



SCHOOL OF COMPUTING AND INFORMATION TECHNOLOGY

REPORT ON JOB RECOMMENDATION SYSTEM

Submitted in partial fulfilment of the requirements for the award of the Degree of

BACHELOR OF TECHNOLOGY IN INFORMATION SCIENCE AND TECHNOLOGY

Submitted by

Gokul M R23EQ033

Kethan Ram Amara R23EQ052

Jeevan C R23EQ044

Under the guidance of

Prof. Bhavana N

Professor, School of C&IT

Project Proposal: Enhancing the ML Candidate Ranking System for Production Readiness

1.0 Introduction: Building on a Successful Foundation

The initial Machine Learning Candidate Ranking System project successfully met its core objectives, delivering a functional, lightweight, and interpretable prototype. This initial version demonstrated the viability of a hybrid approach that combines classical NLP with modern transformer-based embeddings to rank candidates effectively without relying on cloud infrastructure or paid services. This proposal outlines the next logical phase of development, which aims to evolve this successful prototype into a robust, fair, and scalable system ready for real-world application.

Recap of the Current System

The current system's success is built on a practical and efficient architecture. Its key achievements include:

- **Hybrid Scoring Model:** The system employs a weighted formula, $\text{Final Score} = 0.35 \cdot \text{TFIDF} + 0.40 \cdot \text{EmbeddingSim} + 0.25 \cdot \text{LRProbability}$, that blends three distinct signals: lexical alignment (TF-IDF), semantic similarity (Embedding Similarity), and experience-based relevance (Logistic Regression).
- **Lightweight & Efficient:** By utilizing TF-IDF and a compact DistilBERT model (approximately 66MB), the system ensures rapid execution in under 10 seconds and maintains low resource usage, operating entirely offline without cloud dependencies.
- **Interpretability:** A core design principle is transparency. The system outputs ranked lists accompanied by explanations for each candidate's score, a crucial feature for any tool intended to support human evaluation in contexts like hiring.
- **Prototype Success:** The project has successfully proven a practical methodology for candidate ranking, making it an ideal foundation for development in academic or startup environments.

While the prototype stands as a significant accomplishment, strategic enhancements are now required to unlock its full potential and prepare it for enterprise-level deployment.

2.0 The Strategic Imperative: Moving from Prototype to Production

The strategic importance of this next phase lies in elevating the system from a functional proof-of-concept to a production-ready asset. Moving beyond the prototype stage requires us to address key enterprise-level concerns, including fairness, scalability, and seamless integration into existing workflows. By tackling the current system's inherent limitations, we can deliver a tool that is not only powerful but also trustworthy and efficient at scale.

The following table outlines the current limitations and frames them as high-value opportunities for enhancement.

Current Limitation	Proposed Opportunity
Bias Potential: The current model includes an explicit bias towards experience (≥ 3 years) and lacks formal fairness checks.	Enhance Trust & Fairness: Implement systematic fairness checks to detect and mitigate bias related to sensitive attributes, increasing system trust and utility in HR contexts.
Basic Explanations: Interpretability is a goal, but explanations are based on raw component scores.	Deeper Semantic Insight: Evolve from simple scores to richer, more nuanced insights using graph-based embeddings and LLM-powered summary explanations.
Individual Processing: The system is designed to rank a small batch of candidates for a single job description.	Improve Efficiency at Scale: Introduce job-candidate clustering to efficiently manage and analyze large talent pools, enabling broader talent discovery.
Standalone Functionality: The system operates as a standalone pipeline, limiting its integration into other workflows.	Enable Enterprise Integration: Deploy the system as a scalable API (using FastAPI or Flask) to allow seamless integration with existing HR software and internal tools.

This proposal details a specific, phased roadmap designed to methodically realize these opportunities and mature the system into an enterprise-grade solution.

3.0 Proposed Enhancements: A Phased Roadmap

The proposed work is structured as a two-phase roadmap. This approach prioritizes foundational improvements in fairness and intelligence before focusing on broader scalability and integration. This ensures that as the system scales, it does so on a responsible and highly effective technical foundation.

3.1.1 Phase 1: Foundational Enhancements — Fairness and Intelligence

Fairness Auditing and Bias Detection This enhancement focuses on integrating systematic checks to analyze the system's outputs against sensitive demographic attributes, such as age or location, which are available in the synthetic training data. Its purpose is to actively identify and flag potential biases in the ranking logic, ensuring more equitable and defensible outcomes. This is a critical step for building a responsible AI tool, enhancing trust with HR stakeholders, and mitigating the risks associated with automated decision support.

Advanced Semantic Understanding This enhancement involves two significant upgrades to the core ranking intelligence:

1. **Graph-based skill embeddings:** While the current DistilBERT embeddings capture general semantic meaning, they do not explicitly model the complex, hierarchical relationships between skills. Graph-based embeddings will allow us to represent this network of abilities, capturing non-obvious connections between a candidate's skills and the job requirements.
2. **LLM-based summary explanations:** This upgrade will leverage a Large Language Model (LLM) to translate the system's quantitative scores into natural, human-readable justifications for each candidate's ranking. This will make the system's outputs far more intuitive and actionable for end-users.

3.1.2 Phase 2: Scalability and Integration

Job-Candidate Clustering This feature will implement unsupervised learning techniques to automatically group similar candidates and/or job descriptions. This moves the system beyond its current one-to-many ranking paradigm (one job vs. multiple candidates) to a many-to-many analysis. The strategic benefit is immense: it transforms the system from a simple "ranker" for a single job into a "discovery engine." This will allow users to identify entire talent pools, understand skill relationships within large candidate datasets, and proactively map talent markets.

API Deployment The final and most critical step for production readiness is to wrap the system's logic in a production-grade API, using a modern framework like FastAPI or

Flask. An API abstracts the complexity of the ML pipeline and provides a simple, stable endpoint that other applications can call. This is the key that unlocks seamless integration with existing HR software, internal dashboards, and other enterprise systems, allowing the ranking service to be used on-demand and scaled effectively.

These enhancements, executed in sequence, will systematically transform the prototype into a powerful, reliable, and integrated solution.

4.0 Summary of Anticipated Value

This section summarizes the tangible benefits that project sponsors and technical leadership can expect upon completion of the proposed roadmap. The enhancements are designed to deliver compounding value, improving not only the system's technical capabilities but also its direct business impact.

Proposed Enhancement	Business Value	Technical Value
Fairness Checks	Increased user trust; Reduced risk of biased recommendations; Alignment with responsible AI principles.	Creates a more robust and defensible model; Adds a critical validation layer to the pipeline.
Advanced Semantic Understanding	Deeper, more accurate insights into candidate skills; More intuitive and actionable explanations for recruiters.	Moves beyond simple similarity scores to capture complex relationships and improve model accuracy.
Job-Candidate Clustering	Improved efficiency in managing large talent pools; Enables proactive talent discovery and market mapping.	Adds powerful unsupervised learning capabilities; Allows the system to handle much larger datasets.
API Deployment	Unlocks integration with existing HR systems; Enables scalable, on-demand use; Increases ROI of the system.	Transforms the standalone script into a scalable, reusable microservice; Establishes a production-ready architecture.

These developments will collectively elevate the ranking system into a strategic asset for talent acquisition and management.

5.0 Recommendation and Next Steps

The initial prototype has successfully validated our core approach to ML-driven candidate ranking. The system is fast, interpretable, and effective. Now is the opportune moment to build upon this strong foundation and mature the tool for production use, where its value can be fully realized.

We formally propose the approval of this project, beginning with **Phase 1:**

Foundational Enhancements. Completing this phase will significantly advance the system's core capabilities in fairness and intelligence. It will deliver a more trustworthy and insightful tool while establishing the robust groundwork necessary for the subsequent scalability and integration efforts outlined in Phase 2.

6.0 References and Links

Google collab demo

<https://colab.research.google.com/drive/1CDVZCc0WZhl9JVSejabJchtPtDMmfA3h>

[1] Gunjan Aggarwal

Intelligent Hiring Software- Job Recommendation, 2022

System and profile builder

[2] M. N. Tentue, "The Multi Agent System for Job Recommendation,"
in IOP Conference, Semarang, Indonesia, 2020.

[3] K. C. Kara, "Design and implementation of a job recommender
system," in IEEE, Antalya Turkey, 2017