# AI-Enhanced Recruitment Systems: Transforming Online Hiring Processes with Intelligent Automation for Improved Talent Acquisition and Efficiency

*Dr. Shaik Asif Hussain*[1], *Prof. Shashi Kant Gupta*[2],
[1] Post Doc researcher, centre for post graduate studies, Lincoln University College, Malaysia;
[2] Professor, centre for post graduate studies, Lincoln University College, Malaysia;
sah.ssk@gmail.com; pdf.asifhussainshaik@lincoln.edu.my; shashigupta@lincoln.edu.my

**Abstract:** AI-Enhanced Employment Systems: With Intelligent Robotics for Improved Acquisition of Talent and Efficiency — As talent acquisition trends evolve and the recruitment landscape changes, the demand to sustain the online hiring process is real. This paper proposes a new framework that incorporates machine learning techniques i.e. TF-IDF (Term Frequency Inverse Document Frequency) & KNN classifier to enhance and automate the assessment of resumes across multiple domains like analytics, data science, machine learning and Java. The system has a comprehensive data preprocessing technology to ensure the data quality, and finally the efficient text vector processing and feature extraction implemented by TF-IDF. KNN ranks resumes relevant to designer and subject matter experts. It has shown significant accuracy of 89% on the trainer data and 82% on the test data. Word cloud image helps to visualize word distribution, and filtered phrases, competencies contribute to analysing the data, while simple streamlit UI makes it easy to upload resumes and analyse them. In addition to offering holistic ratings by domain, it offers tactical advice about whether to code-match or not-code match someone based on their coding criteria. The benefit of this strategy is that it is a useful tool for the modern HR process and helps improve the objectivity of hiring and the efficiency of decision-making while also speeding up the hiring process.

**Keywords**: Acceptance, Automation, Classification, Machine Learning, Resume Evaluation, Streamlit, Talent Acquisition and Word-cloud

**Introduction**

Machine learning and artificial intelligence are transforming several industries, including recruiting. Traditional recruiting procedures often include screening of handwritten resumes, which can prove tedious and biased. Sophisticated recruiting systems rely on intelligent automation [1] to address these problems and streamline the hiring process. As these systems are based on natural language processing (NLP) they often rely on methodologies such as TF-IDF (Term Frequency-Inverse Document Frequency) [2] when mapping textual data to the ML model. These technologies aim to make hiring more impartial and efficient by evaluating resumes based solely on their content and how well they match job descriptions. Based on current industry statistics, an average efficiency of 50% is estimated for corporations owing to the adoption of AI-based technologies [3]. In addition, they can address the common issue of lengthy hiring processes by reducing time-to-hire by roughly 30%. AI-powered recruiting solutions can assist organisations in making better hiring decisions through automating the CV screening process and providing data-driven insights. This increases hiring efficiency and reduces the firm's operating costs. The paper examines a new approach to modernizing hiring practices. TF-IDF and KNN[4] are used in this study to evaluate and rank resumes in various fields such as analytics, machine learning, data science, Java, etc. KNN explain how good this method is at evaluating the contents of resumes and matching it to job criteria, with success rates of 89% on training data and 82% on test data success. Moreover, the system contains a word cloud that represents the crucial ideas and skills [4], and also provides a streamlit user interface [5] for easy shortlisting and evaluation.

**Related work**

In the past Employee Automation systems were an extremely fast way for hiring teams as they used many different machine learning & NLP techniques to make the process more efficient. The main emphasis of early efforts was screening resumes using keyword-based filters [6] and simple text matching algorithms. These algorithms compared resume content with predetermined job requirements in an effort to find qualified individuals. Unfortunately, these methods frequently suffered from shortcomings as a result of their incapacity to fully grasp the text's complex context and semantic meaning, which resulted in less than ideal performance when it came to finding qualified applicants. Krishnan et al. examines how AI affects HR productivity and worker satisfaction, with a particular emphasis on how AI is integrated into hiring, onboarding, performance management, and talent development procedures [7]. It offers important information for businesses looking to use AI in people development, including insights on the benefits and difficulties of integrating AI. Natarajan et al. investigates the use of artificial intelligence (AI) to personnel management with the goal of streamlining the hiring, training, and retention procedures. Recruitment, predictive analytics, tailored instruction, managing performance, retention, management of succession, and diversity efforts [8] are among the AI-powered tactics covered in this article.

As technology developed, more advanced techniques were added, including machine learning models like Random Forests and Support Vector Machines (SVMs). These models improved the capacity to rank and categorize resumes according to several characteristics that were taken from the text, such as education, experience, and talents. Rayyan et al. report looks at how AI and ML are affecting hiring in Indonesia's e-commerce industry, and it identifies issues [9] with data quality, model accuracy, and system flexibility. Gupta et al. built an application of AI in hiring procedures is examined in this thesis, with an emphasis on candidate quality and efficiency [10]. The usefulness of these systems has been improved by the addition of visualization tools, such as word cloud, and user-friendly interfaces [11], which make it simpler for HR professionals to understand the data and make wise decisions. Safitri et al. investigates the effects of AI integration and a high-performance work system on hiring and performance reviews of employees. It emphasizes AI's potential to improve HRM procedures, especially in hiring and performance review [12], by drawing on research on healthcare. Halid et al. proposed an AI usage in HR procedures is increasing as a result of organizations' adaptation to online platforms and IT developments [13].

**Key Contribution**

This paper suggests a hybrid NLP-based resume screening framework by integrating TF-IDF text vectorization with one of the classifiers which is KNN (K-Nearest Neighbours) to generate a domain specific rank list of resumes with an accuracy of 89% during training and 82% on test sets for different job domains (e.g., analytics, data science, machine learning, java). Unlike conventional keyword screening methods, the system offers unbiased, data-driven assessments that enhance efficiency in hiring decisions and eliminate hiring bias. It is also accompanied with a Word Cloud visualization module and a web application with streamlit, to aid the tool in being user friendly, even for those HRs who are not technically inclined, and does not require any programming knowledge.

**Data Collection and Preprocessing**

The main goal of the AI-enhanced recruiting system's data gathering and preparation phase (as shown in Fig.1) is to get textual data from job descriptions and resumes ready for efficient analysis and modeling. The first step in this approach is to compile a varied and representative collection of job descriptions and resumes from a range of industries, including analytics, machine learning, data science, and Java [14]. Data collection is done from multiple sources, including resume databases, corporate career pages, online job boards, and so forth in order to ensure that a broad and diverse sample is obtained. This variety is important for training a good model to properly address different résumé format and job description. Once the data is collected, it is cleaned to remove irrelevant or redundant data. At this stage, duplicates are removed, formatting errors are corrected, and all non-textual elements that add no value to the analysis are also removed.] This cleaning process makes sure that the data used for further processing is uniform and relevant, which is crucial to building a reliable and effective model. Text normalisation is what comes next step, which entails converting the unprocessed text data into a standardized format that can be used for analysis.

Several crucial methods are used in this normalization process, including lowercasing all text to guarantee consistency, getting rid of the punctuation and unique characters [15] that might not add to the semantic meaning, and getting rid of stop words, which are words like "and," "the," or "is" that are commonly used but have little meaning when it comes to text analysis. Tokenizing a CV into individual words, for example, enables the system to assess the frequency and significance of particular concepts and abilities. Applying other text processing methods, such feature extraction and phrase frequency analysis, is also made easier by tokenization. The textual input must next be transformed into numerical representations so that machine learning algorithms can process it after tokenization is finished. Word embeddings and Term Frequency-Inverse Document Frequency (TF-IDF) are two methods used to accomplish this conversion. To measure a term's significance in a text in relation to its frequency throughout the full dataset, TF-IDF is utilized.

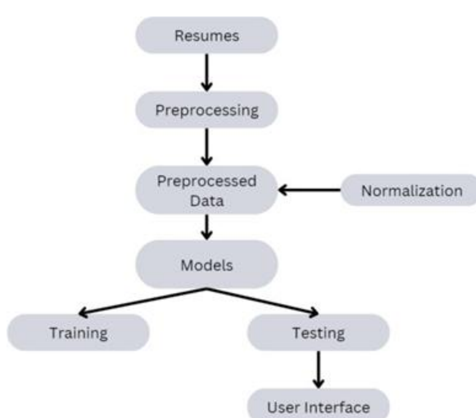


*Figure 1. Architecture*

## Principles and Methods

### A. TF IDF Vectorizer

Term Frequency-Inverse Document Frequency (TF-IDF) a well-known approach in text processing and information retrieval which measures a word's importance in a document relative to a corpus, or set of documents. TF-IDF is mainly used to find relevant words in a certain text while minimizing the effect of terms that are domains of all articles. This helps in distinguishing between terms that are important within individual documents and terms that are common across the whole collection. TF-IDF is built on two key components: Term Frequency (TF) and Inverse Document Frequency (IDF). Term frequency is the number of times that term occurs in a document. This is based on the assumption that terms that appear frequently in an article are more likely to be relevant to the content of that document. So, if the word "Data Science" appears many times on a resume, the implication is that this is important to the document. But TF alone does not consider how common or rare a term is in the complete set of texts. TF-IDF is a good method of turning text input into numerical features a machine learning model may use. It assigns a TF-IDF score to terms in a document indicating how relevant they are. These scores are used to construct the feature matrix for the document collection, with each row representing an item of paper and each column representing a phrase from the corpus. This matrix serves as the input for many machine learning algorithms. One advantage of TF-IDF is that it is easy to use and efficient for determining the importance of phrases within a document. TF-IDF, in contrast to more complex models, provides a straightforward way of representing text, which balances how common a word is across all documents with how rare it is within a particular set of documents. It's this configuration that makes it an easy and powerful tool for many text-based analytics tasks, especially when dealing with large text datasets. Despite the effectiveness of TF-IDF, it has its own defects.

### B. KNN Classifier

K-Neighbors-Classifier (a.k.a lazy learners): Lazy learners are members of the instance-based learning family of algorithms, where the decision to remember a case instance of the training dataset is made until it has been created. Unlike many other methods, which build a model from the training set, the KNN makes predictions directly from the training data. The principle works on the basis of similarity, where it uses the idea of "nearest neighbours" to categorize the data points that are new. For simplicity and interpretability, it is thus a commonly used choice for classification tasks. The idea behind the KNN, in fact, is to make predictions on new data points based on what its closest neighbors in the feature space have done. When a new sample has to be classified, the method determines the 'k' nearest training samples from a particular location. Common metrics used are Euclidean distances, however, depending on the given task and type of data, it is also possible to use other distance measures such as Manhattan or Minkowski.

K is a significant config in KNN, so selection of K i.e. number of neighbours to consider is significant. If k is too small, overfitting occurs, as it would cause the model to be too relying on noise in the data. On the other hand, if 'k' is large enough, it can lead to underfitting as the model can be more generalized because it considers a large number of neighbors to account the prediction. The choice of the best number for 'k' is often a balance between variance and bias heckled with plenty of trial and error and validating techniques such as cross-validation. This is a distance-based classification, where the K-Neighbors-Classifier averages the labels of the 'k' nearest instances from the new data point to determine the class

[20]. This aggregation typically uses a majority voting technique, in which case the class that appears most often among those neighbours is assigned to the new sample. Alternatively, you can use weighted voting, where neighbors closer to the unseen location can be given more weight in producing the final categorization result.

**Results**

The TF-IDF and K-Neighbors Classifier-powered AI-enhanced hiring system results replicate the fact that it automates and enhances the recruitment process. The system aims to provide a data driven, objective evaluation of resumes based on how relevant they are to specific job types such as analytics, machine learning, data science and Java. The key outcome of our work will be the construction of a robust tool that accurately ranks and classifies resumes, enabling HR managers to make better hiring decisions. The TF-IDF implementation also notes a word's relevance, allowing the system to weigh words in resumes and job descriptions when determining importance. The K-Neighbors-Classifier achieves high metrics in performance when classifying the resumes using the scores gathered from the TF-IDF. Overall, the findings prove the classifier can VA a compelling applicant simply based on their resume content with 89% accuracy on the train data and 82% accuracy on the test data.

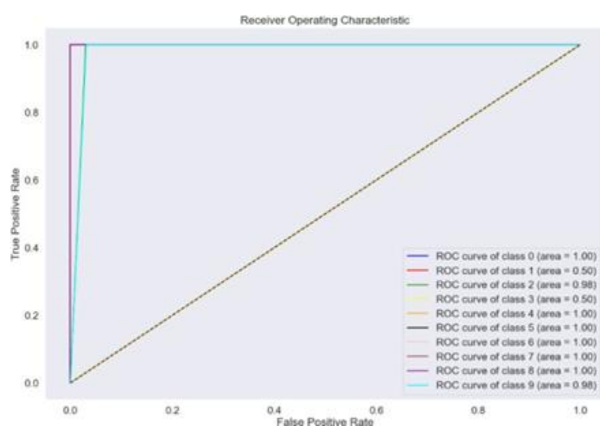


*Figure 2. ROC Curve*

In addition to that, the system comes with a streamlit interface that makes the product intuitive to use and therefore more accessible to some extent. HR mangers and users can also have the ease to deal with the system via the streamlit UI, upload CV, and get feedback in real time regarding how well a candidate matches the skills for certain job categories. Our interface is designed to alleviate the burden of resume screening and is very user friendly, so, without requiring a high level of technical knowledge, users can easily assess resumes. The streamlit UI aims for a smooth user experience with components such as interactive displays, file upload buttons, and educational visualizations. Direct resumes can be submitted by the users via the portal. These files are used by the system to assign domain- specific ratings.


*Figure 3. Streamlit User Interface*

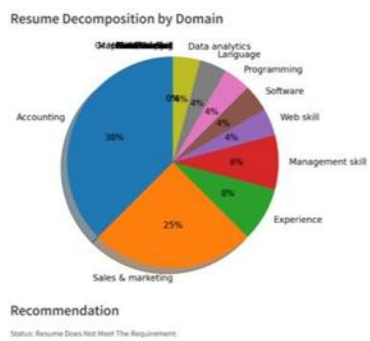


*Figure 4. Resume Screening*

In addition, this recruiting system employing the streamlit UI demonstrates how friendly interfaces can enhance access to state-of-the-art tools. A simple user interface (UI) for efficient resume review helps facilitate system adoption and HR personnel to utilize more of the system's features. According to performance data and user reviews of the system, the strategy can drastically enhance the way hiring is done, which translates to better efficiency all around. AI-enhanced recruiting functionalities demonstrate how TF-IDF and K-Neighbors-Classifier can effectively integrate to perform computerized evaluation of resumes. Thanks to its high accuracy and the streamlit UI that it provides further it a user-friendly interface on web (as shown in Fig. 3&4). It solves the greatest hurdles of hiring and offers fresh human resources specialists at some of the best-caliber organizations a more innovative and efficient recruiting experience, delivering credibility for categorization, discernible patterns, and a seamless user interface that serves for better decision-making.

**Conclusions**

AI-enhanced employment system capable of automation and optimization with K-Neighbors-Classifier and TF-IDF privacy. A TF-IDF along with K-Neighbors Classifier has shown a very good performance with

89%(Learning) and 82%(Testing) classification accuracy on resumes by jointly measuring/extracting term significance as well as its fine classfication. Such a systematic technique allows candidates to fill gaps in traditional recruiting methods by categorizing applicants based on eligibility and qualifications for multiple job types. And with an easy streamlit interface, HR can quickly upload resumes, view holistic evaluation by AI, and choose better candidates. In addition, this is a tool that provides an objective and data-driven way to recruit which makes the hiring process easier as it automates the screening of resumes and reduces biases while also offering meaningful visualizations.

**References**


[1] Jha, Nishant, Deepak Prashar, and Amandeep Nagpal. "Combining artificial intelligence with robotic process automation—an intelligent automation approach." *Deep learning and big data for intelligent transportation: enabling technologies and future trends* (2021): 245-264.

[2] Alfarizi, Muhammad Ibnu, Lailis Syafaah, and Merinda Lestandy. "Emotional text classification using tf-idf (term frequency-inverse document frequency) and lstm (long short- term memory)." *JUITA: Jurnal Informatika* 10.2 (2022): 225-
232.

[3] Huang, Ziqi, et al. "A survey on AI-driven digital twins in industry 4.0: Smart manufacturing and advanced robotics." *Sensors* 21.19 (2021): 6340.

[4] Sari, I., and R. Kosasih. "Prediction types of legal cases in Indonesia using the TF-IDF method and the KNN algorithm." *AIP Conference Proceedings*. Vol. 2689. No. 1. AIP Publishing, 2023.

[5] Data, Build Web-Based, and Sujay Raghavendra. "Beginner's Guide to Streamlit with Python."

[6] Daryani, Chirag, et al. "An automated resume screening system using natural language processing and similarity." *ETHICS AND INFORMATION TECHNOLOGY [Internet]. VOLKSON PRESS* (2020): 99-103.

[7] Krishnan, L. R. K., K. Praveen, and S. Poorani. "Artificial Intelligence in Human Resource Management: Enhancing Efficiency & Transforming Employee Experience." *LRK Krishnan, Praveen K & Poorani S* (2024).

[8] Sundarapandiyan Natarajan, Dr Korapu Sattibabu, et al. "AI- Powered Strategies for Talent Management Optimization." *Journal of Informatics Education and Research* 4.2 (2024).

[9] Rayyan, Mohammad, Nakayenga Sharifah, and Rini Kuswati. "Revolutionizing Talent Acquisition in Indonesia's E- Commerce Industry: The Transformative Impact of AI and Machine Learning." *Journal of Humanities and Social Sciences Studies* 6.4 (2024): 01-12.

[10] Gupta, Aishwarya, and Kimiya Rahimi Ata. "Data-Driven Hiring: Implementing AI and Assessing the Impact of AI on Recruitment Efficiency and Candidate Quality." (2024).

[11] Jin, Yuping. "Development of word cloud generator software based on python." *Procedia engineering* 174 (2017): 788-792.

[12] Safitri, Indah, Hafidz Widodo, and Muafi Muafi. "Application of High-Performance Work System with AI in Employee Recruitment and Performance Evaluation." *Technology- Driven Business Innovation: Unleashing the Digital Advantage, Volume 1*. Cham: Springer Nature Switzerland, 2024. 465-473.

[13] Boddapati, Mohan Sai Dinesh, et al. "Creating a Protected Virtual Learning Space: A Comprehensive Strategy for Security and User Experience in Online Education." International Conference on Cognitive Computing and Cyber Physical Systems. Cham: Springer Nature Switzerland, 2023.
[14] Kamath, Uday, and Krishna Choppella. *Mastering java machine learning*. Packt Publishing Ltd, 2017.
[15] Lapworth, Katherine. *The Writer's Guide to Good Style: A 21st Century guide to improving your punctuation, pace, grammar and style*. Hachette UK, 2019.