ADAPTIVE SKILL DEVELOPMENT SYSTEM

Harsh Patel1, Vidit Gala1, Shruthi Prabhu1, Pratham Shah1,
Nancy Nadar1

1*Computer Engineering, D.J. Sanghvi College of Engineering, Vile Parle, Mumbai, 400056, Maharashtra, India.

Contributing authors: harshpatel2209200@gmail.com; vidit250703@gmail.com; pshruthi2003@gmail.com; prathamshah1006@gmail.com; iiiauthor@gmail.com;

Abstract

In the rapidly evolving tech industry, engineering students often struggle to choose the right career path and acquire relevant skills. This project aims to provide data-driven career guidance by leveraging web scraping, sentiment analysis, and machine learning. Our system offers insights into emerging tech trends, personalized career roadmaps, skill and learning resource recommendations, and job/internship matching. By analyzing industry trends and aligning them with individual preferences, our solution empowers students to make informed career decisions and enhances their employability in the competitive job market.

Keywords: Skill Development, Machine Learning, Insights

1. Introduction

In today's rapidly evolving technological landscape, engineering students, particularly those in computer science and related fields, face challenges in selecting the right career path and acquiring the necessary skills. The dynamic nature of the job market, driven by the emerging technologies and industry trends, necessitates a data-driven approach to career guidance.

Traditional career counseling methods often lack real-time insights into job market trends and fail to provide personalized recommendations. Moreover, students struggle to identify the most relevant skills, learning resources, and job opportunities that align with their aspirations. To bridge this gap, we propose an AI-driven system that leverages web scraping, sentiment analysis, and machine learning to offer comprehensive career guidance.

The proposed system provides:

Insights on Tech Industry Trends – By scraping articles, blog posts, and reports related to technology, the system performs sentiment analysis to predict the potential rise or decline of various fields in computer science.
Personalized Career Roadmaps – Based on user preferences and market insights, a structured roadmap is generated to guide students in becoming Software Developers, Machine Learning Engineers, or other tech professionals.
Skill and Learning Resource Recommendations – The system not only identifies essential skills for a chosen career path but also suggests learning resources from platforms such as YouTube, Coursera, and other educational websites.
Job and Internship Matching – By scraping job postings from platforms like Naukri.com, the system provides job recommendations aligned with the student's acquired skills.
By integrating these components, the system aims to provide data-driven, real-time, and personalized career guidance to engineering students. This paper discusses the methodology, implementation, and evaluation of our system, highlighting its potential in automating career counseling, improving job readiness, and enhancing employability.

2. Literature Review

1] https://ieeexplore.ieee.org/document/10477957

In their study, "Personalized Career-Path Recommendation Model for Information Technology Students in Indonesia," the authors present a model designed to assist IT students in selecting suitable job specializations. The model addresses the challenge students face in making career decisions, which often rely on subjective perceptions from peers. By providing personalized recommendations, the model aims to guide students toward appropriate career paths in the IT sector.

21]https://www.irjmets.com/uploadedfiles/paper//issue_11_november_2024/63607/final/fin_irjmets1731301418.pdf

In their study, "AI-Driven Career Counseling Platform," Agrawal et al. introduce a system tailored for 11th and 12th-grade students, utilizing advanced AI algorithms to assess individual interests, academic performance, and current job market trends. The platform offers personalized career recommendations, suggesting suitable courses, skill development opportunities, and higher education pathways. It aims to empower students to make informed decisions aligned with evolving industry demands.

3]https://www.researchgate.net/publication/361830477_Evaluating_a_Natural_Language_Processing_Approach_to_Estimating_KSA_and_Interest_Job_Analysis_Ratings

In their study, "Evaluating a Natural Language Processing Approach to Estimating KSA and Interest Job Analysis Ratings," Putka et al. developed an NLP-based model to predict the importance of knowledge, skills, abilities, and other characteristics (KSAOs) for various occupations. Utilizing job descriptions and task statements as inputs, the model achieved cross-validated correlations with subject matter expert ratings of .74 for knowledge, .80 for skills, .75 for abilities, and .84 for interests. This approach streamlines job analysis by providing accurate, data-driven KSAO estimations, reducing reliance on traditional, resource-intensive methods.

4]https://www.researchgate.net/publication/315848045_A_Recommendation_System_for_Online_Courses

In "A Recommendation System for Online Courses," Estrela et al. develop a system that suggests online courses to users by analyzing their profiles and drawing parallels with other users. The system employs three recommendation techniques: Content-Based Filtering, Collaborative Filtering, and a Hybrid approach. By integrating these methods, the system aims to deliver precise and personalized course recommendations, enhancing user engagement and ensuring that learners receive content aligned with their interests.

5] https://ieeexplore.ieee.org/abstract/document/10533556

The study employs **Twitter scraping and sentiment analysis** to assess the correlation between public sentiment and stock price fluctuations. Using **natural language processing (NLP)** techniques, the model quantifies sentiment polarity and evaluates its influence on market trends. The findings highlight the significance of real-time sentiment analysis in **predicting financial market behavior** and emphasize the broader applicability of sentiment-driven forecasting in various domains.

6]
https://www.researchgate.net/publication/359655090_Recommender_Systems_in_E-learning

In "Recommender Systems in E-learning," Zhang et al. systematically review the primary recommendation techniques—**content-based**, **collaborative filtering-based**, and **knowledge-based**—employed to personalize learning experiences. The study highlights how these methods identify learner preferences to suggest appropriate courses and materials, addressing the challenge of navigating vast online educational resources. The authors also discuss emerging research directions, emphasizing the need for adaptive and context-aware recommender systems to enhance learner engagement and outcomes.

7] https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050923X00143/1-s2.0-S1877050923020793/main.pdf

In "Web Scraping using Natural Language Processing: Exploiting Unstructured Text for Data Extraction and Analysis," Pichiyan et al. explore the integration of web scraping techniques with NLP to transform unstructured web data into structured, actionable insights. The study provides an overview of web scraping methods, including rule-based parsing and XPath queries, and delves into applying NLP techniques for processing and analyzing the extracted textual data. This approach addresses challenges in handling unstructured data, enabling more efficient information retrieval and analysis.

8] https://www.sciencedirect.com/science/article/abs/pii/S0957417424009679

In "A Personalized Career Path Recommendation System for IT Students Using Machine Learning," the authors present a model that leverages machine learning algorithms to analyze student profiles and job market trends, providing tailored career path suggestions. The system aims to bridge the gap between academic training and industry requirements, enhancing employability by aligning student skills with market demands. This approach underscores the importance of data-driven decision-making in career counseling.

9] https://ijritcc.org/index.php/ijritcc/article/view/10561/8003

In "A Review of Resume Analysis and Job Description Matching Using Machine Learning," the authors provide a comprehensive review of machine learning (ML) and natural language processing (NLP) techniques for optimizing resume-job description matching in talent acquisition. The paper highlights the significance of ML algorithms, including neural networks, support vector machines (SVM), and deep learning architectures, for resume parsing and semantic analysis. It also discusses challenges such as the need for diverse datasets and explores future research directions. This study aligns with our project's goal of enhancing job matching by leveraging advanced ML and NLP techniques to improve the accuracy of skill-based recommendations for students.

10] https://link.springer.com/article/10.1007/s10869-022-09824-0

In "Evaluating a Natural Language Processing Approach to Estimating KSA and Interest Job Analysis Ratings," the authors propose a streamlined NLP-based method to estimate the importance of knowledge, skills, abilities, and other characteristics (KSAOs) for jobs using job description and task statement text as input. The study leverages data from the O*NET system and an independent organization, demonstrating strong correlations between machine-predicted KSAO ratings and subject matter expert (SME) ratings. The findings highlight the validity of machine-based predictions, supported by meaningful regression coefficients and conceptual relevance of predictor models. This research aligns with our project's objective of using NLP to analyze job descriptions and recommend skills, providing a foundation for automating skill extraction and matching processes to enhance career guidance for students.

3. Methods Used

a) Datasets

To build our career guidance system, we have created and utilized multiple datasets sourced through web scraping, APIs, and publicly available repositories. First, we developed a dataset by scraping articles, blog posts, and reports using an NLP-based scraper, which was then analyzed to generate insights on emerging technologies. These insights are structured into a dataset containing technology names, sentiment classifications (positive, negative, or neutral), and relevant percentages. For roadmap generation, we leveraged a Kaggle dataset that maps job roles to required skills and supplemented it with job postings scraped from various employment websites. Additionally, we used the YouTube API to compile a dataset of educational videos, including metadata such as views, comments, and timestamps, enabling personalized video recommendations. A separate dataset was created by scraping Coursera courses, capturing details such as course ratings, instructors, pricing, and links for course recommendations. Furthermore, job postings from Naukri.com were scraped to form a dataset containing job titles, salaries, locations, experience requirements, and required skills, aiding in targeted job and internship recommendations. These datasets collectively empower our system to provide real-time, data-driven career insights, structured learning paths, and job opportunities tailored to students' aspirations.

b) Preprocessing

The preprocessing phase of our project is a critical step that ensures the raw data collected from various sources is cleaned, standardized, and transformed into a format suitable for analysis and further processing. Given the diverse nature of the data sources—ranging from job postings and articles to resumes and course descriptions—the preprocessing pipeline is designed to handle multiple data types and formats, ensuring consistency and accuracy throughout the project.

1. Text Extraction and Cleaning
The first step involves extracting text from unstructured data sources such as PDF resumes, job postings, and articles. For resumes, we use the PyPDF2 library to extract text from PDF files, ensuring that all relevant information, including skills and experience, is captured. The extracted text is then cleaned to remove unnecessary whitespace, special characters, and non-alphanumeric symbols using regular expressions. This step ensures that the text is in a consistent format for further processing.

For job postings and articles, we employ web scraping techniques using libraries like BeautifulSoup and Selenium to extract relevant content from HTML pages. The extracted text is cleaned to remove HTML tags, advertisements, and other irrelevant content. Additionally, we use the newspaper3k library to parse articles and extract metadata such as titles, authors, and publication dates, which are stored for later analysis.

2. Skill Extraction and Standardization
Skill extraction is a crucial part of the preprocessing pipeline, as it directly impacts the accuracy of job matching and roadmap generation. We use a combination of keyword-based matching and natural language processing (NLP) techniques to extract skills from resumes and job descriptions. A predefined set of skill keywords (e.g., "Python," "Machine Learning," "React.js") is used to identify relevant skills in the text. Additionally, we employ spaCy's named entity recognition (NER) to identify job roles and technologies mentioned in the text.

To handle variations in skill terminology (e.g., "app development" vs. "mobile app development"), we implement fuzzy matching using the fuzzywuzzy library. This ensures that similar skills are recognized as matches, even if they are phrased differently. The extracted skills are then standardized to lowercase and stripped of any leading or trailing spaces to ensure consistency.

3. Data Normalization and Transformation
Once the skills and other relevant information are extracted, the data is normalized to ensure uniformity across different datasets. For example, job postings from Naukri.com and Coursera course descriptions are stored in separate CSV files, each with its own structure. We use the pandas library to merge and transform these datasets into a unified format, ensuring that all fields (e.g., job titles, skills, locations) are consistently named and formatted.

In cases where skills are stored as comma-separated strings, we use Python's ast.literal_eval function to convert them into sets, making it easier to perform set operations like intersection and union. This is particularly useful for calculating the percentage match between a user's resume and job requirements.

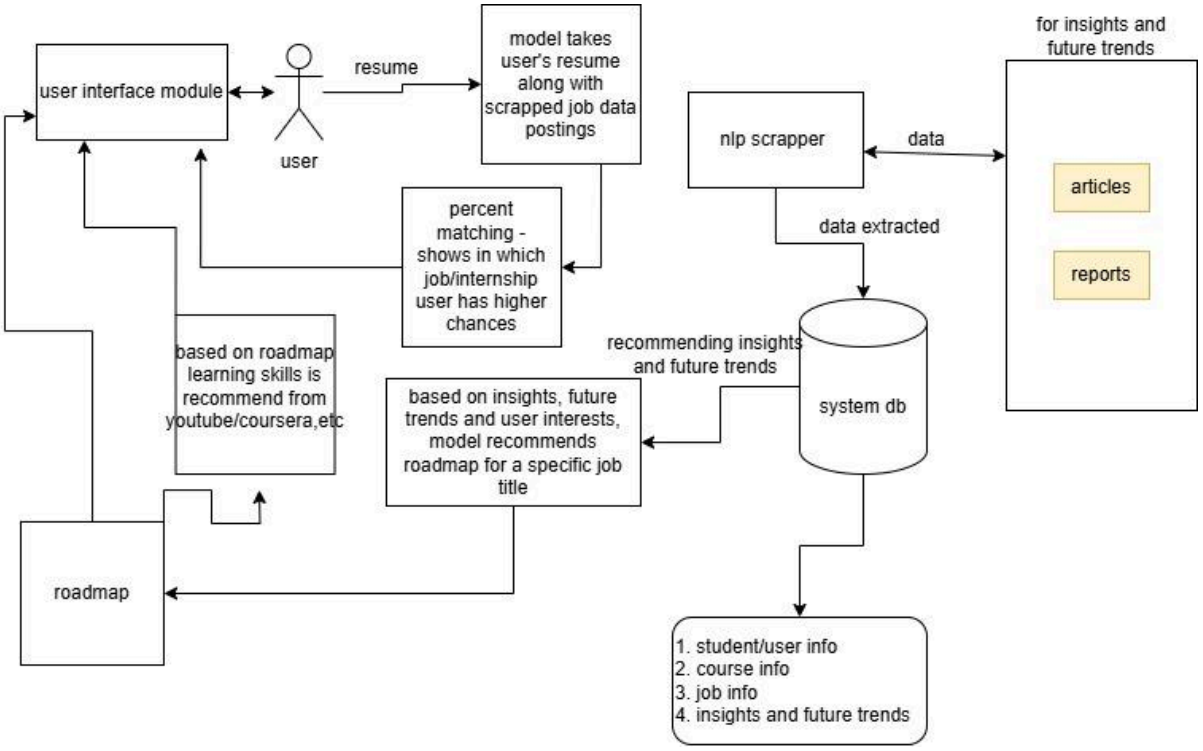
4. Sentiment Analysis and Trend Identification
For articles and blogs, we perform sentiment analysis to identify trends in the job market. Using the TextBlob library, we analyze the sentiment of sentences containing keywords related to job roles and technologies. Positive and negative keywords (e.g., "growth," "decline") are used to determine whether a particular skill or job role is on the rise or decline. This information is then used to generate insights for users, helping them make informed decisions about which skills to learn.

5. Data Storage and Integration

Finally, the preprocessed data is stored in structured formats such as CSV files and Excel sheets for easy access and further analysis. We ensure that all datasets are properly indexed and organized, allowing for seamless integration with other components of the project, such as the roadmap generator and job matching module.

By the end of the preprocessing phase, the raw data is transformed into a clean, structured, and standardized format, ready for analysis and visualization. This meticulous preprocessing ensures that the subsequent stages of the project—such as generating insights, creating roadmaps, and recommending jobs—are built on a solid foundation of high-quality data.

c) Proposed Architecture



The architecture diagram in Fig. 1 represents a Career Guidance System for engineering students, focusing on job trends, skill recommendations, and job matching.

Key Components:

1. User Interface Module – Users upload resumes and receive personalized recommendations.

2. Job Matching Model – Analyzes resumes and scraped job postings to show job/internship opportunities with higher selection chances.

3. NLP Scraper – Extracts insights from articles and reports on tech trends, stored in the system database.

4. Career Roadmap – Generates learning paths based on user preferences, future trends, and job role requirements.

5. Learning Resource Recommendations – Suggests YouTube, Coursera, and other platforms to help users acquire skills.

6. Final Job Suggestions – Once skills are acquired, users receive job and internship recommendations based on resume matching.

d) Models and Methods

This section outlines the methodologies employed in our system, integrating web scraping, Natural Language Processing (NLP)-based sentiment analysis, and similarity-based recommendation techniques to provide career guidance, skill enhancement recommendations, and job application readiness evaluation.

A. Web Scraping for Job Market Analysis and Course Recommendations

To extract structured data on job market trends and learning resources, web scraping techniques are utilized:

- **Job Listings Extraction:** Using **Selenium** and **BeautifulSoup**, job postings from *Naukri.com* are scraped, capturing job roles, required skills, and descriptions.

- **Course Data Collection:** Information from *Coursera* is retrieved, including course names, ratings, skill coverage, and provider details.
- **Implementation Details:**
    - **Selenium WebDriver** automates interaction with dynamically rendered web content.
    - **BeautifulSoup** parses the extracted HTML to retrieve structured information.
    - Data is stored in **CSV format** for subsequent processing.

B. Sentiment Analysis for Job Trend Detection

To quantify job demand trends, sentiment analysis is applied to textual job descriptions and industry articles.

1. **Preprocessing Pipeline:**
    - Tokenization, lemmatization, and stopword removal are performed using **NLTK**.
    - Special characters and numerical values are removed to refine textual features.
2. **Sentiment Classification:**
    - The **VADER (Valence Aware Dictionary and sEntiment Reasoner)** model computes sentiment polarity scores.
    - Job descriptions are categorized into **positive, neutral, or negative sentiment** based on compound scores.
3. **Trend Estimation:**
    - Roles with consistently high positive sentiment indicate **high-demand career paths**, whereas declining sentiment suggests **diminishing demand**.

C. Career and Skill Recommendation System

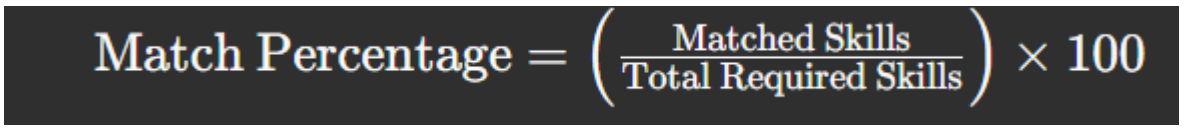
The system recommends job roles and upskilling courses using **information retrieval techniques** and **vector similarity models**.

1. **Job Role Recommendation:**
    - **TF-IDF (Term Frequency-Inverse Document Frequency)** is applied to vectorize job descriptions and user profiles.
    - **Cosine Similarity** measures the closeness between job role descriptions and a candidate's existing skills, retrieving the top-*k* relevant job matches.
2. **Skill-to-Course Mapping:**
    - **Keyword Extraction:** Relevant skills are extracted from job postings and matched against course syllabi.

- **Jaccard Similarity** is used to rank and recommend the most relevant courses for identified skill gaps.

D. Job Application Readiness Evaluation

To assess how well a candidate's resume aligns with a job role, a **resume-to-job similarity scoring model** is implemented.

1. **Resume Processing:**
   - **PyPDF2** extracts text-based skills and experience from a candidate's resume.
   - Extracted content is structured into a vectorized format for comparison.
2. **Skill Match Computation:**
   - The **set intersection** method determines overlapping skills between the candidate's profile and job requirements.
   - The **match percentage** is computed as follows:
3. \begin{equation}
   Match\ Percentage = \left(\frac{|\text{Matched Skills}|}{|\text{Total Required Skills}|} \right) \times 100
   \end{equation}
   - 

4. **Threshold-Based Classification:**
   - If **Match Percentage > 75%**, the job is classified as a **highly relevant opportunity**.
   - If **50% <= Match Percentage <= 75%**, the job is labeled as **moderately suitable**, prompting skill improvement recommendations.
   - If **Match Percentage < 50%**, the system suggests **upskilling courses** to bridge the knowledge gap.

E. Data Storage and Visualization

- Extracted data is stored in **CSV format** for structured analysis.
- **Visualization Tools:**
  - **Matplotlib** generates bar charts depicting job demand fluctuations.
  - **WordCloud** visualizes high-demand skills from job descriptions.

F. Key Contributions

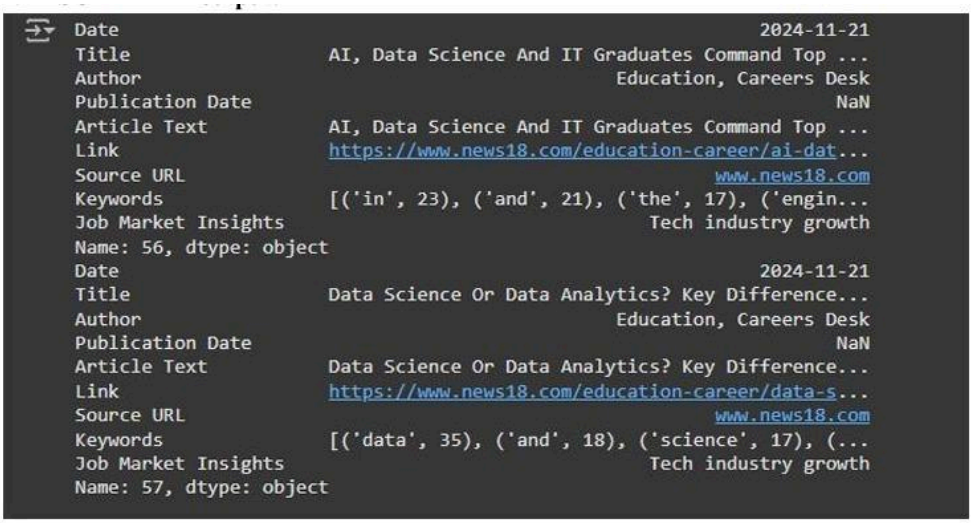
The proposed system provides:

1. **Automated job market insights** via web scraping and NLP-based sentiment analysis.

2. **Personalized career recommendations** using **TF-IDF and cosine similarity**.
3. **Skill-to-course mapping** using **Jaccard similarity** and keyword matching.
4. **Resume-based job matching** through **set intersection and semantic similarity**.
5. **Trend forecasting** leveraging **VADER sentiment scores**.

This data-driven methodology enhances employability by identifying **high-growth career opportunities, optimizing job applications, and guiding skill development** for engineering students and professionals.

4. Results

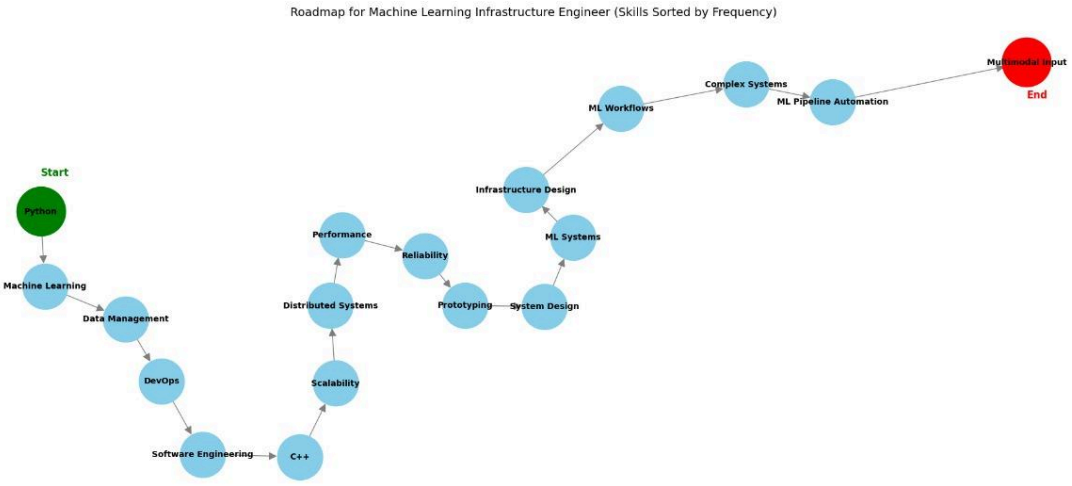


```
Date                                                    2024-11-21
Title                 AI, Data Science And IT Graduates Command Top ...
Author                                        Education, Careers Desk
Publication Date                                                NaN
Article Text          AI, Data Science And IT Graduates Command Top ...
Link                  https://www.news18.com/education-career/ai-dat...
Source URL                                              www.news18.com
Keywords              [('in', 23), ('and', 21), ('the', 17), ('engin...
Job Market Insights                             Tech industry growth
Name: 56, dtype: object
Date                                                    2024-11-21
Title                 Data Science Or Data Analytics? Key Difference...
Author                                        Education, Careers Desk
Publication Date                                                NaN
Article Text          Data Science Or Data Analytics? Key Difference...
Link                  https://www.news18.com/education-career/data-s...
Source URL                                              www.news18.com
Keywords              [('data', 35), ('and', 18), ('science', 17), (...
Job Market Insights                             Tech industry growth
Name: 57, dtype: object
```

The output shown in the image is the result of **news article scraper** built using **Google Custom Search API and the Newspaper3k library**. The script searches for articles related to the **future job market in computer science & related fields**, extracts article metadata (title, author, publication date, text, and keywords), and stores the results in a **CSV file**. The **keywords** are derived using **NLTK's frequency distribution**, and a simple rule is applied to classify articles under **job market insights**. The results are sorted by **publication date** and saved for further analysis.

```
Word: information, Sentiment Score: -1, Final Percentage: -68.00%
Word: technology, Sentiment Score: 1, Final Percentage: 0.00%
Word: computer science, Sentiment Score: 3, Final Percentage: 66.00%
Word: machine, Sentiment Score: 1, Final Percentage: 58.00%
Word: chemical, Sentiment Score: 1, Final Percentage: 0.00%
Word: artificial, Sentiment Score: 2, Final Percentage: 0.00%
Word: tech, Sentiment Score: 0, Final Percentage: 0.00%
Word: electronics, Sentiment Score: 1, Final Percentage: 0.00%
Word: india, Sentiment Score: 2, Final Percentage: 0.00%
Word: ece, Sentiment Score: 1, Final Percentage: 0.00%
Word: desk, Sentiment Score: 0, Final Percentage: 0.00%
Word: intelligence, Sentiment Score: 1, Final Percentage: 0.00%
Word: communication, Sentiment Score: 2, Final Percentage: 58.00%
Word: graduates, Sentiment Score: 2, Final Percentage: 0.00%
Word: data science, Sentiment Score: 3, Final Percentage: 66.00%
Word: global, Sentiment Score: 1, Final Percentage: 75.00%
Word: engineering, Sentiment Score: 2, Final Percentage: 68.00%
Word: cyber, Sentiment Score: 1, Final Percentage: 0.00%
Word: data, Sentiment Score: 1, Final Percentage: 0.00%
Word: mechanical, Sentiment Score: 0, Final Percentage: 0.00%
Word: it, Sentiment Score: 1, Final Percentage: 66.00%
Word: civil, Sentiment Score: 0, Final Percentage: 0.00%
Word: employment, Sentiment Score: 3, Final Percentage: 0.00%
Word: jobs, Sentiment Score: 0, Final Percentage: 0.00%
Word: news, Sentiment Score: 0, Final Percentage: 0.00%
Word: btech, Sentiment Score: 1, Final Percentage: 0.00%
Word: first, Sentiment Score: 0, Final Percentage: 0.00%
Word: learning, Sentiment Score: 1, Final Percentage: 0.00%
Word: b.tech, Sentiment Score: 1, Final Percentage: 0.00%
```

The code **analyzes job market trends** in an article by identifying job roles/technologies and their context. It calculates a **sentiment score** (positive, negative, or neutral) for each mentioned entity based on surrounding keywords related to growth or decline. **RAKE** is used to extract keywords from the context before and after the matched word, and these percentages are included in the final sentiment score. Finally, it generates a summary of the article and prints a sentiment analysis for each entity, including a final percentage reflecting the overall trend. This percentage is based on the weighted sum of positive and negative keyword mentions, with a cap at 100%.



The output is a **roadmap for a Machine Learning Infrastructure Engineer**, where skills are sorted by their frequency of occurrence across job roles. The roadmap starts with **Python**, followed by foundational skills like **Machine Learning, Data Management, and DevOps**. It then progresses through **Distributed Systems, Performance, and Scalability**, leading to more advanced topics like **ML Workflows, System Design, and ML Pipeline Automation**. The path ends at **Multimodal Input**, highlighting the culmination of expertise. The green node represents the **starting skill**, while the red node signifies the **final advanced skill** in the roadmap.

```
Rank: 10
Title: Java Full Course for free 🥥
Link: https://www.youtube.com/watch?v=xk4_1vDrzzo
Composite Score: 0.8888

Rank: 35
Title: This mat helped me learn Java so fast 🤯 #coding #java #programming #computer
Link: https://www.youtube.com/watch?v=-GOyQJXfxWg
Composite Score: 0.7789

Rank: 3
Title: Introduction to Java Language | Lecture 1 | Complete Placement Course
Link: https://www.youtube.com/watch?v=yRpLlJmRo2w
Composite Score: 0.7533

Rank: 1
Title: Java Full Course for Beginners
Link: https://www.youtube.com/watch?v=eIrMbAQSU34
Composite Score: 0.6436
```

The result displays the top 4 YouTube video recommendations based on a **composite score** calculated from normalized **view count, like count, and comment count** with assigned weights (0.5, 0.3, and 0.2, respectively). The

MinMaxScaler ensures all values are scaled between 0 and 1 before computing the final score. Despite the ranking order in the output, the composite score determines the actual relevance, leading to cases where a lower-ranked video (like Rank 10) has a higher score than Rank 1. This approach prioritizes videos with higher engagement metrics, making it an effective ranking strategy.

	Free_Paid	Link	Name	From	SkillsGain	rating	reviews	ExpType_Time
0	Paid	https://www.coursera.org/learn/python-for-appl...	Python for Data Science, AI & Development	IBM	[Skills you'll gain: Jupyter, Automation, Web ...	4.6 Rating, 4.6 out of 5 stars	40K reviews	[Beginner · Course · 1 - 3 Months]
1	Status: FreeFree	https://www.coursera.org/projects/showcase-bui...	Build a Website using an API with HTML, JavaSc...	Coursera Project Network	[Skills you'll gain: JSON, Restful API, Web De...	3.9 Rating, 3.9 out of 5 stars	24 reviews	[Intermediate · Guided Project · Less Than 2 H...
2	Paid	https://www.coursera.org/specializations/html-...	HTML, CSS, and Javascript for Web Developers	Johns Hopkins University	[Skills you'll gain: Cascading Style Sheets (C...	4.7 Rating, 4.7 out of 5 stars	17K reviews	[Beginner · Specialization · 1 - 3 Months]
3	Paid	https://www.coursera.org/projects/build-portfo...	Build Your Portfolio Website with HTML and CSS	Coursera Project Network	[Skills you'll gain: Hypertext Markup Language...	4.5 Rating, 4.5 out of 5 stars	2K reviews	[Beginner · Guided Project · Less Than 2 Hours]
4	Status: FreeFree	https://www.coursera.org/projects/backend-web-...	Backend Web Development with Go: Build an E-Ma...	Coursera Project Network	[Skills you'll gain: Secure Coding, Web Server...		None	[Intermediate · Guided Project · Less Than 2 H...

The extracted data represents a structured table of Coursera courses with details such as **Free/Paid status, course link, course name, provider, skills gained, rating, reviews, and expected time to complete**. The code scrapes this information using **BeautifulSoup** by extracting relevant HTML elements from the page. The extracted details are stored in a pandas **DataFrame**, and finally, the data is saved as a CSV file. The script efficiently processes multiple courses in a list format and ensures proper data structuring for easy analysis. Let me know if you need further modifications!

	Link	Name	CompName	loc	sal	exp	Skills
0	https://www.naukri.com/job-listings-software-d...	Software Developer (Training Focus) - MTech CS...	Nxtwave Disruptive Technologies	Hyderabad, Chennai, Bengaluru	Not disclosed	0-2 Yrs	[Google Sheets, Java, Javascript, HTML, Python...
1	https://www.naukri.com/job-listings-software-d...	Software Developer (C#)	Siemens	Bengaluru	Not disclosed	0-3 Yrs	[Power system, Concept design, Multithreading,...
2	https://www.naukri.com/job-listings-software-d...	Software Developer Engineer II	Amazon	Bengaluru	Not disclosed	0-7 Yrs	[automation framework, Object oriented design,...
3	https://www.naukri.com/job-listings-software-d...	Software Developer Intern / Trainee	Techavidus	Gandhinagar	Not disclosed	0-1 Yrs	[Asp.Net MVC, Java, Software development, Node...
4	https://www.naukri.com/job-listings-software-d...	Software Developer / Software Engineer	Intelligent Motion Technology Pvt Ltd	Pune	2-5.5 Lacs PA	0-3 Yrs	[JavaFx, C#, Core Java, VB.Net, C and C++, Hib...

The extracted data represents job postings for software developer roles from **Naukri.com** using **Selenium and BeautifulSoup**. The script scrapes job details such as **job title, company name, location, salary, experience required, and skills** from multiple pages. The data is structured into a **pandas DataFrame** and saved as a **CSV file** for further analysis. The extracted job postings include various companies like **Amazon, Siemens, and Nxtwave Disruptive Technologies**, covering different experience levels and skill sets. The image output shows a tabular

representation of the collected data, making it easy to analyze job trends and requirements.

CompName	Name	Match Percentage
Saffiretech	Software Developer	62.5
Mysuru Consulting Group	Software Developer	37.5
Dawn Verse Cloud Technologies	Software Developer (Cloud Solutions)	37.5
Design Centric	Software Developer	37.5
IntouchCX	Software Developer	37.5
Cisco	Software Engineer - C/C++ Developer	25
Cascades Data Solutions India	USA - H1B Sponsorship For Software Developers/Engineers	25
Brandmaker333	Software Developer	25
Dima Engineering	Software Engineer	25
IDZ Digital	Software Developer	12.5
In Solutions Global (ISG)	Software Developer	12.5
Catts Apac	Software Developer	0
PES Hr Services A Div Of Amal Infosystems	Urgent Openings For - Software Developer- Permanent	0
Infor	Software Engineer	0

The output in the image represents a **job matching system** that compares a candidate's resume skills with job postings. The **Match Percentage** column indicates how well a job aligns with the candidate's skills. Companies like **Saffiretech** have a **62.5% match**, meaning their job requirements closely align with the candidate's resume. Others, like **Infor**, have a **0% match**, suggesting no skill overlap. This approach helps job seekers prioritize applications where they have a higher chance of selection.

5. Conclusions

The study presents an innovative AI-driven career counseling platform designed to assist engineering students in navigating their career paths effectively. By leveraging advanced algorithms and data analysis, the system provides personalized recommendations that align with individual interests and market demands. The integration of web scraping and sentiment analysis allows for real-time insights into job market trends, enhancing the relevance of the guidance offered. Furthermore, the platform's ability to match students with suitable job opportunities based on their skills significantly improves their employability prospects. Overall, this approach not only automates career counseling but also equips students with the necessary tools to make informed decisions about their futures, ultimately fostering a more skilled and job-ready workforce. The findings underscore the potential of technology in

transforming career guidance and supporting students in achieving their professional goals.

6. References

1. https://ieeexplore.ieee.org/document/10477957

2.https://www.irjmets.com/uploadedfiles/paper//issue_11_november_2024/63607/final/fin_irjmets1731301418.pdf

3.https://www.researchgate.net/publication/361830477_Evaluating_a_Natural_Language_Processing_Approach_to_Estimating_KSA_and_Interest_Job_Analysis_Ratings

4.https://www.researchgate.net/publication/315848045_A_Recommendation_System_for_Online_Courses

5. https://ieeexplore.ieee.org/abstract/document/10533556

6.https://www.researchgate.net/publication/359655090_Recommender_Systems_in_E-learning

7.https://pdf.sciencedirectassets.com/280203/1-s2.0-S1877050923X00143/1-s2.0-S1877050923020793/main.pdf

8. https://www.sciencedirect.com/science/article/abs/pii/S0957417424009679

9. https://ijritcc.org/index.php/ijritcc/article/view/10561/8003

10. https://link.springer.com/article/10.1007/s10869-022-09824-0