extract useful information by extracting irrelevant data for predictive analytics. Selective use of relevant domain techniques and tools facilitates facili-tation and effective text mining. Domain integration, different notions of granularity, multilingual text refinement, and NLP complexity are major problems and difficulties in the text mining process.

Yuanyuan chen et al. [4] takes a recruitment website as an example and builds a recruitment position information visualization platform based on text data mining, data analysis mining and other related issues.

Fantayc Aycle et al. [5] mentions various techniques of data mining, but among the techniques of text mining, information extraction technique is an effective technique for extracting valuable information from text. This technique focuses on extracting information from the actual text. The purpose of text mining is to discover knowledge from unstructured text. Related tasks in IE include finding the information you need and converting unstructured text into a structured database. Hand-made information extraction systems have existed for a long time, but recently, automatic construction of information extraction systems using M1 is also being done.

Rehmat Ullah et al. [6] have worked on paper. In this paper, the analysis is performed automatically when the data are provided in the required format, compared to other statistical approaches reported in the literature. Our thoughts can be employed to analyze various trends in other areas of life. Similarly, this idea can be extended to look at a country's entire labor market. The same can be used for teachers and course evaluations at universities. After receiving student feedback, you can perform qualitative analysis

Florentia et al. [7] uses suggestions from other approaches. As an alternative to job requirements analysis, you can eval-uate job descriptions, titles, or services provided by the com-pany. Through this white paper, we have provided a detailed overview of the competency requirements.

IV. REFERENCE ARCHITECTURE

The idea is to offer a solution for encrypting and hiding data as chunks on several nodes

A. *Architecture*

1) The first step is to collect and aggregate relevant data. Once the data is collected, it needs to be pre-processed to ensure it is in a suitable format for analysis.The pre-processed data is then stored in a suitable data man-agement system, such as a relational database, NoSQL database, or a data lake. This allows for efficient retrieval and processing of data during the recommendation pro-cess.

2) The relevant features are extracted from the pre-processed data. These features can be derived from user behavior, item attributes, or other contextual information. Feature extraction helps in representing the data in a format that can be used by machine learning algorithms. Rec-ommendation algorithms are at the core of a recom-mendation system. Various algorithms can be employed, such as collaborative filtering, content-based filtering, matrix factorization, deep learning models (e.g., neural networks), or hybrid approaches that combine multiple techniques. These algorithms analyze the data and gen-erate recommendations based on user preferences, item characteristics, and historical patterns. The recommen-dation algorithms are trained using historical data. This involves optimizing model parameters and hyperparam-eters to improve the accuracy and performance of the recommendations.

3) Once the models are trained, they can be deployed to generate real-time recommendations. User queries or requests trigger the recommendation system to process the available data, apply the trained models, and generate personalized recommendations. The system may also consider additional factors such as diversity, novelty, or business rules. The recommendations generated in the previous step may undergo post-processing and filtering to refine the results. Finally the product that we are working on can succesfully start running.

4) Data: The input data will be a paragraph or set if statements which contain user skills proficiencies and areas of interest

5) Data Preprocessing: The data will under go preprocess-ing which will involve the use of Natural Language processing tools like nltk for removing stop words and punctuation, lemmatization, POS tagging, etc.

6) Keyword Extraction: All the key words will be extracted from the input data through text mining and text analysis. The keywords extracted in this case will be the words that can most appropriately identify user interests and relevant user skills.

7) Filtering Component: It consists of the database and the Machine learning model.

8) Database: The mongo db database consists of all tye various career paths that exist in the IT and Computer science sector along with a short description of each and a list of requirements for each career.
9) Machine learning Model: The machine learning model uses the SVM algorithm to match and indentify appro-priate career paths from the database according to the key words extracted.
10) Recommendations: The matched career paths are then displayed along with a short description of each for the user to see.
11) Profile: This is the user profile which contains the set of career paths that were found suited for them.

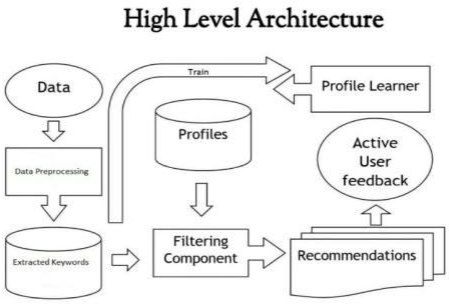


Fig. 1. Architecture Diagram

B. Security Evaluation

In the proposed system, security is provided by data dis-tribution and hiding. The system is reliable as it contains multiple copies of chunks. Apart from these, following are some criteria's for security analysis.

1) Availability - information requested should be available.
2) Usability - Information must to be accessible in a way that makes it easy to read.
3) Integrity - Integrity is the ability to ensure that a system and its data has not suffered unauthorized modification.
4) Authenticity - the owner of the information should have access to the author of the given information.
5) Confidentiality - keeping in place legal limitations on disclosure and access, as well as measures to safeguard confidential and proprietary information.
6) Ownership - the owner of the information should have access to it.

To summarize:

a) The user have to Login the page, here there has to be completed the registration part before Login. For regis-tration the user have to enter the name, email, password. After registration the user get the chance for the Login.
b) For the Authentication and Verification , first of all the email verification is done it will check every aspect of the email like Alpha-numeric charecters by the NodeMellor library in NodeJS.
c) Every entries in the MongoDb database is checked and verified by the authenticators library which will be done by the various NodeJS libraries.
d) After all the process is done by the Authentications part user will provide the input to the machine learning model through the framework called as streamlit. It will run the Model at the back-end and return the output at the front-end.

V. IMPLEMENTATION

We have used dataset Modified.csv which we have got from kaggle.com. The dataset gives information of Job Title, Job Requirements, Job description and Required Qualifications for various types of jobs. In data preprocessing we have done removing null values, removing unnecessary entries which are not related to IT sectors, Lowerization the text,removing stopwords and stemming.

For feature extraction we have used Term Frequency Inverse Document Frequency. We have used Support Vector Machine(SVM) algorithm for training and for recommondation. We have saved the SVM model with the help of pickle.

The implementation consists of various modules which work together to each other as a complete system. Below each module is explained in detail:

A. SVM

One of the most well-liked supervised learning algorithms, Support Vector Machine, or SVM, is used to solve Classifica-tion and Regression problems. However, it is largely employed in Machine Learning Classification issues.

The SVM algorithm's objective is to establish the best line or decision boundary that can divide n-dimensional space into classes, allowing us to quickly classify fresh data points in the future. A hyperplane is the name given to this optimal decision boundary.The SVM is giving accuracy of 88.52 when we take test size equal to 30 percent and using Stemming and TFIDF.

B. Streamlit Framework

Streamlit is an open-source Python framework that allows you to build interactive web applications for data science and machine learning tasks. It simplifies the process of creating and deploying web applications by providing an intuitive and declarative interface.

Create a new Python script for example, main.py, where we write your Streamlit application. Import the necessary libraries, including Streamlit. We use Streamlit's functions to define the structure and layout of the app. Streamlit provides various elements such as buttons, sliders, text inputs, and plots that you can use to create interactive components. This Frame work make our model more compatible to front-end and the back-end scripts like python.

C. Front-end

Design and create a web page or mobile app that in-cludes the login and signup components.Use HTML, CSS, and JavaScript to build the user interface. We used frameworks ReactJS for a more organized and interactive design. The main motive of the making the intreactive desingn of the site for more user convinient.

D. Backend-end

Set up a server-side technology such as NodeJS, ExpressJS for implementing an authentication system that handles user registration, login, and session management. We can Store user credentials securely in a database MongoDB that we can use with MongoDB Compass for Database.

VI. EXPERIMENTAL RESULTS

We have calculated accuracy for various algorithms with varying test size and methods.

NB=Na¨ıve Bayes

DT=Decision Tree SVM=Support Vector Machine KNN=K Nearest Neighbours (N=7) RF=Random Forest

A. Lemmatization and TFIDF

Test Size/Algorithm	NB	DT	SVM	KNN	RF
0.2	57.28	78.47	87.08	77.81	84.1
0.3	58.28	83	87.85	76.15	85.43
0.4	51.9	77.61	86.4	74.95	82.42

B. Lemmatization and Bag of Words

Test Size/Algorithm	NB	DT	SVM	KNN	RF
0.2	81.78	79.8	86.42	71.85	86.09
0.3	79.91	84.98	88	72.62	85.2
0.4	76.78	81.26	87.39	69	84.9

C. Stemming and TFIDF

Test Size/Algorithm	NB	DT	SVM	KNN	RF
0.2	54.3	81.12	88.41	80.13	83.77
0.3	56.95	78.14	88.52	75.71	84.32
0.4	51.9	78.6	87.06	75.12	82.42

D. Stemming and Bag of Words

Test Size/Algorithm	NB	DT	SVM	KNN	RF
0.2	82.78	84.1	86	76.15	82.78
0.3	80.35	83.88	87.19	75.27	84.32
0.4	78.1	79.93	86.06	72.96	83.08

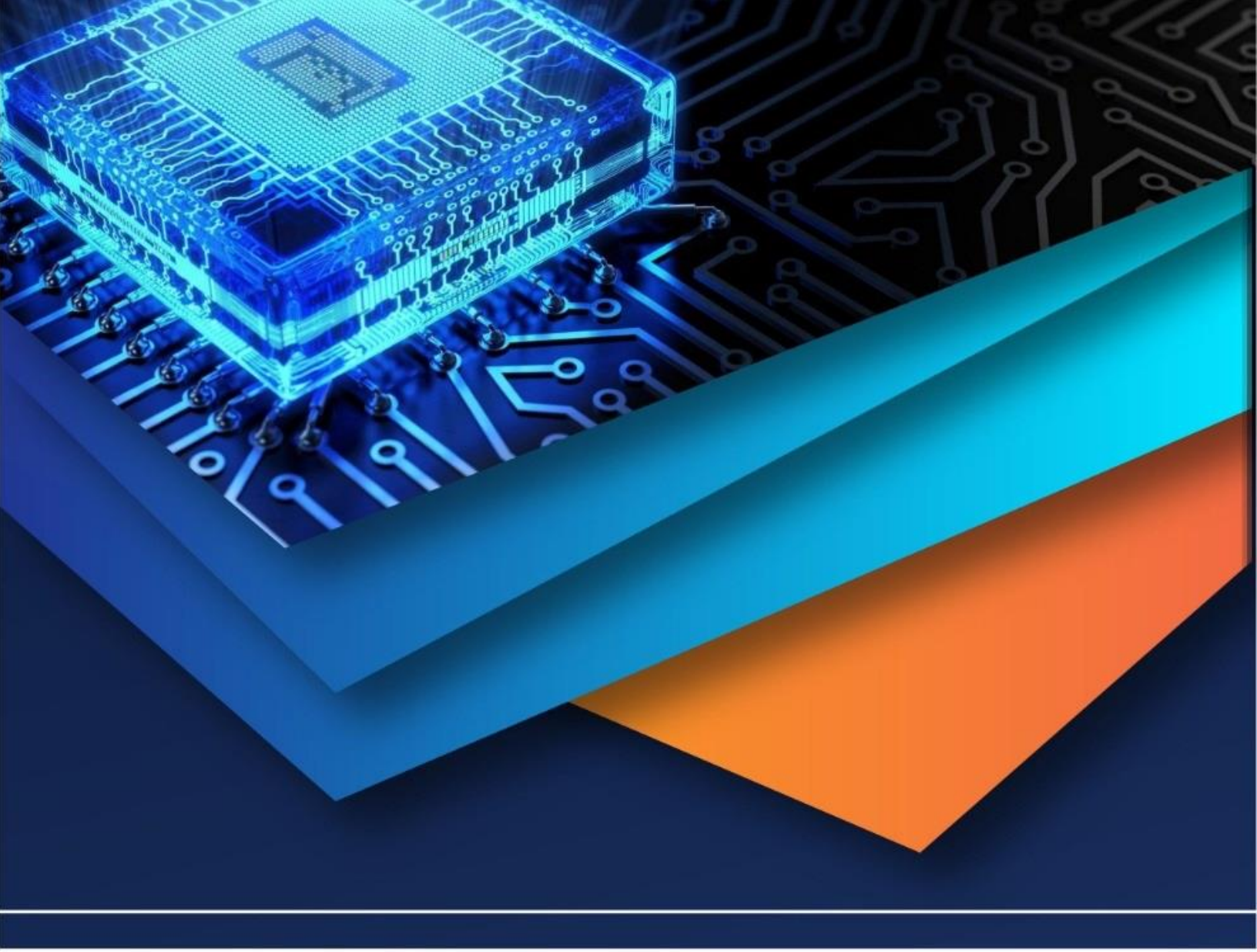
VII. CONCLUSION

The proposed system can use machine learning to identify an individual's career interests based on user-entered data. You can sort everyone based on their interests and match them with the right careers. Thus, users skip the steps of searching and searching different websites and changing search inputs for hours to get the desired answer. Chatbots can be used and language bots can be implemented in local languages so that comprehension is not hindered. You can also upload your resume. Our website informs the user what is missing and what needs to be added based on the user's chosen career path and takes the necessary steps to start the user's chosen career path. make it possible. [?].

REFERENCES

[1] S Vignesh, C Shivani Priyanka, H Shree Manju and K Mythili, "An Intelligent Career guidance system using Machine Learning," pp.2079-984, 19-20 March 2021.
[2] Kartikey Joshi, Amit Kumar Goel and Tapas Kumar, "Online Career Counsellor System based on Artificial Intelligence: An approach," 2021, 7th International Conference on Smart Structures and Systems (ICSSS), 2020 pp. 199-957, 23-24 July 2020.
[3] G Ravi Kumar, S Rahamat Basha, Surya Bhupal Rao, "A SUMMA- RIZATION ON TEXT MINING TECHNIQUES FOR INFORMATION EXTRACTING FROM APPLICATIONS AND ISSUES ,", J. Mech. Cont. Math. Sci., Special Issue, No.-5, January 2020 pp 324-332 , January 2020.
[4] Yuanyuan Chen, Ruijie Pan, "Research on Data Analysis and Visual-ization of Recruitment Positions Based on Text Mining," Advances in Multimedia 2022(23), pp. 904-202 ,July 2022.
[5] Aswani Pavithran, Z. A. Ahamad Ashraf, "Identification of Career Interest using Text Mining Techniques,", Mukt Shabd Journal, Volume IX, Issue V, MAY/2020, ISSN NO : 2347-3150, pp. 38-43, May 2022.
[6] Fantaye Ayele, "Text Mining Technique for Driving Potentially Valuable Information from Text,", In ISSN (Paper)2224-5758 ISSN (Online)2224-896X, , pp. 127-131, January 2020.
[7] Rehmat Ullah, Khawaja Muhammad Yahya, Samir Hussain Abdul-Jauwad, "Text Mining with Information Extraction in Career Coun-seling," In International Conference on Information and Intelligent Computing At: Hong KongVolume: 18, pp. 33-39, November 2020.
[8] Florentina, Marcu, "JOB REQUIREMENTS ANALYSIS WITH DATA MINING TECHNIQUES.," In Annals of 'Constantin Brancusi' Univer-sity of Targu-Jiu. Engineering Series . 2020, Issue 1, p8-13. 6p., pp. 457-462, July 2020.

©IJRASET: All Rights are Reserved | SJ Impact Factor 7.538 | ISRA Journal Impact Factor 7.894 |



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)