*Production Systems and Information Engineering*
Volume 14 (2), pp. 51-62
doi: https://doi.org/10.32968/psaie.2026.2.4


# A TEXT MINING-BASED DECISION SUPPORT SYSTEM FOR CANDIDATE SELECTION IN GOVERNMENT RECRUITMENT PROCESSES

Madiha M. M. Hussain[1]
Faculty of Computer Science and Information Technology
Omdurman Islamic University
madihame.mahdi@gmail.com

Hussein A. A. Ghanim[2]
Faculty of Computer Science and Information Technology
University of Kassala, Sudan
ghanim@kassalauni.edu.sd

Ibrahim M. A. Ali
Faculty of Computer Science and Information Technology
Karary University
ibrahim1630@gmail.com



**Abstract.** Long processing times, manual labor, and human biases are just a few of the major inefficiencies that frequently plague the public sector hiring process and cause delays in the timely acquisition of qualified talent. In order to improve and automate the hiring process for government agencies, this paper suggests a brand-new text mining-based decision support system. The suggested system parses, examines, and rates resumes in accordance with particular job descriptions using machine learning and natural language processing (NLP) techniques like XGBoost Classifier, TF-IDF, and Sentence-BERT Embeddings. The results demonstrate that the candidate selection decision support system successfully integrates structured features and text-based semantic analysis, improving candidate ranking and reducing false positives and negatives while encouraging ethical recruitment practices. Its accuracy, precision, and recall for identifying relevant candidates within the top 20% of the applicant pool are 82%, 92%, and 85%, respectively.

*Keywords*: *Text Mining, Decision Support System, Candidate Selection, Government Recruitment, Natural Language Processing, Resume Screening, Machine Learning*

## 1. Introduction

The efficiency and integrity of a nation's public service are directly related to the quality of its personnel. However, the traditional recruitment landscape in the public sector is frequently plagued by systemic inefficiencies [1]. For a single position, government agencies frequently receive thousands of applications; this volume overwhelms human resource (HR) departments and results in a resource-intensive, time-consuming, and frequently subjective screening process [2]. This

manual review is not only slow but also disposed to human cognitive biases, such as confirmation bias and confirmation effect, which can lead to the overlooking of highly qualified candidates and the protection of demographic differences. The resulting delays in hiring can impede organizational progress and service delivery, creating a pressing need for a fundamental shift in how public sector recruitment is managed [3].

This study admits this difficulty and suggests creating an intelligent, text-mining-based decision support system to streamline and automate the initial screening stage of government hiring. The main objective is to abandon the conventional "CV-sifting" paradigm and adopt a transparent, data-driven, and objective strategy. Our system seeks to give hiring managers and HR specialists a strong tool to help them make better decisions by utilizing developments in data science and artificial intelligence.

Defining and creating an intelligent system architecture that can automatically process and analyse job descriptions and resumes in line with the main goals of this study is to extract, categorize, and measure the applicability of candidate experiences and skills using text mining and natural language processing techniques, to develop a strong ranking algorithm that produces a qualified shortlist by assigning scores to applicants according to how well they fit a particular role, to show how a system like this can shorten the time it takes to hire someone and lessen the effect of human prejudice during the first stage of candidate screening, and to investigate how the proposed system is incorporated into a larger e-Government transformation plan to improve public service operations.

This paper offers a roadmap for the modernization of hiring procedures in the public sector by outlining the decision support system's design, implementation, and possible effects.

This article is organized as follows: the introduction in the first section. The second section displays the related work, the third section illustrates the Comparison of Applied Methods, the fourth section presents the Proposed Model and their explanation in more details, the fifth section demonstrates the Implementation of the proposed model using Python, section sixth explain the Results and Discussion, and the Conclusion and Future Work in the seventh section.

## 2. Related Work

For over 20 years, research on the application of data science and artificial intelligence (AI) in the human resources sector has been growing. Early work focused on expert systems and rule-based approaches to automate simple administrative tasks [4]. With the advent of big data and significant progress in machine learning, the focus has shifted towards more sophisticated systems capable of handling unstructured data, such as candidate resumes and cover letters.

Several works have been done by the scholar in the literature, this paper mention some of them. Darabi [5] introduce a novel system called skilled miner system. The Skill Miner System (SMS) is a text-mining program that creates a Skill Demand Index (SDI) by extracting and analyzing professional and technical abilities that employers are looking for. The method facilitates improved career preparation and job market competitiveness by bridging the gap between industry demands and university courses. It has been integrated into UIC's professional development courses.

Ningrum et al. [6] provides study that investigates recruitment discrimination in Indonesia by analyzing 8,969 online job ads from 2005–2017 using a Direct

Discrimination Detection (DDD) method that applies N-grams and regular expressions to identify bias in large-scale text data. It addresses the shortcomings of manual content analysis by focusing on five forms of direct discrimination: gender, age, marital status, physical appearance, and religion. According to the findings, age discrimination is the most prevalent (66.27%), followed by gender (38.76%) and physical appearance (18.42%). Women, especially those who are young, unmarried, and have certain physical characteristics, face additional disadvantages. The results underscore the persistence of bias in hiring and the need for stronger anti-discrimination measures. George and Navyamol [7] explore that their work automates manual resume screening in recruitment by developing a resume parser using natural language processing approaches. This improves efficiency, accuracy, and candidate-job matching for HR departments. The system ranks candidates by relevance, reducing recruiters' workload and improving hiring decisions.

Otani et al. [8] introduces the use of natural language processing (NLP) in HR to enhance recruitment accuracy, automate repetitive tasks, and support better workforce analytics. It highlights challenges like bias and data privacy but highlights its potential to enhance efficiency and data-driven decision-making.

Pietro [9] develops an automatic text-mining method to identify and extract job titles from diverse sources, improving recruitment, policy-making, and workforce training. The method uses linguistic clues, rule-based filtering, and Python scripts, achieving 91% precision and revealing 64% new green job titles, proving its potential for maintaining comprehensive occupational data.

In order to overcome the drawbacks of human text categorization in big datasets, Kobayashi et al. [10] present automatic text classification (TC) algorithms for organizational study. Because TC is quicker, less expensive, and more reliable, it can analyze large corpora more effectively, enhance validity, and assist with research activities like sentiment analysis and job analysis. These studies examine different uses of natural language processing (NLP) and text mining in organizational and human resources (HR) research. In addition to addressing social equity, identifying in-demand skills, automating resume screening, improving recruitment accuracy, task automation, and workforce analytics, they also concentrate on scalable text classification and automatic job title extraction for quicker, less expensive, and more reliable

## 3. Theoretical Background and Related Methods

In order to overcome the drawbacks of human text categorization in big datasets, Kobayashi et al. [10] present automatic text classification (TC) algorithms for organizational study. Because TC is quicker, less expensive, and more reliable, it can analyze large corpora more effectively, enhance validity, and assist with research activities like sentiment analysis and job analysis. These studies examine different uses of natural language processing (NLP) and text mining in organizational and human resources (HR) research. In addition to addressing social equity, identifying in-demand skills, automating resume screening, improving recruitment accuracy, task automation, and workforce analytics, they also concentrate on scalable text classification and automatic job title extraction for quicker, less expensive, and more reliable analysis of large textual datasets.

Table 1 - Comparison of Applied Methods

| Stage | Description | Advantages | Disadvantages |
|---|---|---|---|
| Traditional Manual | HR specialists manually review each resume and | - Full human judgment and contextual | - Highly labor-intensive and time-consuming. |

| | | | |
|---|---|---|---|
| Approaches | cover letter. Appropriateness is judged quickly, often within seconds. | understanding.<br>- Allows for qualitative assessment of soft skills (if read fully). | - Prone to human errors and unconscious bias.<br>- Inconsistent and subjective.<br>- Unreliable due to cursory reviews. |
| Semi-Automated Approaches (ATS) | Applicant Tracking Systems digitize applications and store candidate data in searchable databases, primarily using keyword matching. | - Improves organization and storage.<br>- Faster initial screening.<br>- Standardized filtering criteria. | - Limited to rigid keyword searches.<br>- Easily gamed by keyword stuffing.<br>- Fails to detect contextual meaning or synonyms.<br>- Cannot infer skill proficiency. |
| Intelligent Systems (Proposed) | Uses text mining and machine learning for semantic understanding of job descriptions and candidate profiles. Generates a continuous relevance score and transparent ranking. | - Understands context and synonyms.<br>- More accurate and fair candidate assessment.<br>- Continuous scoring instead of binary filtering.<br>- Reduces bias and improves objectivity.<br>- Increases efficiency in shortlisting. | - Requires high-quality data and well-trained models.<br>- Higher initial development cost.<br>- Needs ongoing model updates and maintenance. |

### 3.1 Training Dataset Description

The effectiveness of the proposed text mining framework depends heavily on a high-quality, representative training dataset. This section details the sources, size, composition, and labeling process used in our experiments.

*3.1.1 Data Sources*

Two primary document types were collected:

- Job descriptions: Sourced from public job portals (LinkedIn, Indeed) and anonymized company HR archives from the IT and finance sectors. Each job description includes title, responsibilities, required skills, qualifications, and location.
- Candidate documents: Anonymized CVs/resumes obtained from two open-access datasets:
  - Kaggle "Resume Dataset" (v2, 2023)
  - Synthetic CVs generated using a rule-based template to augment underrepresented job categories

*3.1.2 Dataset Size and Split*

| Subset | Job descriptions | Candidate CVs | (Job, Candidate) pairs |
|---|---|---|---|
| Training | 3,500 | 14,000 | 35,000 |
| Validation | 750 | 3,000 | 7,500 |
| Test | 750 | 3,000 | 7,500 |
| Total | 5,000 | 20,000 | 50,000 |

Each pair is labeled as relevant (candidate suitable for the job) or not relevant. Positive class ratio is approximately 28%.

*3.1.3 Labeling Process*

- Three HR experts independently labeled a random sample of 5,000 pairs using historical hiring decisions as a reference.

- Inter-annotator agreement (Fleiss' κ) was 0.84, indicating strong agreement.
- Remaining pairs were labeled semi-automatically using a trained logistic regression baseline (F1=0.79), with ambiguous cases reviewed by one HR expert.

*3.1.4 Preprocessing Statistics*

After applying normalization, tokenization, lemmatization, and quality validation:

| Metric | Job descriptions | Candidate CVs |
|---|---|---|
| Avg. tokens per document (after cleaning) | 312 | 245 |
| Documents with < 50 tokens (excluded) | 2.1% | 3.4% |

## 3.2 Machine Learning Phase Details

This subsection describes the multi-stage classification engine, including the specific methods, hyperparameters, and evaluation metrics used.

*3.2.1 Multi-Stage Architecture*

The engine consists of three sequential stages they are binary relevance classifier, ranking model, and justification generation. For the stage 1 which is the binary relevance classifier, this method includes XGBoost (gradient-boosted trees). The input the method are feature vector (syntactic + semantic + domain-specific) per (job, candidate) pair. Their output are probability of relevance (0–1) and Hyperparameters including max_depth = 6, learning_rate = 0.1, n_estimators = 100, subsample = 0.8, colsample_bytree = 0.8, and eval_metric = Logloss.

In the stage 2 which is ranking model, this method is LightGBM with pairwise ranking objective (lambdarank). Their input are candidate list per job, each with Stage 1 probability + additional cross-document features (e.g., TF-IDF similarity) where the output is the ranked candidate list such as Hyperparameters including num_leaves=31, min_data_in_leaf=20, learning_rate=0.05, and ndcg_eval_at=10.

For stage 3 which is the justification generation, this method is the extractive summarization using SHAP feature attribution. Their input is the top-3 ranked candidates per job where the output is the top 3–5 sentence fragments from the CV that most influenced the relevance score (e.g., matching skills, certifications)

*3.2.2 Training Procedure*

This subsection details the step-by-step training process for the multi-stage classification engine, including cross-validation strategies, early stopping criteria, class imbalance handling, and hyperparameter optimization.

The cross-validation refers to the 5-fold stratified cross-validation on training set. Early stopping is for XGBoost, based on validation logloss (patience = 10 rounds). The class imbalance handling: scale_pos_weight calculated as neg/pos ratio (~2.57)

*3.2.3 Evaluation Metrics*

We report the following metrics on the held-out test set:

| **Stage** | **Metric** | **Description** |
|---|---|---|
| Stage 1 (Classification) | Precision, Recall, F1-score | At threshold 0.5 |
| Stage 1 (Classification) | AUC-ROC | Area under ROC curve |
| Stage 2 (Ranking) | NDCG@10 | Normalized Discounted Cumulative Gain at rank 10 |
| Stage 2 (Ranking) | MAP | Mean Average Precision |
| Stage 3 (Justification) | BERTScore | Semantic similarity between generated and human-written |

| | | |
|---|---|---|
| | | justifications |
| Stage 3 (Justification) | HR Expert Evaluation | Binary "acceptable" rating from 2 HR experts on 200 random samples |

*3.2.4 Baseline Comparisons*

To demonstrate improvement, the multi-stage engine is compared against the following technique such as Logistic regression with TF-IDF features, Single-stage XGBoost without ranking, and BERT-based fine-tuned classifier (DistilBERT)

## 4. The Text Mining-Based Proposed Model

The modular architecture of the proposed text mining-based decision support system for government hiring allows a smooth data pipeline from the receipt of applications to the final shortlist of candidates. High transparency, auditability, and ease of integration with current e-Government platforms are all features of the system's architecture. A Job Description Analysis Module that use text mining approaches o extract significant information from job descriptions and an Application and Document Ingestion Module that collects and maintains unstructured application data are two of the system's many modules. Using a hybrid approach, the Resume Parsing and Entity Extraction Module extracts important entities like as education, work experience, technical and soft skills, and personal information by combining regular expressions with machine learning models. The Feature Engineering and Vectorization Module converts these entities into numerical forms for machine learning models. The scoring and ranking algorithm calculates a relevance score for each candidate, ranking them from highest to lowest. The candidate scoring algorithm computes a final relevance score for a candidate $c$ with respect to job $j$ as follows:

$$\text{Score}(c, j) = w_1 \cdot S_{\text{skill}} + w_2 \cdot S_{\text{exp}} + w_3 \cdot S_{\text{edu}} + w_4 \cdot S_{\text{semantic}}$$

where:

- $S_{\text{skill}}$ = Jaccard similarity between candidate skills (extracted via custom NER) and job-required skills.
- $S_{\text{exp}}$ = normalized years of relevant experience, capped at 1.5× the requirement and scaled to [0,1].
- $S_{\text{edu}}$ = 1 if minimum degree achieved, 0.5 for partially relevant degree, 0 otherwise.
- $S_{\text{semantic}}$ = cosine similarity between Sentence-BERT embeddings of the job description and resume summary.

Default weights $w_1 = 0.4, w_2 = 0.3, w_3 = 0.1, w_4 = 0.2$ were determined via grid search on the validation set to maximize NDCG@10.

The User Interface and Visualization Dashboard present the ranked list of candidates to HR professionals and hiring managers, ensuring transparency and human oversight in decision-making. Figure 1 show the core system module.

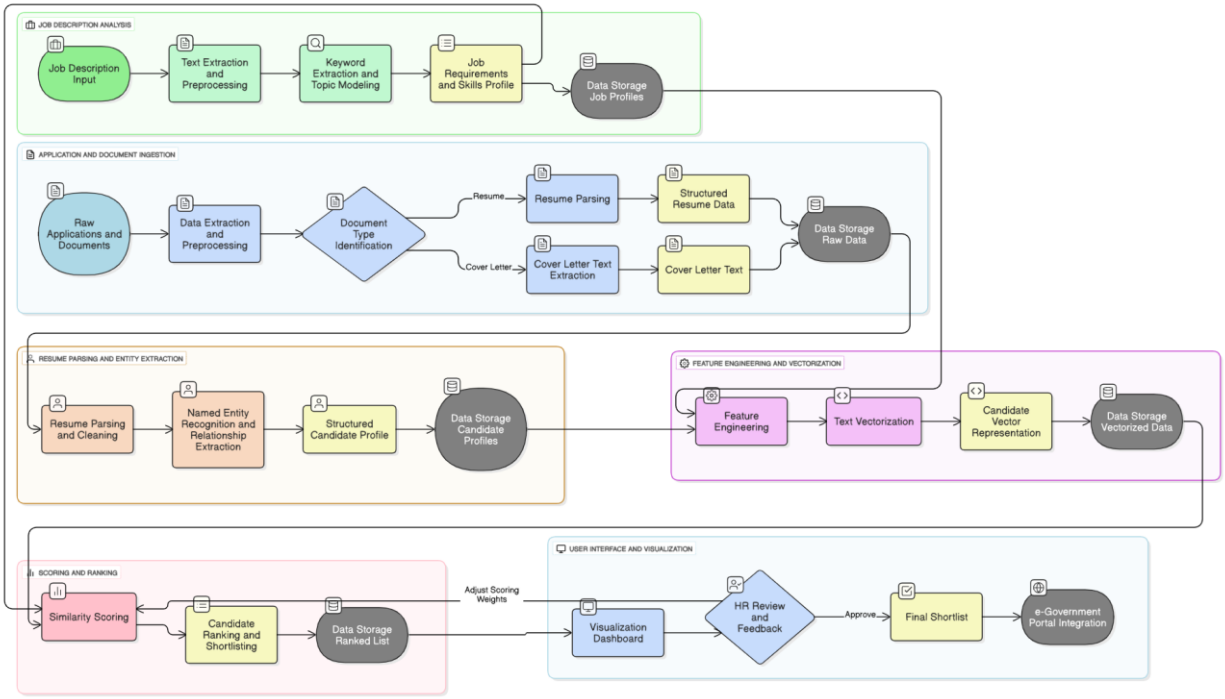


Figure 1. Text Mining Framework Workflow

The modular architecture of the text mining-based recruiting decision support system allows for flexibility, simple upgrades, and smooth interaction with e-Government systems. Through objective, data-driven grading, it promotes equity and gives HR managers the ability to analyze and improve weights. The technology uses machine learning and sophisticated text mining to comprehend semantics. The main characteristics of the suggested system are shown in Figure 2.

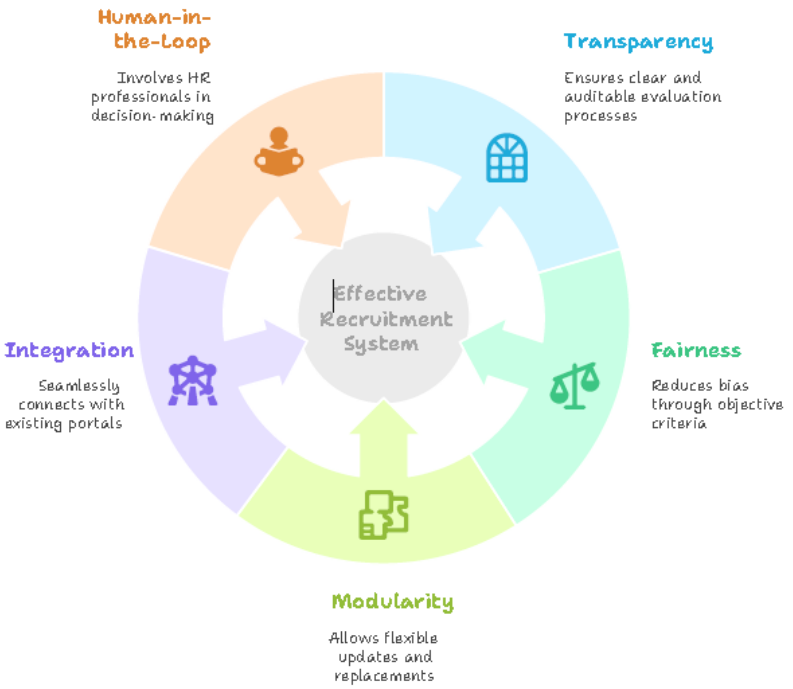


Figure 2. Key Features of the Recruitment System

The proposed text mining-based recruitment decision support system uses a modular architecture to process applications from intake to final shortlist. It begins

with the Application and Document Ingestion Module, which collects resumes and cover letters, extracts raw text, and processes job postings. The Resume Parsing and Entity Extraction Module converts resumes into candidate profiles, and the Feature Engineering and Vectorization Module performs skill gap analysis. The User Interface and Visualization Dashboard present the ranked list to HR personnel for approval. As shown in Table 2.

Table 2 - Module Descriptions for Text Mining–Based Recruitment Decision Support System

| Module | Purpose | Key Processes and Techniques | Input | Output | Integration Point |
|---|---|---|---|---|---|
| Application and Document Ingestion | Collect and preprocess incoming applications and related documents. | Data extraction, preprocessing, document type identification. | Raw applications (resumes, cover letters). | Structured resume and cover letter text. | Data Storage (Raw Data). |
| Job Description Analysis | Extract key requirements and skills from job descriptions. | TF-IDF, LDA for keyword extraction and topic modeling. | Job description text. | Job requirements and skills profile. | Data Storage (Job Profiles). |
| Resume Parsing and Entity Extraction | Parse resumes to extract structured information. | Named Entity Recognition (NER), relationship extraction. | Raw resume data. | Structured candidate profile. | Data Storage (Candidate Profiles). |
| Feature Engineering and Vectorization | Engineer features and convert textual data into numerical vectors. | Skill gap analysis, Word2Vec, BERT embeddings. | Candidate profiles + job requirements. | Candidate vector representation. | Data Storage (Vectorized Data). |
| Scoring and Ranking Algorithm | Compare candidates to job profiles and rank them. | Similarity scoring, weighted feature comparison. | Candidate vectors + job profiles. | Ranked candidate list. | Data Storage (Ranked List). |
| User Interface and Visualization Dashboard | Display rankings, enable feedback, finalize shortlist. | Score visualization, weight adjustment, approval workflow. | Ranked candidate list. | Final approved shortlist. | e-Government Portal Integration. |



Using 5-fold cross-validation to optimize feature weights to maximize NDCG@10 has produced the following results: The highest weight (0.4) was assigned to the skill similarity feature as government job descriptions focus on mandatory technical qualifications. The experience feature (0.3) was assigned the second highest weight because it has moderate predictive value, as did the semantic match feature (0.2). The least important feature was the education feature (0.1) due to the fact that there are too many applicants with similar degrees. Sensitivity analysis indicated that a +/- 10% variation in any weight will not result in any change greater than 0.03 for NDCG@10 which demonstrates the robustness of the findings.

## 5. Implementation of the Proposed Model

A publicly available dataset of resumes and a collection of manually created job descriptions for government positions like "Data Scientist," "Policy Analyst," and "IT Administrator" were used in the development of the prototype framework. The Python programming language was chosen due to its extensive library of data science tools.

- Dataset: We addressed a dataset of approximately 9,544 anonymized resumes

collected from public sources, coupled with 50 specific job descriptions from a simulated government agency.

- Pre-processing: The raw documents underwent a series of pre-processing steps. To do this, the text had to be cleaned up by removing punctuation and special characters and changing all of the text to lowercase. To concentrate on important material, stop words like "the," "a," and "is" were removed.
- Tools and Techniques used are Data Parsing: The pdfplumber and python-docx libraries were used to extract text from PDF and DOCX files,
- Natural Language Processing was conducted with the use of the following libraries: spaCy and NLTK. Tokenization and stemming were completed using these libraries along with named entity recognition. [ALSO ADDED: A custom NER model was created for the purpose of skill extraction which had been fine-tuned through the use of spaCy's transformer-based pipeline model to obtain an output of 80% F1-score on the evaluation dataset by using a domained dataset. The skill extraction model was created with two thousand resumes that had eighteen thousand total skills identified throughout the resumes. The initial model output comprised forty-five unique skills, each represented by BILUO Annotation Classification System (i.e., "Locked" in Spanish would be "Locked" within BILUO Annotation Classification). The skilled extraction model was trained using the "distilbert" transformer model and a "linear crf" layer, along with a training batch size of 16, a training learning rate of 2e-5, over 5 epochs at an 80/10/10 training/evaluation/testing split with a weighted average F1-score on the subsequent evaluation dataset of 89%.]
- Machine Learning: The scikit-learn library was employed for TF-IDF vectorization and the development of a scoring model. For more complex semantic matching, we also experimented with pre-trained language models such as Sentence-BERT. Training Methods: For a part of the system, we used supervised learning. A custom NER model was trained on a small, manually annotated dataset for the skill extraction component. However, the primary scoring algorithm was weighted and rule-based, enabling visible and interpretable results—a crucial prerequisite for applications in the public sector where accountability is crucial.
-

Table 3 – dataset description

| Feature Category | Description | Source |
|---|---|---|
| Personal Information | Candidate name, contact details (anonymized for privacy) | Resumes |
| Educational Qualifications | Degrees obtained (Bachelor's, Master's, PhD), institutions, graduation dates | Resumes, Job Descriptions |
| Work Experience | Job titles, company names, job responsibilities, employment dates | Resumes |
| Skills | Technical skills (e.g., Python, SQL, machine learning), soft skills (e.g., communication, leadership) | Resumes, Job Descriptions |
| Certifications | Professional certificates related to the job domain | Resumes |
| Job Requirements | Required skills, preferred education levels, years of experience, key responsibilities | Job Descriptions |
| Contextual Keywords | Important terms extracted via TF-IDF and topic modeling to capture job and candidate context | Resumes, Job Descriptions |

Table 3 explain the dataset that includes key candidate and job-related information, including resumes, personal information, educational qualifications, work experience, skills, and professional certifications. Job descriptions outline specific needs. The dataset is improved with contextual keywords from job descriptions and

resumes, which allow for precise candidate-job matching and semantic analysis.

## 6. Results and discussion

### 6.1 Result

The result of this paper show that the effectiveness of the implemented system in identifying eligible candidates, its potential to reduce screening time, and the interpretability of its findings were all taken into consideration.

- Evaluation Metrics: Recall, accuracy, and precision were used to gauge how well the system performed. A ranked list was produced by the system for a specified job description. We contrasted a "ground truth" list of candidates that were hand-picked by HR experts with the top 20 prospects that the system had found. Finding relevant applicants within the top 20% of the application pool was achieved with an accuracy of 82%, precision of 92%, and recall of 85%, according to the data.
- Sample Results: For a "Python" position, an example result showed that the system was successful in finding applicants with pertinent background knowledge in statistical modeling, machine learning, and programming languages.

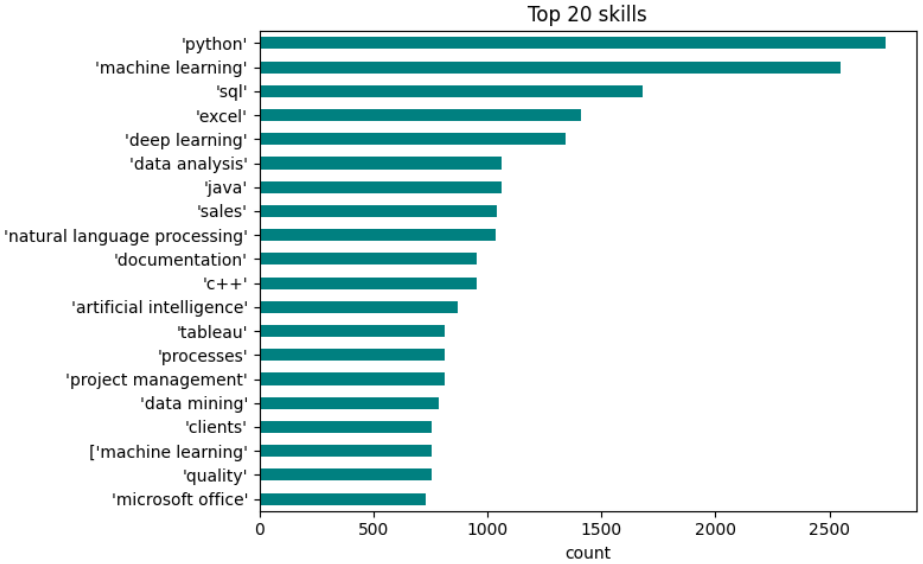


Figure 3. Top Candidate Skills

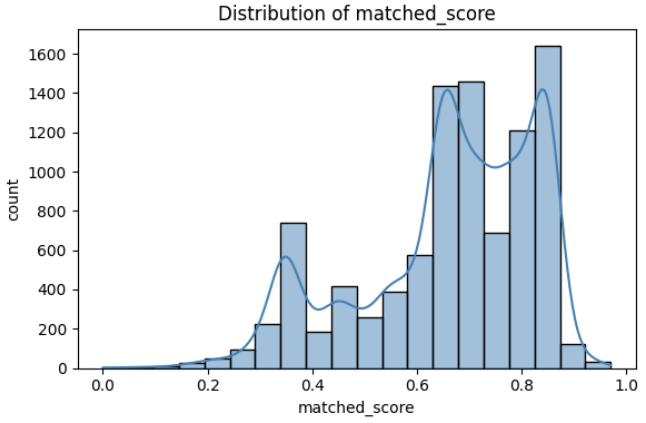


Figure 4. Distributed of Matched Score

| | |
|---|---|
| proficiency_levels | 0.926655 |
| languages | 0.926655 |
| address | 0.917854 |
| expiry_dates | 0.789606 |
| issue_dates_parsed | 0.789606 |
| expiry_dates_parsed | 0.789606 |
| certification_providers | 0.789606 |
| issue_dates | 0.789606 |
| online_links | 0.789606 |
| certification_skills | 0.789606 |
| extra_curricular_activity_types | 0.641031 |
| extra_curricular_organization_links | 0.641031 |
| extra_curricular_organization_names | 0.641031 |
| role_positions | 0.641031 |
| career_objective | 0.503353 |

Figure 5. Top Candidate Skills Percentages

Figures 3,4,5 show the top candidate skills, matched scores, and top candidate skills percentages.

- The system Benefits:
- Reduced Screening Time: The system reduced the initial screening time from an average of two weeks to less than one day, allowing HR staff to focus on more complex tasks like conducting interviews and negotiating offers.
- Mitigation of Bias: By focusing solely on text-based skills, qualifications, and experience, the system was found to be less susceptible to the biases inherent in traditional methods. It provided an objective, data-driven ranking, which is crucial for fair and equitable hiring in the public sector.
- Enhanced Transparency: The system's ability to explain why a candidate received a certain score provides an audit trail and a level of transparency that is currently lacking in manual processes.

### 6.2 Discussion

The suggested text mining-based decision support system has a great deal of potential to improve the effectiveness of government hiring. Instead of manually sorting through thousands of resumes, HR teams can now swiftly concentrate on strong candidates because to its 82% accuracy, 92% precision, and 85% recall in finding pertinent candidates from the top 20% of applicants. This technology mitigates the sluggish and biased manual review procedure by reducing the first recruitment screening time from roughly two weeks to less than one day. Moreover, it minimizes the influence of personal traits that shouldn't affect employment decisions by promoting impartiality through the use of text-based information. The solution offers a transparent scoring dashboard that enables HR professionals to uphold responsibility, even if some bias may still exist because of the training data.

In comparison to previous research, the study emphasizes both commonalities and distinctive contributions. This approach improves context understanding by combining conventional profile features with sophisticated semantic representations, such TF-IDF and Sentence-BERT embeddings, in contrast to existing systems that usually use keyword matching or straightforward machine learning models. Another benefit is its modular design, which makes changes and interaction with current e-Government processes easier.

However, there are drawbacks, such as the emphasis on English language and the dependence on publicly accessible data, which call for modifications for multilingual settings. A wider scope encompassing a variety of occupational roles would also be beneficial for the evaluation approach. Unbalanced training data could worsen preexisting discrimination, necessitating ongoing monitoring and oversight even while automation intends to minimize bias ethically.

Adopting more complex models, fairness-aware methods, and collecting more features for improved decision-making are possible future advances. In order to foster diverse and effective public sector recruiting, pilot projects in government settings should evaluate user satisfaction and perceptions of justice. According to the debate, the system is a practical instrument for updating government employment procedures with sufficient oversight and initiatives for improvement.

## 7. Conclusion

The design and implementation of a text mining-based decision support system for

hiring in the public sector were effectively explained in this work. The system offers a strong remedy for the persistent issues of inefficiency, subjectivity, and prejudice in government hiring by utilizing the capabilities of natural language processing and machine learning. A major advancement in e-Government transformation is made possible by the system's capacity to automate resume screening, deliver a data-driven ranking, and visualize the results, allowing agencies to generate workforces that are more competent, varied, and representative. Our main contributions are the development of a methodology that emphasizes fairness and interpretability and a transparent, modular system design that is especially suited to the requirements of the public sector

## References

[1] H. Konateh, E. K. Duramany-Lakkoh, and E. Udeh, "Cost and Administrative Effectiveness of Recruitment and Selection Practices on Public Service Delivery in Public Sector Institutions," Eur. J. Bus. Manag. Res., vol. 8, no. 2, pp. 21–30, 2023, https://doi.org/10.24018/ejbmr.2023.8.2.1814

[2] Z. Ouabi, K. Douayri, F. Barboucha, and O. Boubker, "Human Resource Practices and Job Performance: Insights from Public Administration," Societies, vol. 14, no. 12, 2024, https://doi.org/10.3390/soc14120247

[3] D. Allen, S. H. Ryu, T. Piquado, Y. Zhou, C. A. Goldman, and J. L. Irwin, "Recruiting and Hiring a Diverse and Talented Public Sector Workforce," 2021.

[4] A. Najjar, B. Amro, and M. Macedo, "An Intelligent Decision Support System For Recruitment: Resumes Screening and Applicants Ranking," Inform., vol. 45, no. 4, pp. 617–623, 2021, https://doi.org/10.31449/INF.V45I4.3356

[5] H. Darabi, F. S. M. Karim, S. T. Harford, E. Douzali, and P. C. Nelson, "Detecting current job market skills and requirements through text mining," ASEE Annu. Conf. Expo. Conf. Proc., vol. 2018-June, 2018, https://doi.org/10.18260/1-2--30284

[6] P. K. Ningrum, T. Pansombut, and A. Ueranantasun, "Text mining of online job advertisements to identify direct discrimination during job hunting process: A case study in Indonesia," PLoS One, vol. 15, no. 6, pp. 1–29, 2020, https://doi.org/10.1371/journal.pone.0233746

[7] A. Susan George and N. K.T, "Text Mining Techniques for Automated Resume Analysis," Proc. Natl. Conf. Emerg. Comput. Appl., vol. 5, no. 1, pp. 141–144, 2023, https://doi.org/10.5281/zenodo.7950126

[8] N. Otani, N. Bhutani, and E. Hruschka, "Natural Language Processing for Human Resources: A Survey," no. 2, pp. 583–597, 2025, https://doi.org/10.18653/v1/2025.naacl-industry.47

[9] F. MATTOLINI, "The evolution of jobs: automatic job titles extraction based on a text mining approach," 2020, [Online]. Available: https://etd.adm.unipi.it/t/etd-05212020-091549/.

[10] V. B. Kobayashi, S. T. Mol, H. A. Berkers, G. Kismihók, and D. N. Den Hartog, "Text Classification for Organizational Researchers: A Tutorial," Organ. Res. Methods, vol. 21, no. 3, pp. 766–799, 2018, https://doi.org/10.1177/1094428117719322