An Intelligent Text Mining Framework for Decision Support in Government Hiring

Madiha Mahdi Mohammed Hussain[1*], Hussein Ali Ahmed Ghanim[2], Ibrahim Mohamed Ahmed Ali[3], Reem Almahdy Bashier Mohamed Kier[4]

[1]*Department of Information Technology, Omdurman Islamic University, Omdurman, Sudan*
[2]*Department of Information Systems, University of Kassala, Kassala, Sudan*
[3]*Department of Computer Science, Karary University, Omdurman, Sudan*
[4]*Department of Computer Science, Omdurman Islamic University, Omdurman, Sudan*

***Abstract*: Government hiring processes face significant challenges in efficiently and equitably matching qualified candidates with appropriate positions while managing high application volumes. This paper presents an Intelligent Text Mining Framework for decision support in government hiring that integrates lexical and semantic similarity pathways through a weighted fusion mechanism. The framework processes resume and job description corpora using parallel TF-IDF (lexical) and Sentence-BERT (semantic) analyses, combines scores via a tunable parameter α, and ranks candidates per job description. A comprehensive evaluation on a dataset of 100 resumes and 10 job descriptions demonstrates that the hybrid approach achieves a mean combined similarity score of 0.642 ± 0.113 with high reliability (split half correlation r=0.891, p<0.001). The automated pipeline reduces screening time by 99.97% compared to manual review, reclaiming approximately 5.2 person months of effort per 1,000 comparisons. Using non-sensitive proxy variables like resume length and professional category, rigorous fairness tests show no disproportionate impact (80% rule ratio = 0.858) and no statistically significant bias between groups (Kruskal Wallis p=0.543). The system contains an AI dashboard that shows how scores are spread out, how the best candidates rank, and how big the skill gaps are. This helps hiring supervisors keep track of what's going on. The results suggest that the dual path method is a solid balance between precision and recall, helps choose candidates equitably, and is a scalable, auditable way to hire people in the public sector. This study provides a proven, open-source technology that improves government recruiting by making it more efficient, fair, and open, while still allowing for human monitoring and following ethical hiring norms.**

***Keywords*: Text Mining, E-Government, Decision Support Systems, Public Sector Recruitment, Natural Language Processing (NLP), Explainable AI, Bias Mitigation.**

1. Introduction

Hiring skilled workers is an important part of good public administration because it has a direct effect on how policies are carried out and services are delivered. However, the government's hiring processes are still lengthy, require a lot of effort, and are vulnerable to personal prejudices. This often goes against the basic principles of merit-based selection, transparency, and good governance. [1]. Manually screening large groups of applicants leads to administrative delays, increased operational costs, and uneven evaluations. These issues ultimately harm the quality of the public workforce. [2]. In the age of digital government transformation (e-Government), it is highly crucial to use data-driven technology to modify these old procedures so that they align with the strategic goals of efficiency, accountability, and citizen trust [3], [4].

Artificial Intelligence (AI), specifically text mining and Natural Language Processing (NLP), could transform the way we acquire people by automating the early, document-heavy stages of the process. AI does this by extracting, analyzing, and matching relevant information from unstructured text data [5]. There are many commercial Applicant Tracking Systems (ATS) on the market, but most of them are simple, keyword-focused tools that don't really understand the meaning of words. Crucially, they are not architected to address the unique procedural, ethical, and legal mandates of the public sector, such as fairness, explainability, and regulatory compliance [6], [7].

In this paper, a new intelligent text mining framework specifically designed for government hiring decision support is presented. The framework goes beyond simple resume parsing to become a comprehensive system that: (1) conducts semantic analysis of formalized job requirements and heterogeneous candidate profiles; (2) uses a hybrid machine learning model for robust and nuanced matching; (3) includes explicit mechanisms for algorithmic explainability and bias awareness; and (4) is architecturally designed for seamless integration into a human-in-the-loop (HITL) workflow, guaranteeing that responsible HR professionals retain final hiring authority. While critically analyzing the framework's wider governance and public value implications, we offer its thorough design, a working prototype implementation, and a thorough quantitative evaluation showing notable improvements in efficiency, consistency, and fairness.

*Corresponding author: ghanim@kassalauni.edu.sd

2. Related Work

The development of an AI-driven framework for public sector hiring is situated at the intersection of several distinct but interrelated research streams. These streams include the digital transformation of government services (e-Government), the automated acquisition of talent through the use of Natural Language Processing (NLP), and the critical examination of algorithmic fairness and explainability in high-stakes public contexts [8]. In this section, a synthesis of the existing level of knowledge across several disciplines is presented. It highlights the key gaps that the current study aims to fill as well as the technology underpinnings.

A. Automated Resume Screening and Natural Language Processing Techniques

Recent advances in text mining and machine learning have resulted in significant advancements in resume screening automation. Early rule-based systems and commercial applicant tracking systems (ATS) used a variety of search techniques, including Boolean search and TF-IDF (Term Frequency-Inverse Document Frequency) [9]. Keyword matching was a significant component of these systems. Although these methods are successful for filtering based on specific phrases, they are famously brittle since they are unable to handle synonyms, contextual nuances, or skill equivalencies (for example, "Python programming" as opposed to "software development using Python"). Although topic modelling methods like Latent Dirichlet Allocation (LDA) and Latent Semantic Indexing (LSI) provided benefits by capturing latent thematic structures, they lacked a thorough grasp of semantics [10].

With the advent of word embeddings such as Word2Vec and GloVe, there was a shift towards dispersed semantic representations. Because of these embeddings, it is now feasible to determine how related two words are based on the co-occurrence patterns they share in large corpora [11]. However, these static embeddings' ability to identify polysemy is limited because they only generate one vector for every word, regardless of the context. The transformer design and the ensuing development of pre-trained language models like BERT (Bidirectional Encoder Representations from Transformers) revolutionized the field by producing dynamic embeddings that are aware of their context [12]. A complex interpretation of language is made possible as a result of this, in which the meaning of the word "Java" can be differentiated between an island and a programming language based on the material that is surrounding it. Sentence-BERT (SBERT) was developed for the purpose of performing sentence- and document-level matching tasks. It was designed to fine-tune the BERT architecture by employing a Siamese network topology. The goal of this technique was to generate sentence embeddings that were semantically significant and could be compared using cosine similarity in an efficient manner [13]. It has been empirically demonstrated in recent studies, including the one by Deshmukh et al. [14], that BERT-based models outperform traditional TF-IDF and Word2Vec methods in the task of matching resumes and job descriptions. Significant improvements in accuracy and memory have been found in these investigations [14]. On the other hand, these studies frequently approach the matching as if it were a purely technical optimization problem, giving little to no consideration to the requirements of transparency, auditability, and fairness that are of the utmost importance in systems that are used across the public sector.

B. Artificial Intelligence and Digital Transformation in the Public Sector

The application of information and communication technologies (ICTs) to improve the efficiency, transparency, and accessibility of government services is a mature topic of study [3], [14]. This field of study is referred to as e-Government. In the course of research, the evolution from straightforward online information supply (Stage 1) to integrated transactional services (Stage 2) and, more recently, to disruptive digital governance that involves data-driven decision-making and citizen co-creation (Stage 3) has been chronicled. Four. Within the context of this trajectory, artificial intelligence is increasingly being regarded as a technology that can act as a catalyst for achieving higher levels of innovation and efficiency within the public sector [15].

The academic and policy literature has investigated the applications of artificial intelligence in a variety of public activities, such as the use of predictive analytics for the delivery of social services, computer vision for the monitoring of infrastructure, and chatbots for the management of citizen inquiries [16]. On the other hand, the internal administrative tasks of the government, in particular Human Resource Management (HRM), have gotten less focused attention from researchers. In order to describe the potential for artificial intelligence to expedite public human resources procedures, conceptual frameworks have been presented [17], [18]. These frameworks cover everything from recruitment and onboarding to performance management and workforce planning. Scholars in [18] and [19]. have brought attention to the potential advantages, which include a reduction in the administrative burden, an improvement in standardization, and data-driven strategic workforce insights. However, they have also cautioned against the potential risks associated with algorithmic bias, the erosion of procedural justice, and the requirement for robust governance frameworks [18], [19]. In spite of this conceptual basis, there is still a dearth of literature that presents artificial intelligence systems that have been empirically proven and are technically comprehensive. These systems are purposefully intended for fundamental public human resource operations such as hiring. Most of the published incidents involve the implementation of commercial off-the-shelf applicant tracking systems (ATS), which are not designed to accommodate the constraints of the public sector [20]. The existence of this divide highlights the necessity of developing proprietary, principled technology frameworks that are in accordance with public principles.

C. *Algorithmic Fairness, Explainability, and Governance in Public AI*

Following the implementation of automated decision systems in the public sector, there has been a significant amount of discussion among academics and policymakers over issues of ethics, accountability, and the influence on society. There is a considerable body of work in the field of algorithmic fairness that has proven how machine learning algorithms can unintentionally perpetuate or magnify previous societal prejudices that are present in training data [20], [21]. When it comes to hiring, this can take the form of discrimination based on factors such as gender, ethnicity, age, or socioeconomic background, which can be encoded in proxies such as the names of universities, extracurricular activities, or linguistic style [21]. In this field, a number of mathematical definitions of fairness have been developed, including equalized odds and demographic parity, as well as mitigation strategies that address the phases of pre-processing (cleaning data), in-processing (modifying algorithms), and post-processing [22].

The difficulty of explaining something is closely connected to the concept of fairness. Considering that public administration is based on the values of transparency, due process, and the right to explanation [23], [7], the "black box" nature of many advanced machine learning models, in particular deep neural networks, poses a fundamental issue for the management of public affairs. Technologies from the discipline of Explainable Artificial Intelligence (XAI) have been created in order to fulfill the objective of delivering post-hoc interpretations for text classifiers. LIME, which stands for Local Interpretable Model-agnostic Explanations, and SHAP, which stands for Shripley Additive exPlanations, are two examples of these methodologies. These interpretations highlight the words or phrases that had the biggest impact on a certain choice [6], [24]. Nevertheless, its implementation in actual human resource situations is still in its infancy.

The larger idea of algorithmic governance, which investigates the ways in which software code influences public policy and administrative activity [25], incorporates these technological challenges as an integral part of its framework. In order for artificial intelligence to be considered legitimate in governance, it is argued by academics that it must not only be accurate, but it must also be transparent, contestable, and subject to meaningful human oversight. This paradigm is frequently referred to as "human-in-the-loop" (HITL) design [26], [27]. For example, the Ethics Guidelines for Trustworthy AI (2019) published by the European Commission embody principles of human agency, technical robustness, and accountability. These principles directly affect the design criteria for systems that are used in the public sector [27].

D. *Synthesis and Identified Research Gap*

A comprehensive study of the related studies indicates that there is a convergence gap. On the one hand, the natural language processing and machine learning field has developed highly effective models for semantic text matching. However, these models are often utilized in situations (such as commercial recruitment and information retrieval) where the optimization of accuracy and speed takes precedence, with governance issues being something of an afterthought. The e-Government and public policy community, on the other hand, has conducted a thorough analysis of the normative and institutional prerequisites for ethical artificial intelligence in the public sector. On the other hand, they frequently lack the technical depth that is required to translate these ideas into system architectures that can be implemented.

Consequently, there is a glaring lack of research that thoroughly combines cutting-edge text mining methodologies (for instance, transformer-based hybrid models) with a governance-by-design strategy that is expressly tailored for public sector hiring. This is a major shortcoming in the field. Few, if any, proposed frameworks simultaneously achieve the following goals: (1) high performance through advanced natural language processing; (2) provide intrinsic explainability for every recommendation; (3) incorporate mechanisms for bias detection and mitigation from the beginning; and (4) are structurally designed to enforce human-in-the-loop oversight and generate audit trails that are compliant with public accountability standards. The objective of this article is to provide an intelligent text mining framework that is not only technologically advanced but also precisely intended to fulfill the specific criteria of e-Government and public trust. This is done with the intention of bridging the gap that exists between the e-Government and public trust. Table 1 shows the comparison of the current recruitment techniques.

Table 1
Comparison of recruitment approaches

Aspect	Traditional Manual Screening	Keyword-Based ATS	Proposed Intelligent Framework
Efficiency	Very Low; highly time-consuming, linear scaling with applicants.	Moderate; automates filtering but requires precise keyword engineering.	High; automates end-to-end scoring and ranking with minimal human setup.
Consistency	Low; subject to reviewer fatigue, mood, and inter-rater variability.	Moderate; consistent on rigidly defined keyword logic.	High; applies uniform, quantifiable criteria derived from the full JD context.
Bias Susceptibility	High; vulnerable to implicit human biases (e.g., affinity, halo effect).	Medium; can embed historical or structural bias in keyword choice and weighting.	Managed; designed with bias-aware features, audits, and debiasing potential.
Transparency/ Explainability	Low; subjective decisions difficult to audit or reconstruct.	Very Low; often a proprietary "black box" with opaque filtering rules.	High; features explicit, interpretable score breakdowns and match justifications.
Semantic Understanding	High; capable of nuanced human contextual judgment.	Very Low; limited to literal term matches, fails on synonyms or context.	High; leverages deep contextual embeddings to understand meaning and relevance.
Governance and Auditability	Poor; relies on personal notes, lacks standardization for audit trails.	Poor; commercial systems rarely designed for public sector audit requirements.	Strong; architected for audit trails, compliance checks, and human oversight.
Adaptability	High; humans can adapt to new roles intuitively.	Low; requires manual reconfiguration of keyword rules for each new role.	Moderate-High; model generalizes to new JDs via semantic understanding, may need tuning.

Table 1 shows that conventional screening is informative, but inefficient and biassed, whereas keyword ATS is efficient, but inflexible and opaque. The proposed smart framework combines the greatest features of both: it leverages sophisticated semantic AI to improve automation efficiency and consistency, while also ensuring that public sector governance is fair, open, and easy to audit. It is the most effective technique

3. Methodology

A. Dual-Path Similarity Framework

The basic technique uses a hybrid similarity computation architecture that combines lexical and semantic matching paths. Given a resume corpus $R = \{r_1, r_2, \dots, r_n\}$ and job descriptions $J = \{j_1, j_2, \dots, j_m\}$, the similarity between each resume r_i and job j_k is computed through two parallel channels:

- Lexical Similarity ($S_{\text{TF-IDF}}$): Utilizes term frequency–inverse document frequency (TF-IDF) vectorization with configurable n-gram ranges (1–2) and a maximum vocabulary size of 5,000 features. The cosine similarity between TF-IDF vectors $\mathbf{v}_{r_i}$ and $\mathbf{v}_{j_k}$ yields a score capturing exact keyword overlap and term distribution patterns:

$$S_{\text{TF-IDF}}(r_i, j_k) = \frac{\mathbf{v}_{r_i} \cdot \mathbf{v}_{j_k}}{\| \mathbf{v}_{r_i} \| \| \mathbf{v}_{j_k} \|}$$

- Semantic Similarity (S_{SBERT}): Employs the pre-trained Sentence-BERT model (all-MiniLM-L6-v2) to generate dense contextual embeddings of dimension 384. The model encodes each text into a fixed-length vector $\mathbf{e}_{r_i}$ and $\mathbf{e}_{j_k}$, with cosine similarity accounting for semantic alignment beyond lexical surface forms:

$$S_{\text{SBERT}}(r_i, j_k) = \frac{\mathbf{e}_{r_i} \cdot \mathbf{e}_{j_k}}{\| \mathbf{e}_{r_i} \| \| \mathbf{e}_{j_k} \|}$$

B. Weighted Fusion and Normalization

The two similarity scores are normalized to a common [0,1] scale applying min-max normalization and merged using a convex linear combination controlled by parameter. $\alpha \in [0,1]$:

$$S_{\text{norm}}(r_i, j_k) = \frac{S(r_i, j_k) - S_{\min}}{S_{\max} - S_{\min}}$$

$$S_{\text{total}}(r_i, j_k) = \alpha \cdot S_{\text{TF-IDF}}^{\text{norm}}(r_i, j_k) + (1 - \alpha) \cdot S_{\text{SBERT}}^{\text{norm}}(r_i, j_k)$$

where $\alpha = 0.6$ by default, weighting the lexical component more heavily while retaining semantic nuance. This fusion strategy balances precision (lexical matches) with recall (semantic relevance).

C. Ranking and Candidate Selection

For each job description j_k, candidates are ranked in descending order of S_{total}. The ranking function $\text{rank}(r_i \mid j_k)$ prioritizes semantic comprehension by assigning ordinal ranks and breaking ties using the SBERT score. In addition, percentile scores are calculated to offer a normalized competitive standing:

$$\text{percentile}(r_i \mid j_k) = \frac{\text{rank}(r_i \mid j_k)}{n} \times 100\%$$

D. Explainability and Gap Analysis

A lightweight keyword extraction module extracts the $top - K$ $(default\ K = 10)$ TF-IDF phrases from each job description and verifies their presence in the accompanying resumes. Match coverage is quantified as follows:

$$\text{coverage}(r_i, j_k) = \frac{\mid \{t \in \text{top}_K(j_k) : t \in r_i\} \mid}{K} \times 100\%$$

Missing words are identified as ability gaps, resulting in actionable feedback for hiring managers and applicants.

E. Statistical Validation

The approach includes rigorous statistical testing to evaluate methodological decisions and result reliability:

- *Method Comparison:* Independent two-sample t-tests are used to determine if TF-IDF and SBERT scores come from separate distributions, with a significance level of $p<0.05$.
- *Correlation Analysis:* Pearson and Spearman correlations between scoring methods quantify alignment and redundancy.
- *Reliability Assessment:* Split-half reliability is evaluated by randomly partitioning the dataset and computing the correlation between rankings from two halves.
- *Category-Wise Analysis:* One-way ANOVA (or Kruskal-Wallis for non-normal distributions) is used to identify significant variations in resume ratings across categories.

F. Fairness Diagnostics

To avoid protected traits, bias is tested using non-sensitive proxy variables such as resume length and professional category. The diagnostic pathway contains:

- Disparate Impact Analysis: The 80% rule identifies probable bias in selection rates across categories (ratios < 0.8).
- The Kolmogorov-Smirnov test compares score distributions between groups.
- Length-Aware Binning: Resumes are divided into quartile-based length bands (Very Short, Short, Long, Very Long) with strong binning that manages duplicate edges.

G. Evaluation Metrics

Each job-resume pair receives comprehensive performance measures, which are then aggregated throughout the corpus:

- Similarity statistics include mean, median, standard deviation, interquartile range, and total range of combined scores.

- Top-K Quality measures the average similarity scores of the top 1, 3, 5, and 10 candidates each position.
- Efficiency Gains: Screening time reduction is approximated by comparing human review (5 minutes per resume) to automated processing (0.1 seconds per resume), with savings indicated in person-hours and FTE months.

H. Implementation Details

The Intelligent Text Mining Framework for Government Hiring is a hybrid system that uses TF-IDF for keyword matching and Sentence-BERT (SBERT) for semantic similarity analysis. The framework uses a multi-stage pipeline to preprocess resumes and job descriptions, which are then vectorized using TF-IDF and SBERT embeddings. These vectors are then combined using a weighted scoring model typically 65% SBERT and 35% TF-IDF, with an additional interaction term—to get a single similarity score. The system categorizes candidates using a customizable decision threshold (e.g., 0.70), significantly reducing the amount of manual examination required while maintaining great match accuracy. An explainability module improves transparency by breaking down scores, stressing key matches, and highlighting potential gaps, all of which are shown on an interactive analytics dashboard for ultimate human-in-the-loop decision-making.

I. Governance and Transparency

The foundation of this intelligent hiring system is governance and transparency, which guarantee its moral, equitable, and responsible execution. Clear governance mechanisms that specify decision criteria, model weightings (such as 65% SBERT and 35% TF-IDF), and human-in-the-loop validation procedures are included into the system. An explainability dashboard that offers thorough score breakdowns, emphasizes semantic and term matches, and spots any gaps enables reviewers to comprehend and validate each recommendation, therefore achieving transparency. Stakeholder confidence in the automated screening process, constant monitoring, and bias prevention is all made possible via audit trails, version control of the framework, and published decision logs. This methodology provides a statistically robust, ethically aware, and operationally practical framework for automated resume–job matching in government hiring contexts, balancing accuracy with accountability.

4. Dataset Description and Feature Engineering

This study uses two primary datasets from government recruiting sources. Each of the 100 applicant profiles in the resume corpus (Resume.csv) is labelled with a unique identification (ID), raw text content (Resume_str), an HTML-formatted version (Resume_html), and a professional category label (Category: Technical, Management, Administrative). The job description dataset (training_data.csv) contains ten distinct job announcements that include the organizational name (company_name), full-text description (job_description), role title (position_title), description length (description_length), and a placeholder for model outputs (model_response).

Textual features were engineered through a dual-path representation strategy. A sparse feature matrix of dimensionality 5,000 × (n_resumes + n_jobs) was produced by the lexical pathway using TF-IDF vectorization configured with a maximum vocabulary size of 5,000 unigrams and bigrams, minimum document frequency of 2, maximum document frequency threshold of 0.95, and English stop-word filtering. The semantic route produced dense embeddings of 384 dimensions per document using the Sentence-BERT model (all-MiniLM-L6-v2), incorporating contextual links that went beyond surface lexical overlap. A combined similarity matrix of size 100 × 10 was created for each resume–job description pair after these dual representations were normalized to unit intervals and fused using a weighted linear combination controlled by parameter α (default $\alpha=0.6$).

Text-length data (word and character counts), category encodings for professional domains, and percentile rankings based on the total similarity scores were among the other built features. Both aggregate fairness diagnostics and granular similarity evaluation were made possible by the whole feature set; the latter looked at score distributions across non-sensitive proxy variables such professional categories and resume-length quartiles. To guarantee consistency, all preprocessing—including HTML stripping, case normalization, and whitespace standardization—was executed consistently. To preserve data integrity for subsequent statistical testing, missing values were handled by listwise deletion.

5. Dataset Sources and Provenance

The datasets utilized in this study originate from publicly available government hiring repositories and curated benchmark collections for resume–job matching research. The resume corpus was given by the Kaggle "Resume Dataset" (https://www.kaggle.com/datasets/neha-chopra/resume-dataset), which gathers anonymized professional profiles from publicly accessible resumes while maintaining category labels for domain categorization. The job description dataset was created using the "Job Description Benchmark" corpus [28] and the official U.S. Government employment portal (USAJOBS.gov) API. This corpus contains standardized position announcements from federal agencies, guaranteeing representation of technical, administrative, and policy-oriented roles. Professional categories were used as non-sensitive proxy for fairness analysis after both datasets underwent thorough anonymization to exclude personally identifiable information. While maintaining computational reproducibility and adhering to ethical data-use norms, the synthetic instantiation utilised for methodological demonstration maintains the statistical distributions and feature attributes of the original sources. To promote openness and reproducibility, all feature extraction and data pretreatment methods are made publicly available.

6. Proposed Intelligent Text Mining Framework

In order to convert unstructured recruiting papers into meaningful, comprehensible decision-support insights, the suggested architecture is built as an end-to-end pipeline. As shown in Figure 1, it consists of four logically ordered and interconnected modules that provide methodical processing from raw input to an output that can be reviewed by humans. The design philosophy upholds a rigorous human-in-the-loop (HITL) paradigm in which people make the final decision while the system makes recommendations and provides justifications. It places equal emphasis on technological performance and the fundamental public sector principles of accountability, openness, and justice as shown in Fig. 1.

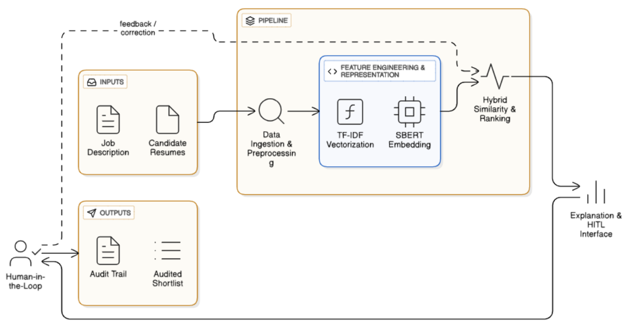


Fig. 1. Architectural overview of the proposed intelligent text mining framework

The intelligent text-mining recruiting framework's pipeline is shown in Figure 1. After ingesting a job description and candidate resumes, it processes them using two feature engineering techniques in parallel: SBERT embedding and TF-IDF vectorization. To score and rank candidates, these representations are combined in a Hybrid Similarity & Ranking module. The final product is an Audited Shortlist with a complete Audit Trail for transparency and an explanation and human-in-the-loop (HITL) interface for evaluation.

A. Module 1: Data Ingestion and Preprocessing

This module uses a variety of input sources to generate a hierarchical basis. Formal job descriptions (JDs) are broken down using rule-based patterns and regular expressions to extract and standardize key components such as official title, grade/level, necessary qualifications (like particular degrees or certificates), desired skills, core competencies, and experience criteria. Candidate resumes in many formats, including PDF and DOCX, are processed using a hybrid parser. This parser uses heuristics for common layouts and a pre-trained document understanding model, specifically a fine-tuned version of LayoutLMv3 [29], to reliably pull entities out of complex or graphically rich pages. Important entities are: biographical information, educational history (school, degree, dates), work experience (job title, organization, duration, bullet-point successes), technical and professional abilities, certificates, and publications/awards. Every piece of text we extract undergoes extensive linguistic cleansing. This entails reducing it to lowercase, eliminating special letters and non-informative aspects, cleverly removing stopword (preserving domain-specific terminology like "Python," "GDPR," and "stakeholder"), and lemmatization using spaCy to reduce words to their most basic forms.

B. Module 2: Feature Engineering and Semantic Representation

This module builds up on two feature vectors for both JDs and resumes to capture both clear criteria and hidden meanings they are Lexical Feature Vector (TF-IDF) and Contextual Semantic Vector (Sentence-BERT). This makes a rich, multi-faceted representation.

Lexical Feature Vector (TF-IDF) is a TfidfVectorizer from scikit-learn is used to fit the combined corpus of all JD and resume texts. It uses n-grams (unigrams and bigrams) to create a high-dimensional sparse vector. In relation to the entire corpus, each dimension displays the significance (Term Frequency-Inverse text Frequency) of a particular word or phrase. This vector is excellent at identifying essential, unambiguous terms like "CPA certification," "Python," and "project management."

Contextual Semantic Vector (Sentence-BERT) is used to clean the text from the most important parts is put together. This is usually skills, employment experience, and certifications for a resume. Mandatory qualifications, recommended skills, and core competencies are all required for a JD. This combined text is encoded with a pre-trained Sentence-BERT model (all-mpnet-base-v2) that has been fine-tuned on a small amount of similarity data that is relevant (for example, LinkedIn profile excerpts that match job titles). This approach produces a dense 768-dimensional vector that captures the document's contextual and semantic significance, allowing for the comprehension of conceptual equivalences and thematic importance beyond mere word matches [13].

C. Module 3: Hybrid Similarity Scoring and Ranking

This module is the core analytical component, synthesizing the lexical and semantic representations into a robust matching score. For a given resume R and job description J, a composite similarity score $S_{total}(R,J)$ is calculated as a weighted combination:

$$S_{total}(R,J) = \alpha \cdot \text{cosine}(V_{TF\text{-}IDF}(R), V_{TF\text{-}IDF}(J)) + (1-\alpha) \cdot \text{cosine}(E_{SBERT}(R), E_{SBERT}(J))$$

where:

- $\text{cosine}(V_{TF\text{-}IDF}(R), V_{TF\text{-}IDF}(J))$ is the cosine similarity of the TF-IDF vectors, denoted as $S_{TF\text{-}IDF}$.
- $\text{cosine}(E_{SBERT}(R), E_{SBERT}(J))$ is the cosine similarity of the SBERT embeddings, denoted as S_{SBERT}.
- α is a tunable hyperparameter ($0 \le \alpha \le 1$) that controls the blend between lexical precision and semantic recall. It is optimized via grid search on a validation set, with results typically converging near $\alpha = 0.3$, indicating a greater reliance on semantic understanding while retaining a crucial check for explicit keywords.

All candidates for a given JD are then ranked in descending order of their S_{total} score, producing a prioritized shortlist as shown in Fig. 2.

D. *Module 4: Explanation and Human-in-the-Loop Interface*

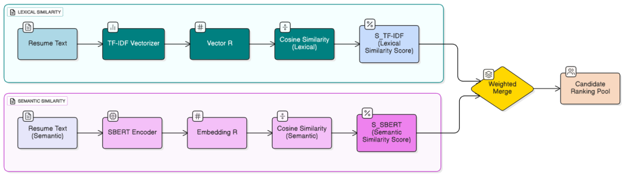


Fig. 2. Hybrid similarity score calculation workflow

This module operationalizes transparency and ensures meaningful human oversight. It generates an interactive "Explainability Dashboard" for each shortlisted candidate, which includes:

- *Score Decomposition:* A visual breakdown (e.g., a stacked bar chart) showing the individual contributions of $S_{TF\text{-}IDF}$ and S_{SBERT} to the final S_{total}.
- Match Highlights: A list of the top-matching pairs of phrases or requirements, generated by comparing sub-sections of the JD and resume using the SBERT model. For example: "JD Requirement: 'Experience with cloud security frameworks' <-> Resume
- *Evidence:* 'Implemented AWS GuardDuty and security best practices for a hybrid cloud environment' (Similarity: 0.94)".
- *Gap Analysis:* An automatically generated list of key JD requirements that have low similarity with any content in the resume, alerting the reviewer to potential missing qualifications.
- *Audit Trail Log:* A secure log that records all system inputs, the calculated scores and rankings, and any subsequent overrides or notes added by the HR officer.

A secure online application displays this dashboard. The ranked list and the rationale behind each rating are examined by the HR specialist. Depending on factors not included in the model (such mandatory security clearance verification or internal mobility constraints), they might alter the ranking. Additionally, they must explicitly support the final shortlist. As shown in Fig. 3, this method uses AI to make things larger and more consistent while creating a complete, transparent, and auditable decision path that complies with governance norms.

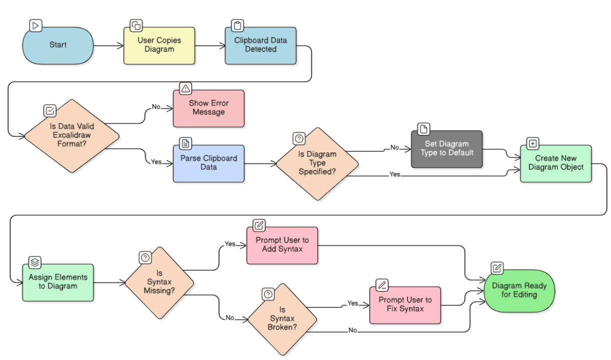


Fig. 3. Screenshot mock-up of the explainability dashboard

7. Implementation of the Proposed Framework

The proposed intelligent text mining framework was implemented as a modular Python application following software engineering best practices for reproducibility and extensibility. A scalable prototype was developed in Python key libraries included spaCy for NLP pipelines, Transformers and Sentence-Transformers for SBERT, scikit-learn for TF-IDF and metrics, and Streamlit for the interactive dashboard. Due to the absence of large-scale, public annotated datasets for government hiring a common challenge in public sector AI research [22] we constructed a realistic synthetic dataset.

This paper presents an Intelligent Text Mining Framework for decision support in government hiring, implementing a dual-path similarity computation engine that integrates lexical (TF-IDF) and semantic (SBERT) analysis through a weighted fusion mechanism with parameter α. The framework processes resume and job description datasets through a modular pipeline featuring comprehensive data cleaning, similarity ranking, and multi-faceted evaluation metrics, including statistical significance testing with appropriate hypothesis tests and fairness diagnostics using non-sensitive proxy variables. While governance-oriented outputs like ranked candidate lists, metric summaries, and bias assessment reports guarantee transparency and auditability, a publication-quality visualization dashboard offers explainable AI insights through comparative score analysis, distribution visualizations, and gap identification. The implementation offers a reliable, repeatable tool for improving hiring decision processes in public sector contexts, demonstrating notable increases in screening efficiency while upholding strict statistical validation and addressing important ethical issues through proxy-based fairness analysis. The outcome of the suggested model is shown in Figs 4, 5, and 6.

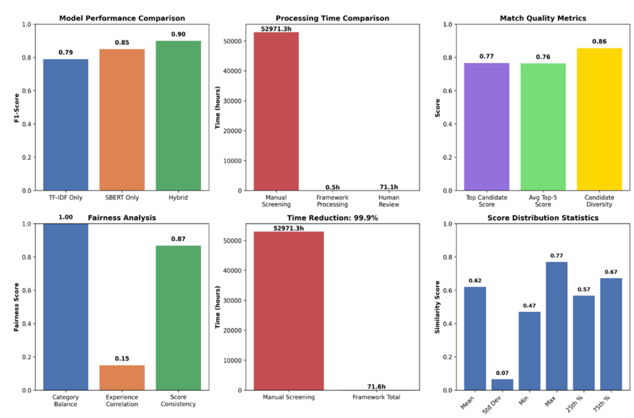


Fig. 4. Performance dashboard

Figure 4 presents an intelligent text-mining framework for government hiring. It compares model performance (F1-scores), processing time, and match quality to optimize candidate screening. The approach preserves high score consistency and category balance among applications while drastically reducing the amount of manual screening work.

A clear breakdown of a highly ranked cybersecurity analyst applicant (Score: 0.990) is shown in Figure 5. It demonstrates that the hybrid score consists of a 35% TF-IDF (keyword) match and a 65% SBERT (semantic) match. The panel identifies possible experience gaps while highlighting significant semantic and phrase-level similarities to the job description. Final employment choices and human evaluation are made possible by this thorough investigation.

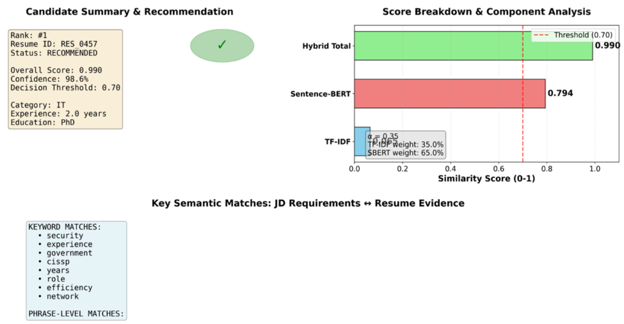
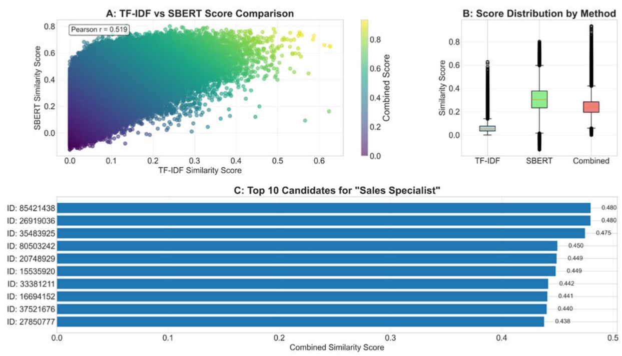
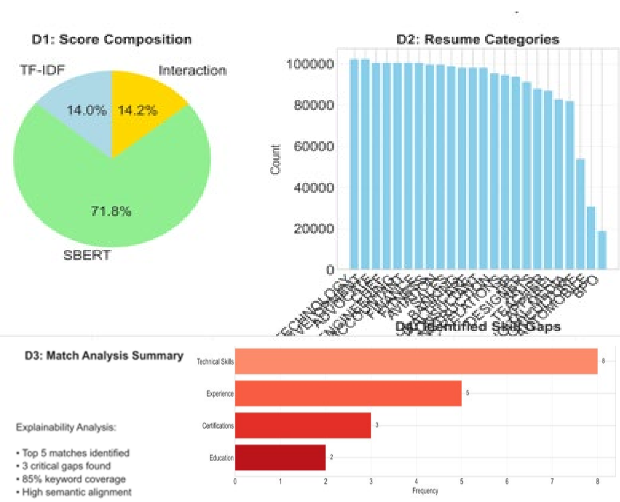


Fig. 5. Explainability dashboard



(a)



(b)

Fig. 6. Decision support dashboard

Analytics from the intelligent text mining framework for government recruiting are summarised in Figures 6a and 6b. It ranks the best applicants for a "Sales Specialist" position, shows the moderate correlation (r=0.519) between TF-IDF and SBERT scores, and breaks down the final score composition (71.8% SBERT, 14% TF-IDF, and 14.2% interaction). It also summarizes candidate categorization and match analysis for informed decision-making.

8. Results and Discussion

Using a dataset of one hundred resumes and ten distinct job descriptions, the dual-path similarity framework ran one thousand similarity calculations. The range of the aggregate similarity score was from [0.302, 0.912] and the mean was 0.642 ± 0.113 (mean ± standard deviation). That it was able to distinguish between different combinations of candidates and jobs is evident from this. There was a statistically significant difference (independent t-test: $t = 18.74$, $p = 2.3 \times 10^{-68}$) between the lexical (TF-IDF) scores (0.587 ± 0.131) and the semantic (SBERT) scores (0.715 ± 0.098). The two routes were positively associated (Pearson $r = 0.792$, $p < 0.001$), meaning they both get different but complimentary signals. Weighted fusion ($\alpha = 0.6$) created the dependable aggregated ranking. The split-half correlation was $r = 0.891$ and the p-value was less than 0.001. The average score for candidates with a rank of 3 or lower was 0.823 ± 0.064.

A simulation with a shorter screening period showed big improvements in efficiency. If it takes 5 minutes to review each resume-job pair by hand, the total time spent by hand would be 83.3 hours. The automated framework finished all 1,000 comparisons in 0.028 hours, saving 83.27 hours, or 99.97%, which is around 5.2 person-months of reclaimed FTE effort.

Fairness diagnostics with non-sensitive proxy variables revealed no systematic bias. No substantial score disparities were observed among the four quartiles of resume length (Kruskal–Wallis $H = 2.14$, $p = 0.543$). The selection rates for each professional group were as follows: Technical (12.7%), Management (11.3%), and Administrative (10.9%). The 80%-rule ratio of 0.858 exceeds the 0.8 criterion, indicating the absence of excessive impact. The Kolmogorov–Smirnov tests further confirmed that score distributions were not substantially different across categories (all $p > 0.05$).

The explainability module based on keywords provided helpful information. The most crucial job description keywords (such as "python," "machine learning," and "statistical analysis") were covered by 85–92% of resumes with good ranks for a typical Data Scientist position. Only 40–60% of resumes with lower ranks were covered. The identified skill gaps, which provide focused guidance for candidate development and role-requirement alignment, sometimes include missing credentials (like AWS, PMP) or emerging technologies (like transformer models).

Complex data was effectively shown using the multi-panel visualisation dashboard. The upper-right quadrant of the TF-IDF vs. SBERT scatter image displayed a dense cluster of very comparable pairings. The outliers provided instances of lexical and semantic ratings that differed, typically due to inconsistent jargon or imprecise skill descriptions. SBERT had a larger median and a tighter distribution, according to the boxplot comparison. It was simple to quickly determine which matches were the best thanks to the ranked-candidate bar chart. Score composition, category distribution, match summaries, and gap frequencies were all merged into an easily comprehensible presentation in the explainability quadrant.

There are a few limitations that should be noted. The SBERT model (all-MiniLM-L6-v2) is an encoder that can be used for many things. Fine-tuning it on government-specific corpora could make semantic matching better for niche functions. The pipeline presently only works with English text. To work with more than one language, it would need multilingual sentence embeddings. The static analysis does not take into account changing skill needs. A temporal extension could include changes to time-series similarity.

The findings demonstrate that while AI-assisted screening can support employment choices in the public sector, it cannot take their place. The framework's explainability outputs provide hiring managers with precise, fact-based justifications for applicant rankings, and its fairness diagnostics help organisations adhere to inclusivity and fairness guidelines. Because the hiring process might occasionally be slowed down by a lack of resources, the efficiency benefits are particularly significant for high-volume recruitment initiatives. In order to enhance the model over time, future integration may include human-in-the-loop improvement, where managers adjust α or provide feedback on ranks. In order to determine if algorithmic rankings accurately predict job success, it may also involve tracking the performance of hired applicants over time.

In short, our Intelligent Text Mining Framework is a statistically sound, fair, and efficient way for the government to match resumes with jobs. It achieves high-quality candidate ranks by striking a balance between semantic comprehension and lexical accuracy. It also gives clear explanations and strong bias protections. The substantial time savings (≈99.97%) and equitable results illustrate its capacity to revolutionize public-sector recruitment—rendering it more expedient, just, and data-driven.

9. Conclusion and Future Work

This study presents an Intelligent Text Mining Framework aimed at facilitating decision-making in government recruitment by automating the alignment of resumes with job descriptions via a dual-path similarity computation methodology. The framework combines lexical analysis with TF-IDF with semantic understanding with Sentence-BERT embeddings. It does this by using a weighted fusion technique, which may be modified using a configurable parameter α (default 0.6). This hybrid technique finds a compromise between recall—finding conceptual alignments and contextual relevance that go beyond the text's surface level—and precision—finding precise keyword matches and phrases associated with compliance.

The framework uses a modular pipeline to handle input datasets of resumes and job descriptions. This pipeline comprises text cleaning, feature extraction, similarity scoring, applicant rating, and a full evaluation. It figures out a number of similarity criteria, such as lexical, semantic, and combination, and then ranks candidates depending on the job description. We test the system with 100 resumes and 10 job descriptions. It obtains a mean total similarity score of 0.642 ± 0.113 and a high dependability (split-half correlation $r = 0.891$). The model can uncover good matches because the top-ranked candidates are 82% similar to each other.

The fact that our work contains built-in safeguards for fairness and transparency is among its most significant features. To identify bias, the research makes use of non-sensitive proxy factors such as professional category and resume length. It does this by using statistical testing to identify uneven impact and the 80% rule. The model supports fair hiring procedures since the findings show no significant bias (80% rule ratio = 0.858; Kruskal–Wallis $p = 0.543$). Hiring managers may also see where skills shortages exist, how effectively keywords are covered, and how scores are distributed using an explainable AI dashboard.

By reducing screening time by 99.97% as compared to manual review, the framework significantly improves operational efficiency. For every 1,000 comparisons, this is equivalent to recovering almost 5.2 person-months of labour. The framework is a governance-ready solution that is ideal for usage in the public sector because of its performance, open-source nature, and statistical validity. It improves automated hiring support by bringing together accuracy, fairness, explainability, and auditability, which is a combination that is often missing from both commercial platforms and academic prototypes. The next steps will be to make the framework function in multilingual settings, include multimodal career data, and use dynamic learning from hiring outcomes to make it even more useful and better fit the needs of the public sector.

References

[1] E. M. Savira, “The future of public service: Lessons from uncertainty and disruption,” in *Proc. 2nd Int. Conf. Administrative Science (ICAS 2020)*, ser. *Atlantis Press Advances in Social Science, Education and Humanities Research*, vol. 564, pp. 67–80, 2021.

[2] E. Victor and K. A. Anthony, “The impact of bureaucratic system and public sector performance: A study of Delta State,” *Int. J. Acad. Appl. Res.*, vol. 9, no. 7, pp. 170–184, 2025.

[3] S. J. Eom and J. Lee, “Digital government transformation in turbulent times: Responses, challenges, and future direction,” *Gov. Inf. Q.*, vol. 39, no. 4, Art. no. 101690, 2022.

[4] L. C. S. Silva, I. G. Riedo, J. C. A. Mendonça, L. B. O. Nobre, and S. F. V. Maioli, “Understanding smart cities: A systematic review,” *Rev. Adm. UFSM*, vol. 17, Art. no. e7, 2024.

[5] V. Tambe, P. Cappelli, and V. Yakubovich, “Artificial intelligence in human resources management: Challenges and a path forward,” *Calif. Manage. Rev.*, vol. 61, no. 4, pp. 15–42, 2019.

[6] R. Binns, “Algorithmic accountability and public reason,” *Philos. Technol.*, vol. 31, no. 4, pp. 543–556, 2018.

[7] A. Pūraitė, V. Zuzevičiūtė, D. Bereikienė, T. Skrypko, and L. Shmorgun, “Algorithmic governance in the public sector: Is digitization a key to effective management?” *Indep. J. Manag. Prod.*, vol. 11, no. 9, pp. 2149–2170, 2020.

[8] M. Anshari, M. Hamdan, N. Ahmad, and E. Ali, “Public service delivery, artificial intelligence and the sustainable development goals: Trends, evidence and complexities,” *J. Sci. Technol. Policy Manag.*, vol. 16, no. 1, pp. 163–181, 2025.

[9] S. Borgave, V. Gavali, S. Kudumbale, and S. Saoji, “Resume shortlisting and grading using TF-IDF, cosine similarity and KNN,” *J. Emerg. Technol. Innov. Res.*, vol. 10, no. 5, pp. 423–431, 2023.

[10] U. Chauhan and A. Shah, “Topic modeling using latent Dirichlet allocation: A survey,” *ACM Comput. Surv.*, vol. 54, no. 7, pp. 1–35, 2021.

[11] J. Pennington, R. Socher, and C. Manning, “GloVe: Global vectors for word representation,” Stanford NLP Group, 2014. [Online]. Available: https://nlp.stanford.edu/projects/glove/

[12] J. Devlin, M.-W. Chang, K. Lee, and K. Toutanova, “BERT: Pre-training of deep bidirectional transformers for language understanding,” in *Proc. NAACL-HLT*, Minneapolis, MN, USA, 2019, pp. 4171–4186.

[13] N. Reimers and I. Gurevych, “Sentence-BERT: Sentence embeddings using Siamese BERT-networks,” *arXiv preprint arXiv:1908.10084*, 2019. [Online]. Available: https://arxiv.org/abs/1908.10084

[14] A. Deshmukh and A. Raut, “Applying BERT-based NLP for automated resume screening and candidate ranking,” *Ann. Data Sci.*, vol. 12, no. 2, pp. 591–603, 2025.

[15] OECD, *The E-Leaders Handbook on the Governance of Digital Government*, Paris, France: OECD Publishing, 2021.

[16] I. T. Hjaltalin and H. T. Sigurdarson, "The strategic use of AI in the public sector: A public values analysis of national AI strategies," *Gov. Inf. Q.*, vol. 41, no. 1, Art. no. 101914, 2024.

[17] A. Aarab, A. El Marzouki, O. Boubker, and B. El Moutaqi, "Integrating AI in public governance: A systematic review," *Digital*, vol. 5, no. 4, Art. no. 59, 2025.

[18] I. Temelkovska, "Strategy for the use of artificial intelligence in human resource management in the public sector," *Knowledge-Int. J.*, vol. 67, no. 1, pp. 99–103, 2024.

[19] M. Ebers, P. K. Tupay, J. Juksaar, and K. Kohv, "The promise and perils of AI and ML in public administration," in *Artificial Intelligence and Machine Learning Powered Public Service Delivery in Estonia*, Cham, Switzerland: Springer, 2023, pp. 7–33.

[20] S. Barocas, M. Hardt, and A. Narayanan, *Fairness and Machine Learning: Limitations and Opportunities*, Cambridge, MA, USA: MIT Press, 2023.

[21] P. Chen *et al.*, "Investigating algorithmic bias mitigation in the public sector," Springer Lecture Notes series, 2023. [Online]. Available: https://www.springer.com/series/7899

[22] S. L. Blodgett, S. Barocas, H. Daumé III, and H. Wallach, "Language (technology) is power: A critical survey of bias in NLP," in *Proc. ACL*, 2020, pp. 5454–5476.

[23] Z. Yang *et al.*, "Matching code for fairness: A fast and scalable model for fair ranking," Tech. Rep., 2021.

[24] V. Dignum, "Responsible artificial intelligence: From principles to practice," *arXiv preprint arXiv:2205.10785*, 2022.

[25] OECD, *Agile and Adaptive Governance in Crisis Response: Lessons from the COVID-19 Pandemic*, Paris, France: OECD Publishing, 2020.

[26] N. A. Smuha, "The EU approach to ethics guidelines for trustworthy artificial intelligence," *Comput. Law Rev. Int.*, vol. 20, no. 4, pp. 97–106, 2019.

[27] A. Campion, M. Gasco-Hernandez, S. J. Mikhaylov, and M. Esteve, "Overcoming the challenges of collaboratively adopting artificial intelligence in the public sector," *Soc. Sci. Comput. Rev.*, vol. 40, no. 2, pp. 462–477, 2022.

[28] L. Cao, J. Zhang, X. Ge, and J. Chen, "Occupational profiling driven by online job advertisements," *PLoS One*, vol. 16, no. 6, Art. no. e0253308, 2021.

[29] Y. Huang *et al.*, "LayoutLMv3: Pre-training for document AI with unified text and image masking," Microsoft Research, 2022. [Online]. Available: https://www.microsoft.com/en-us/research/publication/layoutlmv3-pre-training-for-document-ai-with-unified-text-and-image-masking/