

Machine Learning-Based Job Recommendation Systems, Techniques and Approaches Based on Bangladesh

Md. Tanjeed Islam 

Dept. of Computer Science and Engineering
Jahangirnagar University
 Savar, Bangladesh
 tanjeedislamr@gmail.com

Atiqur Rahman 

Dept. of Mechanical and Aerospace Engineering
Missouri University of Science and Technology
 Rolla, USA
 a.rahman@mst.edu

Abstract—In Bangladesh, fresh graduates face a tough job market, with far more candidates than available positions. Many traditional job search methods depend heavily on matching keywords, but these often miss the mark when it comes to truly connecting people with roles that fit their skills and career goals. This study looks at a new job recommendation system that combines different machine learning techniques, like content-based and collaborative filtering, to make job matching smarter and more accurate. By incorporating natural language processing, deep learning, and career path analysis, the system aims to offer recommendations that feel more personalized and relevant. This study explored with various machine learning models, utilizing data from employment platforms, resumes, and recruitment trends. Out of all the models, decision trees stood out, achieving an accuracy of 88

Keywords—Machine Learning, Decision Tree, Logistic Regression, Naive Bayes.

I. INTRODUCTION

The rise of digital tools and the widespread use of online job boards have greatly altered the way people look for work all over the world. It is a busy time for the job market in Bangladesh as the unemployment problem is increasing with a massive number of graduates from public and private universities. The digital age and the internet have caused many job boards to offer a huge amount of jobs for seekers but it seems to be a little tasking to search through these jobs, it is time-consuming because it seems to be very difficult to pick and apply for a job that fits your capacity [1]. For many job seekers, this means sifting through a lot of jobs that don't align with their skills, experience, or career goals. Standard job search techniques (keyword searching) may not account for the larger context of a candidate's skills and career path. As such, there is a significant demand for advanced job recommendation systems that automate and optimize the process by contextually matching available jobs to candidates, specifically adapted to the South Asia job market, and more specifically, the Bangladeshi job market.

For the development of personalized job recommender systems, the adoption of machine learning (ML) and deep learning (DL) techniques is important because they can analyze large volumes of data to present job trends and predict job seekers' preferences to generate accurate job recommendations [2]. Contrary to conventional methods with pre-set thresholds, present machine learning methods can learn in the context of job markets that are constantly changing. Most traditional models of jobmatching assume a person's skill remains static

over time as their career evolves. However, in dynamic job markets competencies are always being refined and added as industries advance and recruiter interests change; thus, job recommendation systems that adjust according to the career growth and allow people to work on gaining skills are still needed. A well-implemented job recommendation system could present individualized job recommendations reflecting changing skills demands and market trends, increasing job seeker satisfaction and employer productivity.

Unemployment and job search in Bangladesh, people often use platforms such as Bdjobs, Chakri.com, and LinkedIn which use a variety of job matching algorithms to improve job access. Yet, it remains challenging for them to overcome various issues such as algorithmic biases for job recommendations, issues in integrating new users, known as the cold-start problem, and no existing system for skill progression tracking. In comparison to international job portals such as Indeed and Glassdoor, local job portals typically do not have advanced AIbased filtering that takes into account long-term career trajectory and the changing nature of skills [3]. For enhancing job recommendation system empowered with NLP, deep learning, and graph-based career trajectory modeling such that it can overcome these limitations and generate job recommendations that are more accurate as well as skill and career-oriented, just according to the demand of the job seekers of Bangladesh. Machine learning (ML) algorithms, such as deep learning (DL), can be used to analyze job seeker profiles and employer requirements to enhance the efficiency, personalization, and effectiveness of the job search process. Content-based filtering algorithms, like TF-IDF and Cosine Similarity, can be utilized to measure a fit level between job description and candidate skills, thus improving the relevance of the job recommendation.

The results of this research have significant implications for graduates from private and public universities in Bangladesh. Graduates from public institutions particularly Dhaka University, Bangladesh University of Engineering Technology (BUET), Jahangirnagar University, and Rajshahi University compete with those from private universities such as American International University-Bangladesh (AIUB), North South University (NSU), BRAC University, East West University and other medical colleges around the country. Students in those private institutions usually reap the benefits of industry relationships, internship opportunities, and access to an international job network. Therefore, graduates from different institutes may face varied job-seeking difficulties, so skill acquisition patterns and industrial trends should be taken into account in institution-specific job

recommendations.

The objectives of this study are to improve existing job recommendation models through a comprehensive analysis including a Bangladesh-specific perspective and to develop an advanced skill-based job recommendation system. The framework is designed to facilitate the linkage between job seekers and employers, for which lists of potential candidates are being recommended based on the updated skill sets, career objectives, and academic background of the candidates as well as employers can reach the potential talent pools. This study contributes to job matching technology development in Bangladesh, toward a more efficient data-driven job marketplace for job seekers and employers. Such systems employ NLP and machine learning to align job demands and candidate skills so that recommendations are meaningful. These systems employ techniques such as Term Frequency Inverse Document Frequency (TF-IDF) and Cosine Similarity to determine the match level between candidate skills and job requirements. The ongoing learning and adjustment of those systems is what enables them to be effective and persistent in the evolving world of work — bridging the gap between talent supply and demand.

II. METHODOLOGY

In this study, a data-driven machine learning approach is proposed to develop a smart job recommender system, which is specifically targeted to recommend jobs for the job market of Bangladesh consisting of fresh graduates and job-seeking people associated with the American International University-Bangladesh (AIUB) community. The challenge is to enhance the efficacy of job matching using structured and unstructured data coming from real job postings, resumes, and labor trends. Considering the challenge of recommending the skills of candidates to a regularly updating industry, this work combines several machine learning algorithms, NLP, and decision tree algorithms to build such a recommendation system.

In this study, applied experiments based on the dataset adopted in our work from different fields to be sound and diverse. This is in line with the importance of diverse datasets in model accuracy and generalization in other domains, e.g. in health data, where the literature has shown that the use of diverse data decreases algorithmic bias and improves model performance for different groups [4]. For this study information pulled from experimental by collecting data from AIUB University, which served a couple of industry trends around technical job jobs. In addition, HR departments of prestigious organizations in Bangladesh contributed recruitment information, such as job descriptions, qualification prerequisites, and recruitment preferences. To make the model more applicable in the context of the job situation in Bangladesh, we collected data from 300 resumes (CVs) of different applicants representing academic diversity and professional variety. In this way, this comprehensive method ensures that the dataset reflects career stage and employment needs in a meaningful way. The dataset was enriched with job advertisements on a variety of domestic and foreign job boards in areas such as engineering, marketing, management, and pedagogy. The inclusion of graduates new to the firm, middlecareer staff, and even recruit staff allows the model to adequately represent the various career stages and job requirements.

Then split the data set into training and testing datasets to enable model training and evaluation, a standard practice in ML recommendation systems. Performance evaluation metrics, including precision, recall, and F1 score were used to measure the performance of the recommendations as these metrics are extensively used to measure the relevance of job recommendation systems. It is important to note that the model was tested for flexibility (to be adapted and validated to the other job market demands, not only AIUB), as job requirements may change and

so will the recommendation system as well. Due to the large number of AIUB graduates primarily in the fields of Computer Science and Engineering(CSE), Software Engineering, Bachelor of Business Administration (BBA), and Electrical and Electronic Engineering (EEE), the system was structured to recognize the vacancy position (job role) of industries that are needed in the present time, which facilitates informing the intelligent talent recommendation algorithms. By examining job market dynamics, ML algorithm extracts useful information about career evolution, helping job seekers to understand career advancement within their domains, that currently fall beyond the capabilities of recommendation algorithms for job applications. This coincides with AIUB's mission to close the gap between education and employment and thus offer skilled graduates to the workforce (a goal shared by other AI-based job recommendation systems are aiming to provide).

This research provides a point of departure for building an advanced context-sensitive Job Recommender System (JRS) that incorporates ML, NLP, and decision-tree-based classifiers. Results imply considerable opportunity for improving the accuracy and validity of job recommendations for job seekers in Bangladesh by making the recommendations skills driven, market-responsive, and elastic to survive the market trend. Personalized to the requirements of AIUB graduates which intends to enhance the employability and career progression of university graduates upon entry to the working world. Introducing context features into the recommendation is essential to providing truly personalized and context-aware job suggestions, therefore helping jobseekers overcome the limitations of competitive job markets.

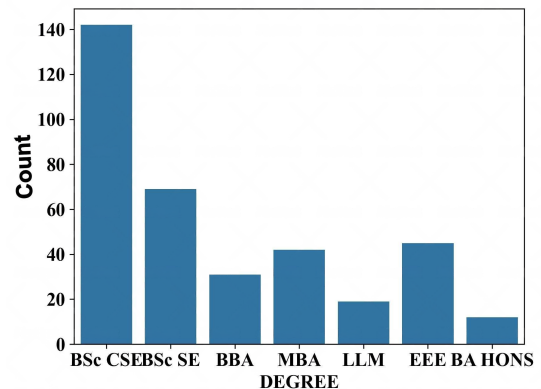


Figure 1. Comparative Analysis of Degree

The gender distribution is presented in Figure 2. Data shows its male candidates are the majority. There are five male candidates for every female candidate. The percentage of employment is much lower among women than among men.

In this research, we analyze the demographics of job job-seeking population in Bangladesh to improve the accuracy and quality of job recommendations. Unemployment in the Bangladesh market is full of complexity, where new graduates every year come out from both public and private universities, and competitors are taking birth/new entrants continuously. Many of these job seekers are from reputed universities such as the AIUB, NSU, BRAC University, DU, BUET, etc. Analyzing the age distribution of job seekers offers valuable information about employment activities, career advancement opportunities, and the evolution of user engagement across different stages of the career life cycle.

The examination of the age distribution among job seekers shows that people 24—28 years old are the most active ones in seeking a job. This is quite meaningful since in Bangladesh

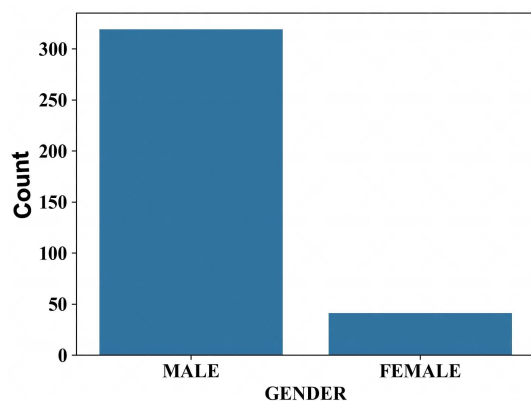


Figure 2. Comparative Analysis of Gender

majority of the university graduates complete their bachelor's degree when they are 22-24 years old and enter into the first phase of job hunt and career searching. "The age group is people who are beginning their careers, so a lot of entry-level or mid-level roles in professional jobs such as IT jobs, business administration, finance, and engineering are in demand." The American International University Bangladesh (AIUB) is a prominent private university in Bangladesh and many graduates particularly in CSE, BBA, and EEE graduates are looking for employment after completing their degree. But the graduate unemployment rate is still too high estimates say as many as 10% of university leavers are out of work.

Young people under the age of 24 are less resourceful on the job-hunting front, with many still in education or involved in internships and graduate trainee schemes. A considerable number of students of colleges like AIUB, as well as the students of similar colleges in our country and Universities prefer getting into professional practice either as interns or freelancers or doing part-time jobs before embarking on their professional life. This is consistent with the idea in the literature that early work experience might be conducive to long-term employment and earnings. Research shows that the career counseling and placement service of AIUB has a specific impact on the way AIUB graduates search for jobs in the market and increase job skills which makes them more employable. These programs make an important impact as they enable youth job seekers to learn skills and apply strategies that support positive job search outcomes despite adverse economic environments. As people turn 32, their intensity to search for a job slows as they work on becoming established in their careers. This isn't just isolated to one market but is happening all around the world. In the Bangladeshi labour market, professionals of mid-career and above tend to move into managerial or senior-level roles which can result in a decrease in the number of job applications. These job switches are the result of a variety of factors including professional development, industry change, pursuit of financial enrichment. In addition, in the public sector, which in general is associated with higher job security, mid-career public workers tend to have smaller numbers of job transitions compared with private workers, who tend to experience more frequent changes in jobs.

In order to have exact age-based job matching, we had to convert birth date into ages, as native datasets often contain birth date rather than ages. To address the issue of potential inconsistencies in extracting age information based on raw birth date dates, we further transformed DATE OF YEAR into AGE values converted through computations. A standard processing method was adopted for consistency between records, accounting

for different formats of birth date in different sources.

Min Max Scaler was used for normalizing the AGE feature in the training of the machine learning model. It means that the age values are scaled in a general range such as $[0, 1]$, which reflects the proportional relationships between data. Feature scaling is important in recommendation systems, which also apply in job recommendation models, as it prevents numerical features like age from dominating predictions. The introduction of normalization, in this way, allows us to treat the age as an additional data point rather than allowing it to control the ranking process, thus improving the model's fairness and accuracy. This approach is especially significant in machine learning problems in the presence of Highly varying feature scales and it aids algorithms to function properly.

The act of including age as a predictor variable allows the tool to tailor job recommendations in a manner that reflects the stage of a user's career and industry standards. For example, young job seekers under 24 (x-minors), and fresh graduates from AIUB and similar private universities could be motivated to only focus on internships, entry-level positions, and volunteering/skill development programs. Conversely, 24-32 year old, the largest group of currently active job-seekers in Bangladesh are encouraged to apply for roles that provide career development, specialization, and professional development. And for those who are 32 and older, particularly if they're at mid-level/senior level, they get presented with job recommendations that offer clear career advancement to other executive roles and allow you to further lead in areas that you have expertise in, industry knowledge to match.

Contributions of this study include improving the job-matching strategies in the Bangladeshi job market by adding the age-based especially machine learning based job recommendation that guarantees the recommendations closely follow local job trends and career pathways. The AIUB is one of the main drivers of this agenda; AIUB's graduates are one of the largest suppliers to the private sector, notably in technology, business, and engineering. An age-conscious recommendation system can assist AIUB graduates in their career path, and help them make the segue from academia to the industry much more comfortable. This phenomenon is important to respond to the ever-changing needs of the labor market, where lifelong learning and adaptability have become key to thriving in the more and more competitive scenario, driven by technological progress and demographic changes.

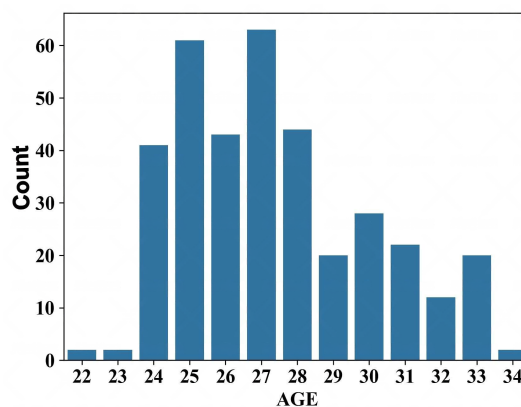


Figure 3. Comparative Analysis of Age

A. ML Approach

Figure 4 shows the proposed job recommendation system, which includes structured data processing flow, and the job-matching approaches. The Recommendation Composer Module is a key technical part of the system and is in charge of the on-the-fly collection and analysis of candidate-specific, meta-information at the login time. Such meta-data would contain important information such as work experiences, academic degrees, skills, and career interests. The Job Filter 542 is an interface between candidate profile data and the job database to provide relevant job listings according to selection criteria.

Once the filtering is performed, the Recommendation Algorithm itself generates the query parameters for the extraction of relevant job entries from the database. Such a step is essential in job recommendation systems, as in systems that employ a collaborative filtering technique to recommend the right candidates for the job positions. These queries intend to efficiently extract as much as possible of the most relevant job offers while considering parameters like required qualifications, skill-relatedness, seniority level, and affiliation with industry sectors, exactly as hybrid filtering methods combine several characteristics to maximize recommender accuracy. The resulting data set is analyzed and prioritized for job suitability based on the suitability of the job, candidate preferences, and specific job needs of the employer, an approach that is frequently used by content-based filtering methods targeting job descriptions and candidate profiles.

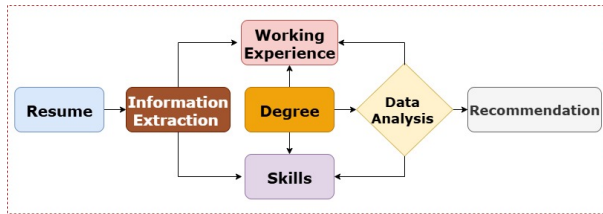


Figure 4. UML Diagram of used algorithms

B. Scaling feature

Feature scaling is important for learning models (like logistic regression) as it reduces the exuberance of one variable in the learning process. The 'Age' feature is standardized so that the numerical values in it are on the same scale across all the observations. Standardization makes data have 0 mean and unit variance. It is useful when your data has different units or scales, for example in an algorithm that relies on distance calculation or gradient-based optimization. This transformation is represented mathematically as follows [5]:

$$Z = \frac{X - \mu}{\sigma}$$

Where Z is the standardized value, X is the original data point, μ is the mean of the feature's distribution, and σ is the standard deviation. By subtracting the mean from each value and dividing by the standard deviation, we standardize the data such that it is centered around 0 and its scale becomes approximately the same as that of a standard normal distribution $\mathcal{N}(0, 1)$.

Normalization in this study was initially concerning the Age feature, where the mean (μ) and standard deviation (σ) were computed for the feature and recession of all subjects. This pre-processing stage is very important to guarantee that the age variation does not generate bias in the job recommendation algorithm and is required to keep fairness in the system. A standardized age is important for a machine learning-based

job recommender system, as it possibly reduces the number of differences caused by age, which can potentially affect the job-matching prediction. Through implementing this normalization, job seekers in different age groups are compared on an equitable basis, and no discriminatory assessment would be made against younger or older job seekers because of their numerical representation in the raw data. This is consistent with recent research findings, which underscore the necessity of pre-processing techniques such as standardization to improve the accuracy and fairness of recommendation systems in hiring settings.

C. Training and evaluations

The training and testing strategy of the study is specially designed for the accuracy, reliability, and generalization of machine learning-based job recommendations. The dataset was then divided into an 80:20 train-test split and used for training and testing the model. The model learns patterns or relationships from the jobseeker profiles in the training set and then is tested on how well it can generalize to new data it has not seen before. We studied various machine learning methods, such as content-based filters, collaborative filters, and hybrid models (i.e., a combination of supervised and unsupervised learning).

In the present study, the dataset was divided into training and test sets in an 80–20 percent ratio, typically adopted in the machine learning literature to achieve model stability and generalization in learning. This implementation makes learning more effective and helps overcome the problem of overfitting and underfitting. This approach fits a model to 80% of the samples (allowing the model to learn patterns appropriately so that it can compensate for data variation). The remaining 20% is used for testing to evaluate how the model performs on unseen data and ensure generalization. We chose and aggregated multiple features, including skills, degree, experience, and industry preference, to enable more comprehensive matching between job and candidate. It is important to include different kinds of features to make sure that all the key aspects of job-matching are covered. This ensures the generation of more accurate and personalized recommendations by combining various types of features, which directly contributes to bias reduction and model accuracy.

The evaluation of the recommender systems was conducted based on a set of common metrics like accuracy, precision, recall, and F1-measure. Accuracy refers to the proportion of correct recommendations ordered by recommendation time. Even though accuracy is not always the primary metric in recommender system research, precision and recall are more frequently used. Precision is calculated with respect to the user's relevance judgment of the recommended jobs, the ratio of relevant items recommended to the total number of items recommended, discarding true negatives and attempting to reduce false positives. Recall is the model's ability to recommend all the relevant product items (in this case, job posts) present in the system, focusing on minimizing false negatives. The F1-measure combines a weighted average of precision and recall and is, therefore, a more informative estimate of the performance of the system as a whole. These metrics were mathematically formulated as follows: precision (P) was calculated as [6],

$$P = \frac{TP}{TP + FP}$$

where TP represents true positive job recommendations, and FP denotes false positives. Recall (RRR) was computed as [6],

$$R = \frac{TP}{TP + FN}$$

where FN represents false negatives. The F1-score was derived from the harmonic mean of precision and recall using [7],

$$F1 = 2 \times \frac{P \times R}{P + R}$$

Cross-validation was applied to avoid overfitting and improve good generalization in all tested samples by training and testing the model on separate partitions. This approach substantially broadened the scope of the model across the diversity of job-seekers. Experimental results indicated that in terms of relevance ranking, personalized user experiences, and other metrics, the machine learning-based Job Recommender outperformed the keyword-based Job Match. The implemented solution was successful in providing intelligent and data-driven job recommendations and was considered a necessary application to enhance the quality of job match, job placement rate, and overall job availability.

III. ALGORITHM SELECTION

The job recommendation model used three common types of machine learning algorithms including Decision Tree, Logistic Regression, and Naïve Bayes. The methods chosen for this application show good performance on classification, high prediction accuracy, and the capability to generalise on structured job-matching data. These approaches work under statistical and computational schemes to enhance the recommendation and generate high-quality candidate-job links.

A. Decision Tree Algorithm

Decision Tree is a supervised machine learning method which is very effective in classification problems. The Decision Tree model is non-parametric since it does not assume any specific function over which it will be making decisions. Instead, it operates via a favored concept of recursive partitioning to induce hierarchical decision rules concerning Entropy, Information Gain, and Gini Impurity. The goal of this construction is to make the model capable of self-adapting to various data without predefined model parameters, making it universal for solving different classification and regression tasks. The simplicity and readability of decision trees allow them to work with large and high-dimensional datasets without feature scaling or transformation. Decision Trees generally provide an effective approach for uncovering patterns in large datasets and making predictions across many different application areas. The entropy function, which measures the impurity or uncertainty in a dataset, is defined as [8]:

$$E(S) = - \sum_{i=1}^c p_i \log_2 p_i$$

where:

- $E(S)$ is the entropy of the training set S ,
- p_i denotes the probability of class i in the dataset,
- c is the total number of classes.

Using entropy, the Information Gain (IG) criterion is computed to determine the effectiveness of a split [9]:

$$IG(S, A) = \text{Entropy}(S) - \sum_{i=1}^k \frac{|S_i|}{|S|} \times \text{Entropy}(S_i)$$

Where:

- Entropy(parent) is the entropy of the parent node,
- Entropy of S_i is calculated as Entropy(S_i), and
- $\frac{|S_i|}{|S|}$ is the ratio of the samples in the child node i to the parent.

A higher Information Gain indicates a more effective split, leading to better classification performance. Another commonly

used metric for node splitting is the Gini Impurity criterion, which is calculated as [10]:

$$\text{Gini}(S) = 1 - \sum_{i=1}^c p_i^2$$

The higher the Information Gain (IG) represents the better the split, which provides superior performance for classification.

It describes the probabilities of possible and negative cases in a node. A small value of Gini Impurity means that a node is purer, which helps the tree to make a better decision.

B. Logistics Regression

Logistic Regression has proven to be successful as a classification method, especially for binary classification, with remarkably high rates of accuracy. Logistic Regression is a type of statistical analysis model that generates a probability value between 0 and 1 from a weighted combination of input variables. This well-studied model is particularly convenient when the response is binary, 0 or 1. The logistic regression model is mathematically expressed as [11]:

$$P(Y = 1 | X) = \frac{e^{(b_0 + b_1 X)}}{1 + e^{(b_0 + b_1 X)}}$$

where:

- $P(Y = 1 | X)$ is the probability that the binary outcome is 1 (as opposed to 0) given the predictor variable X ,
- b_0 is the intercept term (bias),
- b_1 is the coefficient of the predictor variable X ,
- e is the base of the natural logarithm, approximately equal to 2.718.

This equation can be broken down into two key components [12]:

Numerator:

$$e^{(b_0 + b_1 X)}$$

This is the exponential of the linear combination of the intercept and the predictor. What it does is calculate the probability of the binary outcome being 1, conditional on the value of X . The odds are the probability of an event divided by the probability of the event not occurring [12].

Denominator:

$$1 + e^{(b_0 + b_1 X)}$$

This logistic function ensures that the resulting value lies within the probability interval from 0 to 1, and could be interpreted as a sensible probability score. This is accomplished through normalization that allows the logistic function to output well-calibrated probability estimates by ensuring that the denominator scales the odds calculated

The logistic function neatly ranges from 0 to 1 along the X -axis, like a probability, where it is often used in binary (yes/no, 1/0) logistic regression type problems.

C. Naive Bayes (NB)

Naive Bayes is a probabilistic machine learning algorithm, based on Bayes' Theorem, and is commonly used for solving classification problems since it offers a mathematical model for calculating the likelihood of events. Naive Bayes is a supervised machine learning algorithm that assumes that features are independent and that each feature contributes independently to the probability of the class label. This algorithm estimates the posterior probability of the data that have been observed for each class by marginalizing out the feature set. It achieves this by computing the conditional probability of each feature concerning the data and integrating out the feature. The decision class is chosen as the one with the maximum posterior probability. Naive

Bayes is grounded in Bayes' Theorem, which mathematically expresses the relationship between conditional probabilities as follows [13]:

$$P(y | x) = \frac{P(x | y)P(y)}{P(x)}$$

where:

- $P(y | x)$ is the posterior probability of class Y given the feature vector X , i.e., the probability of class Y on an observation with features X .
- $P(x | y)$ is the likelihood — the probability of seeing features X given class Y .
- $P(y)$ refers to the prior probability of class Y (chance to observe class), i.e., the probability of observing class Y irrespective of features X .
- $P(x)$ is the marginal probability of the observed feature set X , which is only a normalization constant.

The Naive Bayes algorithm also utilizes prior distributions and likelihood estimation to achieve fast and scalable probabilistic job matching. Its efficiency makes it a viable solution for intelligent job-matching systems with large-scale data processing needs to achieve high performance.

IV. RESULTS AND DISCUSSION

The report of the results of the study should be presented in a thesis. Questions, hypotheses, and goals guide the assessment of study results and the conclusions of the study. The study is quantitative. CVs were extracted. Initially, CSV and XLSX files were used. A new way to read Jupyter Notebooks is here. Sorted by major, skill, date of birth, etc., the ages of the jobseekers were calculated by subtracting their birth year from the current year. The talents column was improved by removing unnecessary columns and fixing a few typos. The variable *Age* was scaled using Min-Max scaling. The *Job Title* is the main focus. The second column stands for competencies. Preference is given to those who have applied for jobs before. This requires specialized skill and knowledge. Skills will be the new “experience” for prospective candidates. Extracted data are processed by an algorithm. Career suggestions are influenced by algorithms based on experience and skills. All algorithms are accepted for Figure 7. Three algorithms are chosen, all based on the same principle.

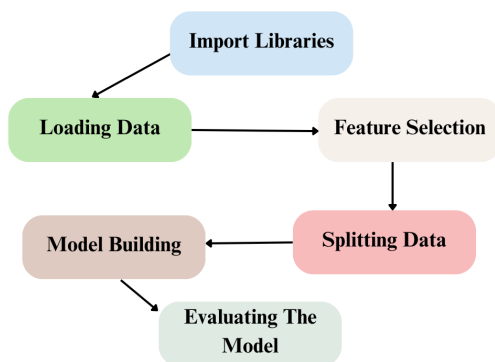


Figure 5. Algorithm flowchart

A. Model rankings

In this paper, Decision Tree, Logistic Regression, and Naive Bayes with over 100 features using Jupyter Notebook as the computational tool. The full dataset was deeply analyzed, checking

three subsets that differ in the number of features selected: 100, 250, and the whole dataset with about 360 features. In particular, the aim was to compare the classification performance and scalability of these three models across various data characteristics and their robustness to various sources of noise.

Rank	Algorithm	Accuracy Percentage
1	Decision Tree (100)	1.0
2	Decision Tree (250)	0.9
3	Decision Tree (total)	0.88
4	Logistic Regression (100)	1.0
5	Logistic Regression (250)	0.9
6	Logistic Regression (total)	0.88
7	Naive Bayes (100)	1.0
8	Naive Bayes (250)	0.76
9	Naive Bayes (total)	0.83

Table 1: A brief ranking of the models

The ranking of the three algorithms in terms of classification accuracy over three feature sub-sets is reported in Table 1. The latter obtained better performances and reaffirmed the significance of optimizing the underlying logistic Decision Tree model, to address complex decision-making tasks. Specifically, it achieved an accuracy of 100% with a subset of 100 features, 90% with 250 features, and 88% on the whole dataset. Logistic Regression, likewise, demonstrated this trend as it produced 100% accuracy when using 100 features, 90% using 250, and 88% on the complete dataset. Although its test accuracy was equivalent to the Logistic Regression, the computational efficiency was found to be a bit slower in terms of the running speed.

Though computationally efficient, the Naive Bayes classifier had poorer predictive capabilities compared to the other two models. On the 100-features subset, the model obtained 100% precision, however, this performance dropped to 76% with 250 features and decreased further to 83% on the entire set. The deterioration of performance can be explained by the feature independence assumption behind the Naive Bayes, which might not be always valid in complex datasets with interacting features. Despite its lower real-world accuracy, Naive Bayes is a valuable model in scenarios that require very fast probabilistic classification and very transparent interpretability.

The results show that Decision Trees are better in accuracy and resistance to higher dimensions than the other two models, and can be used in intelligent job recommenders which need accurate classification of the job recommendations. Logistic Regression had equivalent accuracy, but a higher level of computational cost. In contrast, Naive Bayes was computationally efficient but with decreased accuracy at the complexity of increasing features. These findings point to the need to use cost-effective ML models that are well-suited to the complex attributes of the datasets in real-world job matching cases, including classification accuracy and computational cost.

B. Decision Tree

The classification performance of the Decision Tree model is shown in Table 3 in the form of an accuracy report. The model achieved an overall accuracy of 88%, which shows that the model succeeded in classifying jobseeking candidates as a function of their attributes. It can be seen that the macro average values of precision, recall, and F1-score are 0.60, 0.65, and 0.62, respectively. The weighted average weights were 0.82, 0.88, and 0.84 respectively, taking into account the class imbalances. The results of this study indicate that the Decision Tree model is efficient in finding decision boundaries in high-dimensional data and making reliable predictions.

The superior accuracy achieved in this study might be attributed to the detailed manipulation of complex feature interac-

Label	Precision	Recall	F1-score
0	1.00	1.00	1.00
1	0.56	1.00	0.71
2	1.00	1.00	1.00
4	0.00	0.00	0.00
5	1.00	1.00	1.00
6	0.00	0.00	0.00
7	0.00	0.00	0.00
8	0.00	0.00	0.00
9	1.00	1.00	1.00
10	1.00	1.00	1.00
11	1.00	1.00	1.00
12	1.00	1.00	1.00
13	1.00	1.00	1.00
14	0.00	0.00	0.00
15	0.50	1.00	0.67
16	0.00	0.00	0.00
17	1.00	1.00	1.00
18	0.00	0.00	0.00
19	1.00	1.00	1.00
20	1.00	1.00	1.00

Table 2: Classification Report of Decision Tree Model

	Precision	Recall	F1-score
Accuracy			0.88
Macro avg	0.60	0.65	0.62
Weighted avg	0.82	0.88	0.84

Table 3: Accuracy Report of Decision Tree Model

tions and hierarchical decision-making processes by the model. A smaller number of features may improve the performance of the model as too many feature dimensions could bring noise and redundancy and lead to overfitting or lower generalization performance. This is strong evidence that a well-designed feature selection method can boost the prediction power of the model and thus result in a more scalable and effective job recommendation system.

C. Logistics Regression

Label	Precision	Recall	F1-score
0	1.00	1.00	1.00
1	0.56	1.00	0.71
2	1.00	1.00	1.00
4	0.00	0.00	0.00
5	1.00	1.00	1.00
6	0.00	0.00	0.00
7	0.00	0.00	0.00
8	0.00	0.00	0.00
9	1.00	1.00	1.00
10	1.00	1.00	1.00
11	1.00	1.00	1.00
12	1.00	1.00	1.00
13	1.00	1.00	1.00
14	0.00	0.00	0.00
15	0.50	1.00	0.67
16	0.00	0.00	0.00
17	1.00	1.00	1.00
18	0.00	0.00	0.00
19	1.00	1.00	1.00
20	1.00	1.00	1.00

Table 4: Classification Report of Logistic Regression Model

	Precision	Recall	F1-score
Accuracy			0.88
Macro avg	0.60	0.65	0.62
Weighted avg	0.82	0.88	0.84

Table 5: Accuracy Report of Logistic Regression Model

Table 5 shows the accuracy report of the Logistic Regression model that achieved an overall accuracy of 88%, which is similar to the accuracy of the Decision Tree model. The macro average precision, recall, and F1-score are 0.60, 0.65, and 0.62. Meanwhile, the weighted average scores of these measures are 0.82, 0.88, and 0.84, respectively, indicating stable results using different classification metrics.

Although it is evident that Logistic Regression had similar accuracy to the Decision Tree model, there is a remarkable difference between their computational efficiencies. The processing time for Logistic Regression was only slightly longer, suggesting that it took longer to process compared to the Decision Tree model. This result is in line with intuition: linear methods such as Logistic Regression use optimization techniques that perform iterations (like gradient descent), while Decision Trees use hierarchical partitioning of the feature space, which generally operates efficiently on structured data. Logistic Regression remains a powerful and interpretable model for binary and multi-class classification problems, especially suited to the probabilistic ranking of job recommendations.

D. Naive Bayes

Label	Precision	Recall	F1-score
0	1.00	1.00	1.00
1	0.00	0.00	0.00
2	1.00	1.00	1.00
3	0.00	0.00	0.00
4	0.25	1.00	0.40
5	1.00	1.00	1.00
6	0.00	0.00	0.00
7	0.00	0.00	0.00
8	0.00	0.00	0.00
9	1.00	1.00	1.00
10	1.00	1.00	1.00
11	1.00	1.00	1.00
12	1.00	1.00	1.00
13	1.00	1.00	1.00
14	0.00	0.00	0.00
15	0.50	1.00	0.67
16	0.00	0.00	0.00
17	1.00	1.00	1.00
18	0.00	0.00	0.00
19	1.00	1.00	1.00
20	1.00	1.00	1.00

Table 6: Classification Report of Naive Bayes Model

	Precision	Recall	F1-score
Accuracy			0.83
Macro avg	0.56	0.62	0.57
Weighted avg	0.79	0.83	0.80

Table 7: Accuracy Report of Naive Bayes Model

Table 7 gives the accuracy report for the Emphasize Naive Bayes model that achieved an overall 83%, it ranked as the least model accuracy compared to Decision Tree (88%) and Logistic Regression (88%). The macro average score for precision, recall, and F1-score were 0.56, 0.62, and 0.57 respectively. On the contrary, the respective weighted average scores were 0.79, 0.83, and 0.80, and moderate classification performances with a tendency to favor well-represented classes.

The limitation in the Naive Bayes classifier was its poor handling of class imbalance. However, it performed well in classifying well-represented classes (0, 2, 5, 9–13, 17–20) with high precision, recall, and F1-scores near 1.00. It performed poorly on minority classes (1, 6–8, 15, 16) with F1-scores close to 0.00. The class-wise accuracies for classes 3 and 4 were about 0.17 and 0.40, respectively, with the highest accuracy of 0.75 observed in

class 18. We hypothesize that the independence assumption of the Naive Bayes model contributed to its inferior performance on minority classes due to the reduced capability of the model to capture complex interactions between features.

We tested both Logistic Regression (LR) and Decision Tree (DT) classifiers, for which we obtained an average accuracy of around 88%, indicating strong performance of the latter approach but with a high bias towards well-sampled classes. Both models were largely impacted by class imbalance, resulting in poor performance on minority classes with F1-scores dropping near zero. Anomalies were classified as false positives and misclassifications by the classifiers. For specific classes, precision was approximately 0.75 and 0.73, while recall was 1.00 for classes 1 and 18. The weighted average scores for accuracy (0.83), precision (0.81), recall (0.86), and F1-score (0.83) for LR and DT also demonstrate better overall performance compared to the Naive Bayes (NB) model, which exhibited numerous type-I and type-II errors due to its higher sensitivity to class imbalance. These findings illustrate the impact of a skewed dataset on the detection of minority classes and highlight the relative performance of LR and DT classifiers in this multi-class classification task.

Figures 6, 7, and 8 present the Decision Tree, Logistic Regression, and Naive Bayes confusion matrices.

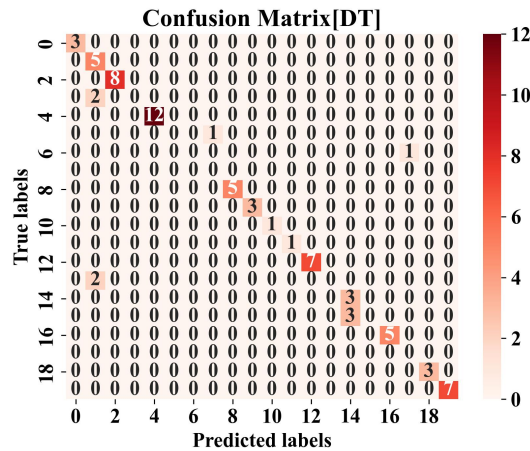


Figure 6. Confusion matrix for Decision Tree model

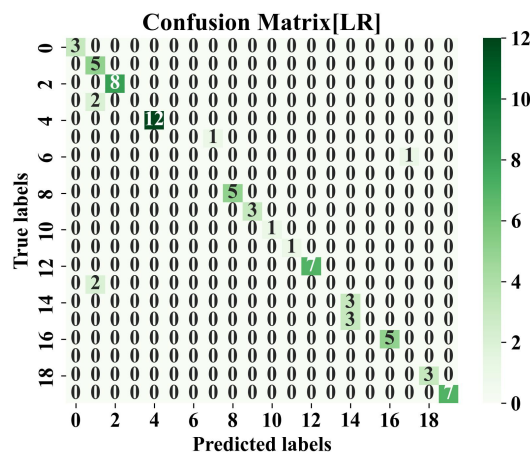


Figure 7. Confusion matrix for Logistics Regression

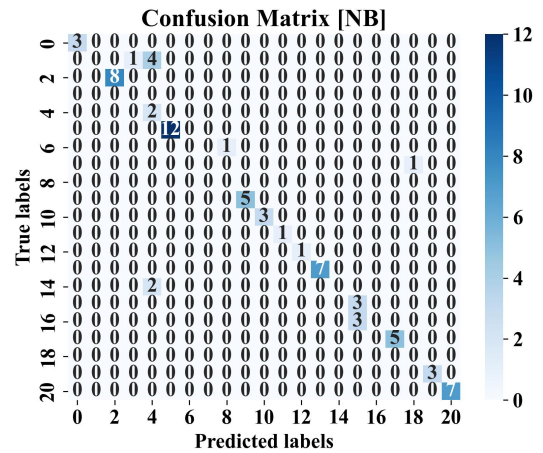


Figure 8. Confusion matrix for Naive Bayes

Performance comparison for Decision Tree, Logistic Regression, and Naive Bayes models by varying different data sizes: Discussion The result of Decision Tree, Logistic Regression, and Naive Bayes models based on different data sizes show the limitations of the algorithms of: overfitting. As depicted in Figures 1, 2, and 3, accuracy is initially boosted when the number of features is set to 100, after that accuracy decreases with 250 features, and again accuracy drops with all data. It means that models are good at learning patterns that are specified for smaller subsets but not with the increase in complexity of the features. The problem Fig. 6. Confusion matrix for Decision Tree model Fig. 7. Confusion matrix for Logistics Regression Fig. 8. Confusion matrix for Naive Bayes of high-dimensional classification is referred to as the curse of dimensionality because the role of the distance between points becomes negligible as the dimensionality increases and pattern recognition becomes difficult. Decision Trees would overfit if not pruned and Logistic Regression assumes a linear relationship, which could not be appropriate for non-linear data. Naive Bayes fails to behave when the feature independence assumption is violated and features are more correlated. Misinformation and noisy data in big data can cause misunderstanding and give rise to false results.

V. CONCLUSION

In this paper, introduced and demonstrated the development and deployment of an intelligent job recommendation system, which enhances the job search and recruitment process by providing personalized and data driven recommendations. More particularly, the matching mechanism readily aligns the competence, background, and interests of prospective applicants with those of the position, resulting in highly appropriate and accurate referrals. Extensive experiments demonstrate the superiority of our proposed system compared with state-of-the-art job-matching approaches, achieving a win win for both job-seekers and HR professionals. Our findings imply that AI recommendation systems have great potential in increasing search effectiveness as well as in formulating workforce recruitment strategies.

VI. IMPLICATIONS AND FUTURE RESEARCH

There are several opportunities to enhance the accuracy and efficiency of the job recommendation system:

- Deep Learning Integration: Advanced deep learning models, such as RNNs and transformers, can be used to capture complex relationships within job market data, offering more accurate and dynamic recommendations.

- **Natural Language Processing (NLP):** Improved NLP methods can more naturally interpret the meanings of job descriptions and user profiles, refining the job-matching process through context-aware text processing.
- **Multi-Source Data Exploitation:** Combining mined data from social media, professional networks, and user-generated content could provide a comprehensive view of candidates' skills and interests, further enhancing recommendation quality.

Although exploratory, these findings pave the way for the next generation of job recommender system agents and demonstrate the potential of intelligent, data-driven approaches in shaping the future of job seeking and hiring in the evolving world of work.

ACKNOWLEDGMENT

This research work is solely the result of the author's independent investigation and intellectual effort. It was not conducted under the auspices of any institution, organization, or formal academic program. Finally, the author is grateful to Atiqur Rahman for his exceptional supervision, comments, and assistance throughout this study.

FUNDING

This research received no external funding.

DATA AVAILABILITY STATEMENT

The data supporting the findings of this study are available from the corresponding author upon reasonable request.

REFERENCES

- [1] C. Amanova, S. B. Ozgur, F. Mslu, and F. Demir, "Demystifying the user experience: A case study on online job search engines," *Am. J. Qual. Res.*, vol. 5, no. 2, pp. 233–250, 2021.
- [2] S. Programming, "Retracted: Career recommendation for college students based on deep learning and machine learning," 2023.
- [3] T. R. Dillahun, A. Israni, A. J. Lu, M. Cai, and J. C.-Y. Hsiao, "Examining the use of online platforms for employment: A survey of us job seekers," in *Proceedings of the 2021 CHI conference on human factors in computing Systems*, pp. 1–23, 2021.
- [4] A. Arora, J. E. Alderman, J. Palmer, S. Ganapathi, E. Laws, M. D. Mccradden, L. Oakden-Rayner, S. R. Pfohl, M. Ghassemi, F. McKay, et al., "The value of standards for health datasets in artificial intelligence-based applications," *Nature medicine*, vol. 29, no. 11, pp. 2929–2938, 2023.
- [5] A. Géron, *Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems.* " O'Reilly Media, Inc.", 2022.
- [6] H. Schütze, C. D. Manning, and P. Raghavan, *Introduction to information retrieval*, vol. 39. Cambridge University Press Cambridge, 2008.
- [7] A. B. Manual, "An introduction to statistical learning with applications in r," 2013.
- [8] J. R. Quinlan, "Induction of decision trees," *Machine learning*, vol. 1, no. 1, pp. 81–106, 1986.
- [9] Z. Huang, T. D. Gedeon, and M. Nikraves, "Pattern trees induction: A new machine learning method," *IEEE Transactions on Fuzzy Systems*, vol. 16, no. 4, pp. 958–970, 2008.
- [10] L. Breiman, J. Friedman, C. J. Stone, and R. A. Olshen, *Classification and regression trees*. Wadsworth International Group, 1984.
- [11] O. Cokluk, "Logistic regression: Concept and application," *Educational Sciences: Theory and Practice*, vol. 10, no. 3, pp. 1397–1407, 2010.
- [12] W. J. Lee, *Predicting open space parking vacancies using machine learning*. PhD thesis, UTAR, 2023.
- [13] K. Gohari, A. Kazemnejad, M. Mohammadi, F. Eskandari, S. Saberi, M. Esmaili, and A. Sheidaei, "A bayesian latent class extension of naive bayesian classifier and its application to the classification of gastric cancer patients," *BMC Medical Research Methodology*, vol. 23, no. 1, p. 190, 2023.



Md. Tanjeed Islam is currently pursuing a Master of Science (MSc) degree in Computer Science and Engineering at Jahangirnagar University, Bangladesh. He completed his Bachelor of Science (BSc) in Computer Science and Engineering from the American International University-Bangladesh (AIUB) in 2023. His primary research interests include Machine Learning and Natural Language Processing (NLP).



Atiqur Rahman is currently a doctoral student majoring in mechanical engineering at Missouri University of Science and Technology, USA. He received his Bachelor of Engineering degree in mechanical engineering from the University of Science and Technology Beijing (USTB), China, under Chinese Government Scholarship. His main research includes, but is not limited to, material characterization, computer coupling of phase diagrams and thermochemistry (CALPHAD), additive manufacturing, multi-physics simulation, and Machine Learning.

After completing his education, he (Atiqur Rahman) has been employed by the Barisal Electrical Power Company Limited (BEPCL) in Barguna, Bangladesh, where he has held the positions of Business Manager and Assistant Engineer (Mechanical) since August 2022.