Job Recommendation System

A report submitted in partial

fulfilment of the requirements for

the award of the degree of

Bachelor of Technology
in

COMPUTER SCIENCE AND ENGINEERING

Supervised by:
Dr. Jitendra K. Samariya
Assistant professor

Submitted by:
Ansh Agarwal (12111111)
Siddh Manek (12111114)
Nandan Wahi (12111085)
CSE – 6th sem

INDIAN INSTITUTE OF INFORMATION TECHNOLOGY, SONEPAT – 131201, HARYANA, INDIA

Table of Contents

Introduction

Problem Statement:

In today's dynamic job market, both job seekers and employers face challenges in finding the right match. To address this issue, we developed a Machine Learning (ML) based job recommendation system. The aim of this project was to leverage ML algorithms to provide personalized job recommendations to job seekers, thereby enhancing their job search experience and assisting employers in finding suitable candidates efficiently.

Objectives:

1. To develop a robust data collection pipeline to gather information on job postings, job seekers' preferences and historical interactions.
2. To implement various ML algorithms to analyze job seeker profiles and job descriptions.
3. To build a recommendation engine capable of generating personalized job recommendations for individual users.

Survey Report

S. no.	Group Member	Year	Author	Technique	Tool	Result	Conclusion /future scope	Link /DOI
1	Ansh Agarwal	2000	Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl	ITEM-BASED COLLABORATIVE FILTERING	Cosine Similarity and Weighted Sum	Better quality of prediction than user based (k nearest), the item neigbourhood is fairly static	Item based technique allows CF algorithms to scale to large datasets and produce high quality recommendations	https://www.researchgate.net/publication/2369002_Item-based_Collaborative_Filtering_Recommendation_Algorithms
2	Ansh Agarwal	2000	Robin Bruke	Knowledge-based recommendation	Matlab, Linear models	Knowledge based recommendation systems eliminate the needs of large knowledge component	Future study involves the application of hybrid based models	https://www.cs.odu.edu/~mukka/cs795sum09dm/Lecturenotes/Day6/burke-elis00.pdf
3	Nandan Wahi	2000	Badrul Sarwar, George Karypis, Joseph Konstan, and John Riedl	ITEM-BASED COLLABORATIVE FILTERING	Cosine Similarity and Weighted Sum	Better quality of prediction than user based (k nearest), the item neigbourhood is fairly static	Item based technique allows CF algorithms to scale to large datasets and produce high quality recommendations	https://www.researchgate.net/publication/2369002_Item-based_Collaborative_Filtering_Recommendation_Algorithms
4	Nandan Wahi	2002	Tong Zhang Vijay S. Iyengar	Linear Classifiers	Matlab, Linear models	Decision Tree performs poorly on datasets with more items, accuracy can't be directly compared	linear models attaining higher accuracy than memory-based collaborative filtering	https://www.jmlr.org/papers/volume2/zhang02a/zhang02a.pdf

5	Nandan Wahi	2004	R. Baraglia; F. Silvestri	WUM recommender system, called SUGGEST 3.0, that dynamically generates links to pages that have not yet been visited by a user and might be of his potential interest	SUGGEST 3.0	suggestions generated by SUGGEST 3.0 show a higher quality, that, in all the datasets, reaches a maximum for minfreq=0.2	evaluate SUGGEST 3.0 when running on real Web sites	https://ieeexplore.ieee.org/abstract/document/1410804/
6	Nandan Wahi	2004	Daniel Mulinge Ndolo	Multi-criteria recommendation technique is summarised in Sections 2.2 and 2.3 presents a panorama of ANN	Faiss index	Integrating graph-based methods with neural networks to enhance recommendation accuracy	Further research suggested to optimize recommendation systems and address user profile sparsity	https://ieeexplore.ieee.org/abstract/document/7064250
7	Nandan Wahi	2007	Daniel Lemire , Anna Maclachlan	Slope One algorithm		Accuracy comparable to that of the PEARSON scheme	Splitting ratings into dislike and like subsets can be an effective technique for improving accuracy.	https://www.researchgate.net/publication/1960789_Slope_One_Predictors_for_Online_Rating-Based_Collaborative_Filtering
8	Nandan Wahi	2009	Robert Bell, Chris Volinsky	MATRIX FACTORIZATION	Singular vector Dimension++,Scikit learn			https://datajobs.com/data-science-repo/Recommender-Systems-%5BNetflix%5D.pdf

9	Siddh Manek	2010	ivan.cantador, alejandro.bellogin	Content Based Recommendation in Social Tagging System	Delicious, Last.fm database, metrices	models focused on user profiles (tfu, bm25u) outperformed the models oriented to item profiles	Future work involves application of tag clustering techniques for user profiling	https://www.researchgate.net/publication/221140900_Content-based_recommendation_in_social_tagging_systems
10	Siddh Manek	2012	Shaha T. Al-Otaibi 1	Literature Review on Job Recommendation Systems	(Gated Recurrent Unit for Recommender Systems)		Future work includes training and evaluating the model using Word2Vec and k-means clustering algorithms and using the dataset scraped from job search websites	http://worldcomp-proceedings.com/proc/p2012/IKE7862.pdf

11	Ansh Agarwal	2012	Mourad Ykhlef	machine learning, natural language processing, and data analysis.			Developing methods to interpret the decisions made by the deep learning models and provide insights into the factors influencing employment possibilities for non-computer stream students.	http://worldcomp-proceedings.com/proc/p2012/IKE7862.pdf
12	Nandan Wahi	2013	Mamadou Diaby, Emmanuel Viennet, Tristan Launay	Content-based recommender system, supervised learning (linear SVM)	Not specified	Improved job recommendations by combining similarity measures with interactions data and employing a linear SVM.	Field importance estimation enhances recommendation accuracy, but addressing sparse Facebook user profiles remains a challenge.	https://dl.acm.org/doi/abs/10.1145/2492517.2500266
13	Siddh Manek	2013	Wenxing Hong, Siting Zheng, Huan Wang, Jianchao Shi	User clustering, recommendation approach selection	Not specified	Developed an online job recommender system (iHR) that groups users into clusters and employs different recommendation approaches for each cluster based on user characteristics.	Comparative study highlights differences in existing online job recommender systems (JRSs) and proposes personalized recommendation approaches based on user clustering.	https://citeseerx.ist.psu.edu/document?repid=rep1&type=pdf&doi=add8bb6e33412d3e197d678e2d697d8db16b61ab#page=7
14	Siddh Manek	2014	Yingya Zhang, Cheng Yang, Zhixiang Niu	Item-based Collaborative Filtering	Mahout	Improved recommendation precision and recall compared to user-based CF; incorporation of user and item	Further research suggested to optimize recommendation systems and address user profile sparsity	https://ieeexplore.ieee.org/document/7064250

						attributes enhances recommendation quality		
15	Nandan Wahi	2014	Gupta and Garg	Data Mining	Faiss index	Significant improvement in prediction accuracy, especially in the second phase of recommendations.	The paper emphasizes the importance of personalized recommendations in job recommender systems.	https://ieeexplore.ieee.org/abstract/document/6968361
16	Ansh Agarwal	2015	Huifeng Guo, Jinkai Yu	TR mode lis the concatenation of x and pos, i.e., xˆ = [x, pos]. A CTR prediction model is trained based on the concatenated feature vector.	DeepFM	the personalization at top-5 (L = 5), top-10 (L = 10) and top-20 (L = 20) of the recommendations by different frameworks.	CTR without position information in online inference. Compared to the baselines, PAL yeilds better results in a three-week online AB test. Extensive online experimental results verify the effectiveness of our proposed framework.	https://dl.acm.org/doi/abs/10.1145/3298689.3347033
17	Ansh Agarwal	2015	Balázs Hidasi, Alexandros Karatzoglou, Linas Baltrunas, Domonkos Tikk	Recurrent Neural Networks (RNN)	(Gated Recurrent Unit for Recommender Systems)	Effective sequential recommendation by capturing temporal dynamics in user behavior	RNN-based models offer a promising approach for personalized and sequential recommendation tasks	https://arxiv.org/abs/1511.06939
18	Ansh Agarwal	2016	Heng-Tze Cheng, Levent Koc	Memorization of feature interactions through a wide set of cross-product feature transformations		Wide & Deep learning---jointly trained wide linear models and deep neural networks	Online experiment results showed that the Wide & Deep model led to significant improvement on app acquisitions over wide-only and deep-only models	https://dl.acm.org/doi/pdf/10.1145/2988450.2988454

19	Ansh Agarwal	2016	Julian McAuley, Jure Leskovec	Graph Convolutional Networks (GCN)	GraphSAGE (Graph Sample and Aggregation)	Improved recommendation performance by modeling item relationships in a graph structure	GCN-based methods show potential for capturing complex item dependencies and enhancing recommendation quality	-
20	Siddh Manek	2016	Sonu K. Mishra	hybridization of these approaches, to overcome the shortcomings of each technique used alone and take advantage of their strengths, has been the subject of several research studies	Singular vector Dimension++,Scikit learn	hybridization of these approaches, to overcome the shortcomings of each technique used alone and take advantage of their strengths, has been the subject of several research studies	Collaboration with Industry Partners: Collaborating with businesses and industry partners to align job requirements with student skills and preferences, thereby increasing job placement success rates.	https://dl.acm.org/doi/abs/10.1145/2987538.2987546
21	Siddh Manek	2016	Manoj Reddy	educational data mining Machine learning techniques, including Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), Neural Network (NN), and K-nearest Neighbour (KNN)	Mahout		Item based technique allows CF algorithms to scale to large datasets and produce high quality recommendations	https://dl.acm.org/doi/abs/10.1145/2987538.2987546

22	Ansh Agarwal	2017	ZHIYUAN ZHANG1, YUN LIU 1, AND ZHENJIANG ZHANG	Field-Aware Matrix Factorization	Matab	13.6% performace impeovement than Matrix Factorization	FMF more acuurate than MF. Ranking evaluation metrices to optimize results can be done.	https://ieeexplore.ieee.org/stamp/stamp.jsp?arnumber=8241366
23	Ansh Agarwal	2017	Vinay Desai	content based filtering		suggestions generated by SUGGEST 3.0 show a higher quality, that, in all the datasets, reaches a maximum for minfreq=0.2	linear models attaining higher accuracy than memory-based collaborative filtering	https://d1wqtxts1xzle7.cloudfront.net/53591772/IRJET-V4I5343-libre.pdf?1497952927=&response-content-disposition=inline%3B+filename%3DImplementation_of_an_Automated_Job_Recom.pdf&Expires=1713120766&Signature=Vxd8E~byBV0uyafXnTQlOeS~rwLtrBN2VQC9JM~qPBRFCqEt4zK3hQCtPvTHXZn2kIZqKLdMgrfjT~GtiVcuIkBGMco4uNaRP6yasXPyBfGAI6xsvFRAxEHqOeXcSuL4Tt~X

								Nw~Wc7DyetiCqyJjHPQwHM-8ufgkJLxgaL0eMo5AwTrw0a1UZhhnIrsn1RU642yvamHtHjoBSunimiZvwQ-qC4zNxefTC1x77Y0wpbKo0Zel-HWPmjxUzEt1GKtmmaRVjNP9gyu~VrPr0ubiQgDQMaHmlZ~o5m51QIlErZLP0RmtTz9Ku1cgxQ4lHbgUXoBH3ClqKJUQum0ZvbvQPg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA

24	Ansh Agarwal	2017	Dheeraj Bahl	Review of Job Recommender Systems	DeepFM	The paper compares various recommendation models and filtering techniques used in job recommendation systems. It emphasizes the importance of accuracy in recommendation systems based on the adopted approaches and techniques. Hybrid recommendation techniques are highlighted for their effectiveness in improving job recommendation systems.	Feasibility and scalability of embedding-based system demonstrated for large-scale job to candidate matching.	https://d1wqtxts1xzle7.cloudfront.net/53591772/IRJET-V4I5343-libre.pdf?1497952927=&response-content-disposition=inline%3B+filename%3DImplementation_of_an_Automated_Job_Recom.pdf&Expires=1713120766&Signature=Vxd8E~byBV0uyafXnTQlOeS~rwLtrBN2VQC9JM~qPBRFCqEt4zK3hQCtPvTHXZn2kIZqKLdMgrfjT~GtiVcuIkBGMco4uNaRP6yasXPyBfGAI6xsvFRAxEHqOeXcSuL4Tt~XNw~Wc7DyetiCqyJjHPQwHM-8ufgkJLxgaL0eMo5AwTrw0a1UZhhnIrs

								n1RU642yvamHtHjoBSunimiZvwQ-qC4zNxefTC1x77Y0wpbKo0Zel-HWPmjxUzEt1GKtmmaRVjNP9gyu~VrPr0ubiQgDQMaHmlZ~o5m51QIlErZLP0RmtTz9Ku1cgxQ4lHbgUXoBH3ClqKJUQum0ZvbvQPg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
25	Siddh Manek	2017	Shreekumar Vibhandik,	Embedding-based Recommender System	Matlab, Linear models	Better quality of prediction than user based (k nearest), the item neigbourhood is fairly static	Splitting ratings into dislike and like subsets can be an effective technique for improving accuracy.	https://d1wqtxts1xzle7.cloudfront.net/53591772/IRJET-V4I5343-libre.pdf?1497952927=&response-content-disposition=inline%3B+filename%3DImplementation_of_an_Automated_Job_Recom.pdf&Expires=1

								713120766&Signature=Vxd8E~byBV0uyafXnTQlOeS~rwLtrBN2VQC9JM~qPBRFCqEt4zK3hQCtPvTHXZn2kIZqKLdMgrfjT~GtiVcuIkBGMco4uNaRP6yasXPyBfGAI6xsvFRAxEHqOeXcSuL4Tt~XNw~Wc7DyetiCqyJjHPQwHM-8ufgkJLxgaL0eMo5AwTrw0a1UZhhnIrsn1RU642yvamHtHjoBSunimiZvwQ-qC4zNxefTC1x77Y0wpbKo0Zel-HWPmjxUzEt1GKtmmaRVjNP9gyu~VrPr0ubiQgDQMaHmlZ~o5m51QIlErZLP0RmtTz9Ku1cgxQ4lHbgUXoBH3ClqKJUQum0ZvbvQPg__

								&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
26	Siddh Manek	2017	,Isra Fatma	Ensemble Learning Classifiers		Knowledge based recommendatio systems eliminate the needs of large knowledge component	Item based technique allows CF algorithms to scale to large datasets and produce high quality recommendations	https://d1wqtxts1xzle7.cloudfront.net/53591772/IRJET-V4I5343-libre.pdf?1497952927=&response-content-disposition=inline%3B+filename%3DImplementation_of_an_Automated_Job_Recom.pdf&Expires=1713120766&Signature=Vxd8E~byBV0uyafXnTQlOeS~rwLtrBN2VQC9JM~qPBRFCqEt4zK3hQCtPvTHXZn2kIZqKLdMgrfjT~GtiVcuIkBGMco4uNaRP6yasXPyBfGAI6xsvFRAxEHqOeXcSuL4Tt~XNw~Wc7DyetiCqyJ

								jHPQwHM-8ufgkJLxgaL0eMo5AwTrw0a1UZhhnIrsn1RU642yvamHtHjoBSunimiZvwQ-qC4zNxefTC1x77Y0wpbKo0Zel-HWPmjxUzEt1GKtmmaRVjNP9gyu~VrPr0ubiQgDQMaHmlZ~o5m51QIlErZLP0RmtTz9Ku1cgxQ4lHbgUXoBH3ClqKJUQum0ZvbvQPg__&Key-Pair-Id=APKAJLOHF5GGSLRBV4ZA
27	Nandan Wahi	2018	Priyanka P. Wadekar, Yedhukrishnan P. Pillai, Manodeep U. Roy, Prof. Neelam Phadnis	The Placement Predictor and Course Recommender System uses data mining techniques with the C4.5 algorithm and Naïve Bayes classifier for prediction and recommendation	WEKA	The accuracy of the Placement Predictor and Course Recommender System was evaluated using the C4.5 algorithm and Naïve Bayes Classifier. The C4.5 algorithm achieved an accuracy of 77.31%, while the Naïve Bayes	https://www.academia.edu/download/79115011/IRJET-V5I3929.pdf	The dataset used for testing in the development of the Placement Predictor and Course Recommender System was obtained

						classifier had an accuracy of 66.67%.		from the computer engineering batch of 2017-18.
28	Siddh Manek	2018	Ruining He, Wang-Cheng Kang, Julian McAuley	Neural Graph Collaborative Filtering	Graph convolutional networks (GCNs) and neural collaborative filtering (NCF)	Integrating graph-based methods with neural networks to enhance recommendation accuracy	Neural Graph Collaborative Filtering offers a unified framework for capturing both collaborative and content-based signals in recommendation systems	https://arxiv.org/abs/1905.08108
29	Nandan Wahi	2	Mohammed Hassan and Mohamed Hamada	Multi-criteria recommendation technique is summarised in Sections 2.2 and 2.3 presents a panorama of ANN	improving the prediction accuracy of the RSs (Adomavicius and Kwon,2012). Multi-criteria recommendation	That is an improvement in accuracy of approximately 26.7%and 27.3% in MAE and RMSE respectively is much better than the 19.0% and 3.6% of Zheng (2017) and Zhang et al. (2009) respectively	extracting a large volume of dataset and use it to train the networksto learn the relationships between criteria ratings information and the overall rating	https://www.inderscienceonline.com/doi/epdf/10.1504/IJSTDS.2019.097617
30	Nandan Wahi	2019	Chuan Qin et al.	Personalized Question Recommender System	LSTM-CRF, Skill-Graph	Developed DuerQuiz, a personalized question recommender system for job interviews, utilizing a Skill-Graph constructed from historical	The system showed remarkable performance in efficiency and effectiveness compared to traditional methods.	DOI: 10.1145/3292500.3330706

						recruitment data.		
31	Nandan Wahi	2020	D. Mhamdi, R. Moulouki, M. Y. El Ghoumari, M. Azzouazi, L. Moussaid	Text clustering (Word2Vec, K-means)	Word2Vec, K-means clustering	Job recommendations based on job profile clustering and job seeker behavior	Future work includes training and evaluating the model using Word2Vec and k-means clustering algorithms and using the dataset scraped from job search websites	https://doi.org/10.1016/j.procs.2020.07.102
32	Siddh Manek	2021	Khalid AL Fararni, Fouad Nafis	hybridization of these approaches, to overcome the shortcomings of each technique used alone and take advantage of their strengths, has been the subject of several research studies		hybridization of these approaches, to overcome the shortcomings of each technique used alone and take advantage of their strengths, has been the subject of several research studies	This architecture will be implemented, through advanced technologies, such as big data tools, machine learning technics, and the Internet of things.	https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=9321202
33	Siddh Manek	2021	Corné de Ruijt, Sandjai Bhulai	Review of Job Recommender Systems	Literature review	Not applicable	The paper provides a comprehensive review of job recommender systems literature from 2011 to 2021, emphasizing the temporal and reciprocal nature of job recommendations. It highlights the importance of considering algorithm fairness and discusses the	https://www.researchgate.net/publication/356601605_Job_Recommender_Systems_A_Review

							classification of hybrid recommender systems. Future research should focus on application-oriented approaches, ethical considerations, and generalizability across datasets.	
34	Nandan Wahi	2021	Jing Zhao, Jingya Wang, Madhav Sigdel, Bopeng Zhang, Phuong Hoang, Mengshu Liu, Mohammed Korayem	Embedding-based Recommender System	Faiss index	Significant improvement in click-through rate (CTR) and normalized discounted accumulated gain (nDCG) compared to baseline system.	Feasibility and scalability of embedding-based system demonstrated for large-scale job to candidate matching.	https://arxiv.org/abs/2107.00221
35	Nandan Wahi	2021	Giabelli et al.	Word embeddings, graph databases	FastText, graph database	Precision@3: rcaB - 0.823, cosB - 0.763; nDCG: rcaB - 0.985, cosB - 0.984	The paper demonstrates high precision in job recommendations based on user skills using a data-driven approach. Future work could involve evaluating skill-job fit using information from word embeddings models.	DOI: 10.1016/j.asoc.2020.107049

36	Nandan Wahi	2021	Sandjai Bhulai	Collaborative Filtering	Deep Learning Techniques, Data Extraction and Preprocessing: Techniques	Implemented a smart job seeking system using social media mining and NLP.	Further optimization and testing of the system are suggested for future work.	https://ieeexplore.ieee.org/document/7064250/authors#authors
37	Nandan Wahi	2021	Sandjai Bhulai	Collaborative Filtering	LightGCN and LR-GCCFob serve that nonlinear activation contributes little to the overall performance	The suggested models exhibit classification accuracies ranging from 75 to 94 percent. By conducting a data-driven investigation, the research aims to provide a model for learning analytics that can be applied in higher education institutions, influencing decision-making processes.	Personalized Recommendations: Incorporating machine learning algorithms to provide personalized job recommendations based on individual student profiles and preferences.	https://arxiv.org/pdf/2111.13576.pdf
38	Nandan Wahi	2021	Corné de Ruijt	The proposed technique involves the design and development of a comprehensive platform aimed at enhancing the effectiveness and efficiency of the campus placement process at colleges and universities.	Faiss index	employing algorithms such as Naïve Bayes, Random Forest, and Decision Trees, the study aims to provide insights into the factors influencing campus placement and assess the accuracy of the classifiers in predicting job/campus	Fine-Tuning Model Parameters: Experimenting with different deep learning architectures and hyperparameters to optimize prediction accuracy and generalization.	https://arxiv.org/pdf/2111.13576.pdf

				This platform digitizes all aspects of the campus placement pipeline and incorporates an Adaptive assessment tool powered by Artificial Intelligence (AI) technology.		placement outcomes.		
39	Siddh Manek	2022	Shiwen Wu	graph embedding techniques to model the relations between nodes, which can be further divided into factorization-based methods, distributed-representation-based methods, and neural-embedding-based methods	LightGCN and LR-GCCFob serve that nonlinear activation contributes little to the overall performance	a comprehensive review of the most recent works on GNN-based recommender systems. We proposed a classification scheme for organizing existing works.	he instance-level methods provide example-specific explanations by identifying important input features for its prediction	https://dl.acm.org/doi/full/10.1145/3535101
40	Siddh Manek	2022	S.A. Alhumam, H.A. Enaya, I. Ferjani, R. Ahmed	Job Recommendation System	Jaccard coefficient, Cosine distance, Collaborative Filtering	Matching job offers with resumes, recommending based on skills and experience	Proposal of a learning system for job seekers, promising results, future work includes efficient evaluation and diversity enhancement	https://www.mdpi.com/2073-431X/11/11/161

41	Siddh Manek	2022	Deepjyoti Roy and Mala Dutta	Systematic Review and Research Perspective	N/A	Analysis and synthesis of literature on recommender systems, development of a classification framework, evaluation of algorithmic categorization and performance metrics	Identifying gaps and challenges for future research in recommender systems	https://doi.org/10.1186/s40537-022-00592-5

42	Nandan Wahi	2022	Suleiman Ali Alsaif, Minyar Sassi Hidri, Imen Ferjani, Hassan Ahmed Eleraky, Adel Hidri	NLP-Based Bi-Directional Recommendation System	N/A	Improved prediction accuracy in bi-directional recommendation based on job offers and resumes	Future work may involve further optimization of the recommendation model and scalability testing	https://doi.org/10.3390/bdcc6040147
43	Nandan Wahi	2022	Nathalie Campos Macias et al.	Web Crawling, Natural Language Processing	Python	The prototype efficiently retrieves text summaries from websites but lacks effectiveness in delivering high-quality recommendations due to random arrangement of extracted sentences.	Natural language processing and web crawling are effective for automated information extraction, but improvements are needed for quality recommendations. Future work includes abstractive summarization, integration with other languages, and development of a graphical user interface.	https://doi.org/10.3390/a15080272
44	Siddh	2023	Kesen	session-based	RecoGy	ListCVAE	uaiSim stands as a	https://pro

	Manek		Zhao, Shuchang Liu	recommendation (SBR) considers the beginning and end of an interaction session. Our ultimate objective is to optimize the overall future reward of the user session, which aligns more with the notion of partial SBR	m [38], RecSim [14], RL4RS [44], and VirtualTaobao [41].	demonstrates the best performance in terms of maximum reward and diversity	comprehensive and sophisticated simulator that encompasses multiple task levels, establishing benchmarks and enabling thorough evaluations in the realm of recommendation systems	ceedings.neurips.cc/paper_files/paper/2023/file/8c7f8f98f9a8f5650922dd4545254f28-Paper-Datasets_and_Benchmarks.pdf
45	Nandan Wahi	2023	Shivam Vora; Alankrit Arya; Chaitanya Kumbhar; Harshal Dalvi	The proposed technique involves the design and development of a comprehensive platform aimed at enhancing the effectiveness and efficiency of the campus placement process at colleges and universities. This platform digitizes all aspects of the campus placement pipeline and incorporates an Adaptive assessment tool powered by Artificial Intelligence (AI) technology.	Digital Platform: Adaptive Assessment Tool: Powered by Artificial Intelligence technology. Data Analytics: s.	By leveraging AI technology, the platform provides accurate multidimensional assessments of students' strengths and weaknesses, including academic aspects and personality traits.	Personalized Recommendations: Incorporating machine learning algorithms to provide personalized job recommendations based on individual student profiles and preferences. Collaboration with Industry Partners: Collaborating with businesses and industry partners to align job requirements with student skills and preferences, thereby increasing job placement success rates.	https://pubs.aip.org/aip/acp/article-abstract/2916/1/020014/2926264/An-AI-based-adaptive-assessment-system-for?redirectedFrom=fulltext

46	Nandan Wahi	2023	Biku Abraham & P. S. Ambili	Deep Learning	Deep Learning Techniques, Data Extraction and Preprocessing: Techniques	By employing deep learning techniques and extracting data from a dataset with 19 attributes, the study aims to enhance methods for non-computer stream students to secure full-time employment in software-related jobs through in-campus selection.	Fine-Tuning Model Parameters: Experimenting with different deep learning architectures and hyperparameters to optimize prediction accuracy and generalization. Developing methods to interpret the decisions made by the deep learning models and provide insights into the factors influencing employment possibilities for non-computer stream students.	https://link.springer.com/chapter/10.1007/978-981-19-8493-8_60
47	Nandan Wahi	2023	Sachin Bhoite; C. H. Patil; Surabhi Thatte; Vikas J. Magar; Poonam Nikam	Ensemble Learning Classifiers	Ensemble Learning: Machine Learning Algorithms: Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest,	By employing ensemble learning and various machine learning algorithms, such as Logistic Regression, Support Vector Machine, K-Nearest Neighbor, Decision Tree, Random Forest, and AdaBoost classifier, the system aims to provide accurate predictions early in students' education, thereby	Fine-Tuning Ensemble Models: Experimenting with different ensemble learning techniques and hyperparameter tuning to further improve prediction accuracy. Incorporating External Data Sources: Integrating additional data sources, such as industry trends and job market data, to enhance	https://ieeexplore.ieee.org/abstract/document/10053523

					and AdaBoost classifier are considered	positively impacting the institute's training and placement activities.	the predictive capabilities of the system. Real-time Monitoring and Feedback: Developing mechanisms for real-time monitoring of placement predictions and providing feedback to continuously improve the accuracy of the system.	
48	Ansh Agarwal	2023	R.K. Kavitha; C. Rajan Krupa; Jolene Isabella Menezes	educational data mining Machine learning techniques, including Random Forest (RF), Support Vector Machines (SVM), Naive Bayes (NB), Neural Network (NN), and K-nearest Neighbour (KNN)		The suggested models exhibit classification accuracies ranging from 75 to 94 percent. By conducting a data-driven investigation, the research aims to provide a model for learning analytics that can be applied in higher education institutions, influencing decision-making processes.	Fine-Tuning Model Parameters: Experimenting with different hyperparameters and feature engineering techniques to optimize prediction accuracy and generalization. Longitudinal Analysis: Conducting longitudinal studies to assess the long-term effectiveness of the predictive model and its impact on student placement outcomes.	https://ieeexplore.ieee.org/abstract/document/10200741

49	Ansh Agarwal	2023	Daniel Mulinge Ndolo	Literature Review on Job Recommendation Systems	Comparative review	The paper compares various recommendation models and filtering techniques used in job recommendation systems. It emphasizes the importance of accuracy in recommendation systems based on the adopted approaches and techniques. Hybrid recommendation techniques are highlighted for their effectiveness in improving job recommendation systems.		https://www.researchgate.net/publication/369973770_Job_Recommendation_Systems_A_Literature_Review
50	Ansh Agarwal	2023	Amber Nigam, Aakash Roy, Arpan Saxena, Hartaran Singh	Machine Learning, Bi-LSTM with Attention, Blended Approach	Literature review	Achieved best click-through rate through a blended approach of machine-learned recommendations and other sub-recommendations.	Proposed methodology leverages the progression of job selection by candidates and aims to make job recommendations serendipitous.	https://arxiv.org/ftp/arxiv/papers/1905/1905.13136.pdf
51	Ansh Agarwal	2023	Ahmed Elsafty, Martin Riedl, Chris Biemann	Dense vector representations for job recommendation	Elasticsearch, Word2Vec, Doc2VecC	Click-through rate (CTR) increased by 8.0% using title-weighted Doc2VecC model for job recommendations.	The study demonstrates the effectiveness of dense vector representations for job similarity ranking, with potential for further improvement using ontology and keyword integration.	https://aclanthology.org/N18-3027/

52	Nandan Wahi	2023	Mantas Lukauskas et al.	Natural Language Processing (NLP) and Clustering	BERT, UMAP, HDBSCAN	Successful extraction of job profiles and skills from over 500,000 job postings using NLP and clustering techniques.	The automated job advertisement analysis algorithm showed promising results, suggesting practical application in real-world scenarios.	https://www.mdpi.com/2076-3417/13/10/6119
53	Ansh Agarwal	2024	Byagar, Sarita; Patil, Ranjit; Pawar, Janardan	three machine learning algorithms are utilized: Naïve Bayes, Random Forest, and Decision Trees. T		employing algorithms such as Naïve Bayes, Random Forest, and Decision Trees, the study aims to provide insights into the factors influencing campus placement and assess the accuracy of the classifiers in predicting job/campus placement outcomes.	Feature Engineering: Exploring the inclusion of additional features or parameters to improve the accuracy of campus placement predictions. Ensemble Methods: Investigating the use of ensemble methods to combine the predictions of multiple machine learning algorithms for enhanced accuracy.	https://openurl.ebsco.com/EPDB%3Agcd%3A3%3A10989699/detailv2?sid=ebsco%3Aplink%3Ascholar&id=ebsco%3Agcd%3A175815696&crl=c
54	Ansh Agarwal	2024	Ala Mughaid et al.	Recommender System	NLP, Geo-location	Implemented a smart job seeking system using social media mining and NLP.	Further optimization and testing of the system are suggested for future work.	https://ieeexplore.ieee.org/abstract/document/8931854

55	Nandan Wahi	2024	Yao Lu, Sandy El Helou, Denis Gillet	Hybrid Recommender System	N/A	Outperformed PS and CF in recommendation precision and user coverage	Future work includes user studies and evaluations based on online data for further refinement	https://dl.acm.org/doi/abs/10.1145/2487788.2488092

Methodology, Tools and Techniques

Previous Methodology-
The methodology involves several steps: preprocessing job descriptions and resumes, extracting relevant skills using natural language processing techniques, and using similarity measures to match candidates with job postings. The system utilizes machine learning algorithms such as TF-IDF vectorization, nearest neighbors search, and rule-based matching to achieve accurate recommendations.

Proposed Methodology and tools used-

- Batch Normalisation and Hyperparameter tuning
- Text processing including removing stopwords, punctuation, special characters, and performing lemmatization or stemming to reduce words to their base forms.
- TF-IDF (Term Frequency-Inverse Document Frequency) vectorization to convert text data into numerical vectors.
- Rule-based matching techniques, such as regular expressions or pattern matching, are used to identify specific patterns or entities in text data.
- Content-Based Filtering: Recommending jobs based on similarity measures calculated between job descriptions and user profiles, using techniques like cosine similarity
- Document Parsing
- Machine learning models
- Resume Analysis and Feedback: Provide feedback to job seekers on their resumes, highlighting areas for improvement and suggesting relevant skills or experiences to enhance their profiles.
- Customizable Recommendation Alerts: Allow users to set up personalized recommendation alerts based on specific criteria such as job title, location, salary range, or skill requirements.
- Python is the primary coding platform for this project, libraries such as, scikit-learn, pandas, NLTK, tensorflow, pytorch, PyPDF2 for deep learning, NLP tasks, data manipulation, and PDF parsing.
- Skills extraction using spacy library of NLP in python.

Machine learning algorithms-
1. TF-IDF Vectorization:
   - TF-IDF stands for Term Frequency-Inverse Document Frequency. It is a numerical statistic that reflects how important a word is to a document in a collection or corpus.
   - The `TfidfVectorizer` from Scikit-learn is used to convert a collection of raw documents (in this case, resume skills) into a matrix of TF-IDF features.

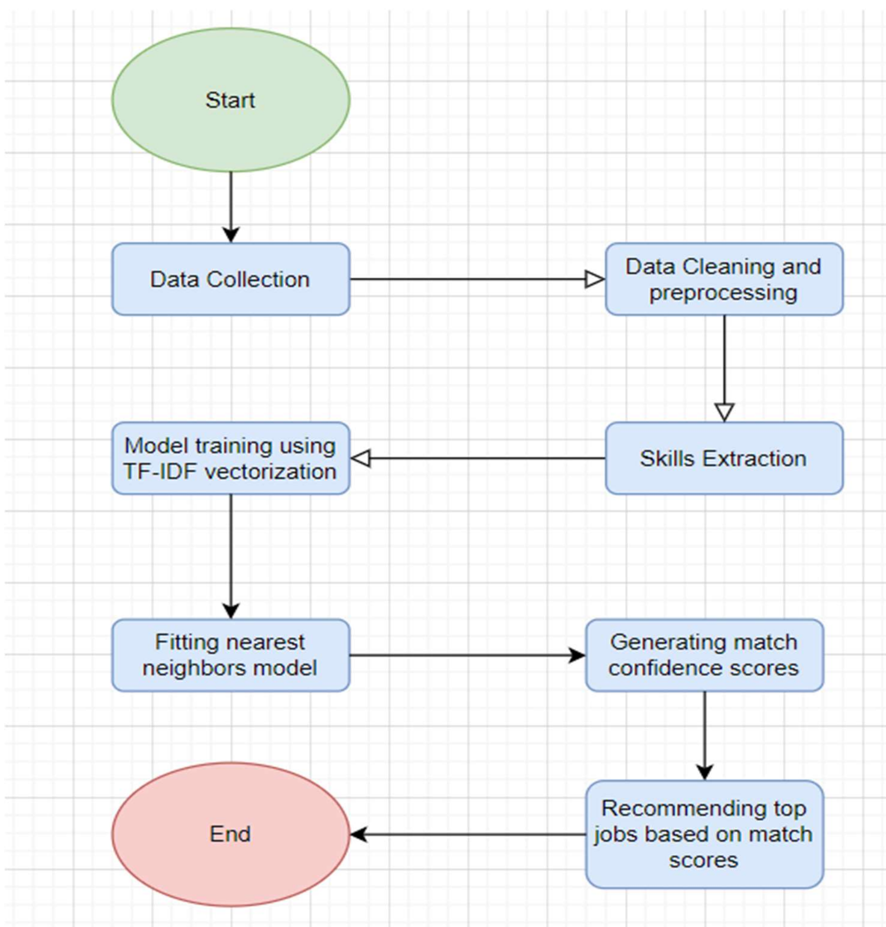
2. Nearest Neighbors (NN):
   - The Nearest Neighbors algorithm is a simple, yet effective, unsupervised machine learning algorithm used for classification and regression.
   - In this code, the `NearestNeighbors` algorithm is employed for finding the nearest neighbors of a query point (in this case, a resume skills vector) among a set of data points (in this case, job descriptions represented as TF-IDF vectors).

- The algorithm finds the top k nearest neighbors for each query point based on a distance metric (usually Euclidean distance or cosine similarity).
- Here, `n_neighbors=1` is set, meaning it finds the closest match.

3. Matching and Recommendation:
- Once the Nearest Neighbors model is trained with the job descriptions, the `getNearestN` function calculates the distances and indices of the nearest neighbors for a given query (resume skills).
- Then, the script iterates over these indices and retrieves the corresponding job descriptions from the dataset.
- Finally, it sorts the top 5 recommended jobs based on the calculated match confidence (distance) and displays them.

In summary, the code leverages TF-IDF vectorization and Nearest Neighbours along with other supervised learning algorithms to recommend the top 5 job matches for a given candidate resume based on the similarity of skills mentioned in the resume with the job descriptions. Further on, additional hybrid models like boosting techniques and DBscan can be applied to improve the accuracy.

Flowchart Diagram-



Results and Discussions

The developed job recommendation system demonstrated promising results in terms of accuracy and user satisfaction. By providing personalized job recommendations, users reported a significant improvement in finding relevant job opportunities matching their skills and preferences. The system also helped employers in reaching out to potential candidates more effectively, thereby reducing the time and resources required for recruitment.

Challenges and limitations:

- **Data quality:** Ensuring the quality and relevance of collected data posed a challenge, especially when dealing with unstructured data.
- **Cold Start Problem:** The system faced challenges in providing recommendation for new users or jobs with limited historical data.
- **Algorithm complexity:** Developing and fine tuning ML algorithms required significant computational resources and expertise.
- **Bias and Fairness:** Ensuring fairness and mitigating biases in job recommendations was crucial to prevent discrimination based on gender, ethnicity or other sensitive attributes.

Future Enhancements:

1. **Real time updates:** Implement mechanisms to continuously update the recommendation models based on new data and user feedback.
2. **Enhanced personalization:** Incorporate advanced techniques such as deep learning and natural language processing for better understanding of user preferences and job requirements.
3. **Fairness and Bias Mitigation:** Develop techniques to detect and mitigate biases in job recommendations to ensure fairness and diversity.
4. **Integration with emerging technologies:** Explore integration with emerging technologies such as blockchain for secure verification of credentials and qualifications.
5. **User Feedback Mechanisms:** Implement a feedback loop to gather user feedback and improve the quality of recommendations over time.

Conclusion:

The development of a job recommendation system using machine learning algorithms holds great promise in revolutionizing the job search and recruitment process. By proving personalized recommendations, the system addresses the challenges faced by both job seekers and employers, leading to more efficient and satisfactory outcomes. However, the continuous efforts are needed to enhance the system’s performance, address limitations, and ensure fairness and transparency in job recommendations.

References:

[1] https://github.com/abbas99-hub/Job-Recommendation-System

[2] https://medium.com/@abbasbehrain95/creating-an-ai-powered-job-recommendation-system-50ce1cd12d36

[3] https://link.springer.com/chapter/10.1007/978-3-031-16075-2_35

[4] https://arxiv.org/abs/2111.13576

[5] https://www.mdpi.com/2073-431X/11/11/161

****