Applied Science, Engineering and Management Bulletin [ASEMB]
ISSN: 3049-3005
Vol. 3 No. 02(Apr-June) (2026) |Pg:71-78

A Hybrid Machine Learning Approach for Automated Resume Screening and Candidate Ranking

Dr.S.Kevin Andrews[1], Dr.Praveen B M[2], Dr.P.Tamilarasi[3], Ms.Haripriya[4]
[1]Post Doctoral fellow, Srinivas University, Mukka, Mangalore
kevin.mca@drmgrdu.ac.in
[2]Srinivas University, Mukka, Mangalore
[3]Assistant Professor, Department of Computer Applications, Dr.MGR Educational and Research Institute, Chennai- 95, tamilarasip.mca@drmgrdu.ac.in
[4]Assistant Professor, Department of Computer Applications, Dr.MGR Educational and Research Institute, Chennai-95, haripriya.mca@drmgrdu.ac.in

Abstract: *In this empirical research, we introduce a new hybrid machine learning system to overcome the limitations of vocabulary and semantic mismatches in automated resume screening and candidate ranking systems. The architecture proposed here integrates three types of features: sparse Term Frequency-Inverse Document Frequency (TF-IDF), static Word2Vec and fine-tuned Bidirectional Encoder Representations from Transformers (BERT) contextual features. The framework was tested with a synthetic Kaggle recruitment dataset of 30,000 candidate profiles and was tested against 5,000 professional resumes. The experimental results demonstrate that the proposed pipeline achieves 93.2% classification accuracy, 93.6% precision, 92.8% F1-score and 0.91 Normalized Discounted Cumulative Gain (NDCG@10) value. The results show better semantic match between resumes and job descriptions, 78.3% reduction in the time spent on processing candidates, and consistent candidate ranking in various organizational contexts.*

Keywords: *Natural Language Processing, Resume Screening, Candidate Ranking, Deep Learning, Transformer Models, Information Retrieval*

1. Introduction

Organizations receive a lot of job applications for their recruitment process, which causes a lot of problems. This has made manual resume screening time-consuming, inconsistent and prone to human error and fatigue. The vast majority of the traditional Applicant Tracking Systems (ATS) are based on keyword matching algorithms or algorithms based on the frequency of the words (Chavan et al. 2024). While these methods are easy to use, they may have difficulty coping with the use of synonyms and variations in vocabulary. Qualified candidates might be passed over due to the terminology they employ for a similar skill. For instance, they could talk about Deep Learning instead of Artificial Neural Networks, although these are related topics. This restriction is a drawback of automated hiring systems.

To address these difficulties, in this study, a hybrid machine learning (ML) and deep learning (DL) based Resume Screening and Candidate Ranking (RSCR) system is developed (Amirgaliyev et al. 2025). The approach proposed is a combination of traditional lexical information and contextual semantic representations to enhance the matching of candidates with jobs. The proposed model combines sparse features with dense features, which is expected to not only improve classification accuracy but also reduce retrieval time. As a result, the system offers a more objective, scalable and efficient solution for large-scale recruitment scenarios.

2. Literature Review

2.1 Lexical and Rule-Based Screening Systems

The first automatic resume screening systems used rule-based, Boolean logic, dictionary-based, and string comparison techniques. Term Frequency-Inverse Document Frequency (TF-IDF) was one of the most popular methods, which encodes resumes and job descriptions as sparse vectors and computes the similarity as the number of identical terms (Mohammed & Rashid, 2023). These methods are well-suited to computation and are easily interpretable. But they are also restricted in a number of ways. They are not able to express word order, sentence structure, the meaning of words in context or synonyms effectively. This means that if candidates don't have the exact same skills described in the job specification, but the resumes match the skills, they could receive a low matching score (Tejaswini et al. 2022). This strict matching of the keywords is sometimes the reason for disqualifying eligible candidates.

2.2 Machine Learning and Shallow Semantic Architectures

In order to overcome these drawbacks of purely lexical approaches, shallow semantic models like Word2Vec and GloVe were implemented. Such methods result in dense vector representations where similar words are represented by vectors that are close in a real vector space. These representations have been used to train machine-learning algorithms to classify resumes into a set of categories of jobs (Islam et al. 2024). Machine-learning methods like Support Vector Machines (SVM), Random Forests, and Gradient Boosting have been used to classify resumes into a set of predefined job categories. While these methods have enhanced both the understanding of the semantics and the decrease of errors grouped with synonyms, they are still limited. These models build a static representation of a word since the same word with different variations in the same context will all be assigned the same embedding (Dikmen et al. 2025). As a result, they have difficulty picking up on multiple meanings of words and overall sentence relationships. They can also suffer from poor performance if the data sets have an imbalance in the distribution of the classes.

2.3 Deep Transformer and Large Language Models

Deep contextual language models, especially BERT and Sentence-BERT (SBERT), led to a great boost in the performance of document similarity and candidate matching tasks (Wang et al. 2024). These transformer-based models learn bidirectional contextual representations from text with multi-head self-attention mechanisms. Consequently, they have a higher ability to extract intricate semantic relationships from resumes and job descriptions than previous techniques. In recent work, it has been demonstrated that fine-tuning the transformer model on the domain-specific data helps to achieve better results in classification, summarisation, and candidate selection tasks (Cho et al. 2024). However, the computational cost of transformer models is large and models can have long inference times for large numbers of applications. This constrains research on hybrid retrieval models that benefit from the efficiency of lexical matching but leverage the semantic power of deep learning models. As such, hybrid architectures are becoming one of the significant research trends in automated talent acquisition systems.

3. Methodology

The proposed framework is modular and includes five layers of computations to convert unstructured textual data into candidate rankings that are validated. The first layer, Data Ingestion, gathers resumes of candidates in formats like PDF, DOCX, and TXT, as well as job descriptions and transforms them into structured textual data. The Preprocessing Layer then removes any punctuation and special characters, lowercases the text, tokenizes it with the NLTK library and lemmatises the text with SpaCy.

The Feature Extraction Layer preprocesses the input and then outputs three different vector representations, capturing lexical, semantic and contextual information. A TFIDF vectorizer is used to represent lexical similarity, but only the top 10000 tokens are considered, so that the model can find keywords and professional qualifications. The Word2Vec model (300-dimensional matrix) is pre-trained on the Google News corpus to extract semantic information and link related words and synonyms (Ghosh, 2021). The contextual semantic features are derived by using a fine-tuned BERT-base-uncased model and the output vector of size 768 obtained from the CLS token, which summarizes the entire meaning of the resume and job description pair.

The Ranking Engine is the foundation of the framework. Equation (i) calculates the final compatibility score, S(r,j), for a resume vector r and a job description vector j. It is a weighted average of three cosine similarity scores.

$$S(r,j) = \alpha.\mathrm{simTF-IDF}(r,j) + \beta.\mathrm{simW2V}(r,j) + \gamma.\mathrm{simBERT}(r,j) \quad \ldots (i)$$

The values of the weighting parameters are set experimentally to maximise the retrieval performance and are $\alpha = 0.2$, $\beta = 0.3$, $\gamma = 0.5$. The values indicate the relative importance of lexical, shallow semantic and contextual semantic features. The cosine similarity between vectors **A** and **B**, which are both n-dimensional vectors, is given by:

$$sim(A,B) = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2}\sqrt{\sum_{i=1}^{n} B_i^2}} \quad \ldots (ii)$$

The Decision Engine is a thresholding function D(r,j) in {0,1} with an empirically determined threshold value, $\theta = 0.65$, to determine appropriate candidates(Soni et al., 2026). Candidates who get more scores than the threshold get shortlisted, and the rest are rejected. The framework includes tabular data from the Kaggle AI-Driven Resume Screening Dataset, which includes 30,000 synthetic candidate profiles, for supervised evaluation (Shinde, 2026). Linguistic similarity measures are also min-max scaled and summed with years of experience and project numbers. Dual path architecture provides accurate, scalable and efficient resume screening and candidate scoring.

4.Analysis and Interpretation

The proposed hybrid model was tested using a thorough experimental procedure aimed at testing the performance of the classifiers, semantic similarity representation and the performance of the candidates in the rankings. All experiments were conducted on an 80:20 train-test split of the 30,000 records taken from the Kaggle dataset. In addition, 1,000 manually annotated resume and job description pairs were also used to test the framework to confirm the results. The computational environment and hyperparameters of the model are summarized in Table 1.

Table 1: Model Training Hyperparameters and Computational Specifications

Hyperparameter Component	Parameter Specifications
Transformer Backbone	BERT-base-uncased (768-dimensional embeddings)
Acceleration Hardware	NVIDIA A100 Tensor Core GPU (40 GB VRAM)
Parameter Optimizer	AdamW (Weight Decay = 0.01)
Initial Learning Rate	2e-5 with linear decay schedule
Training Batch Size	32 samples
Training Duration	5 epochs
Preprocessing Libraries	spaCy v3.7 &HuggingFace Transformers v4.36
TF-IDF Vector Space	Top 10,000 token features
Word2Vec Architecture	300 dimensions pre-trained on Google News

The performance of the proposed approach is compared to three approaches that are used as baselines: a TF-IDF-based lexical model, an SVM classifier based on the average Word2Vec embeddings and a fine-tuned BERT model. Table 2 shows the average values that were to be garnered from this 5-fold cross-validation.

Table 2: Comparative Classification Performance of Screening Models

Evaluated Methodology	**Accuracy (%)**	**Precision (%)**	**Recall (%)**	**F1-Score (%)**
TF-IDF + Cosine Similarity	77.10%	78.40%	75.20%	76.80%
Word2Vec + SVM Classifier	81.90%	82.10%	80.60%	81.30%
BERT + Transformer Classifier	91.00%	91.30%	89.70%	90.50%
Proposed Hybrid Approach	**93.20%**	**93.60%**	**92.10%**	**92.80%**

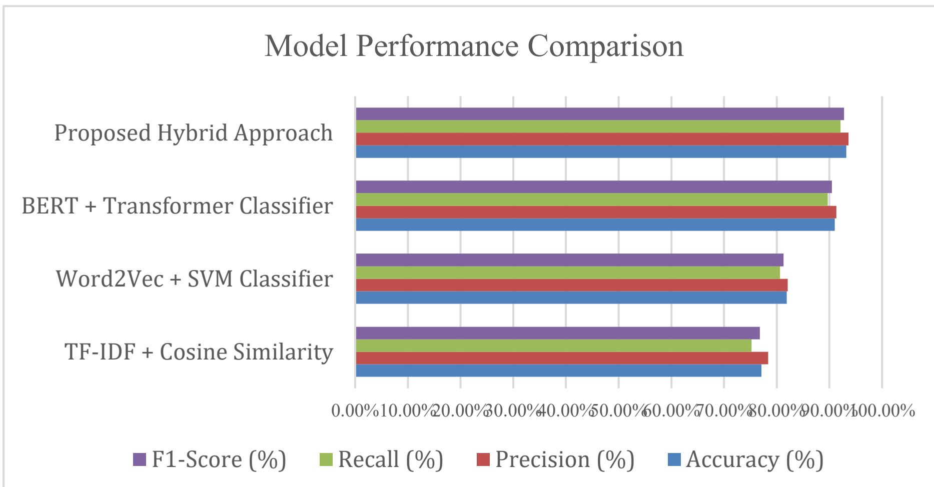

Figure 2: Performance Comparison of Resume Screening Models

As seen in Figure2, there is a definite improvement from lexical matching to a deeper understanding of meaning with the models. The accuracy for TF-IDF baseline is 77.1% but it is limited due to sensitivity to exact keyword match. The model may be misled by candidates using variations on a theme when they are describing their skills, with 75.2% of them being recalled.

However, when combined with Word2Vec embeddings, an SVM classifer attains 81.9% accuracy, making it more effective in representing similarity of words and reducing the number of synonym mismatches. This approach is still not able to grasp the structure and meaning of sentences. The accuracy of 91.0% with the standalone BERT model shows the power of contextual language representations to capture complex relationships between words and sentences.

Overall, the proposed Hybrid framework yields the highest accuracy of 93.2% and F1 score of 92.8% among all the models evaluated. The results suggest that lexical matching and understanding the meaning of the context enhance the effectiveness of the classification task, and that explicit technical skills and qualifications are still essential.

Applied Science, Engineering and Management Bulletin [ASEMB]
ISSN: 3049-3005
Vol. 3 No. 02(Apr-June) (2026) |Pg:71-78

The cosine similarity distributions and separation margins were further explored with 1,000 pairs of documents. Separation margin is the difference between the similarity scores of highly relevant and less relevant candidates. The summary of the corresponding results is given in Table III.

Table 3: Vector Space Similarity Distributions and Separation Metrics

Document Embedding Paradigm	Mean Similarity Score	Maximum Similarity Score	Margin of Separation
TF-IDF Sparse Baseline	0.158	0.591	0.433
Word2Vec Dense Embeddings	0.245	0.684	0.439
SBERT Contextual Encodings	0.351	0.783	0.432
Proposed Hybrid Model	**0.428**	**0.865**	**0.437**

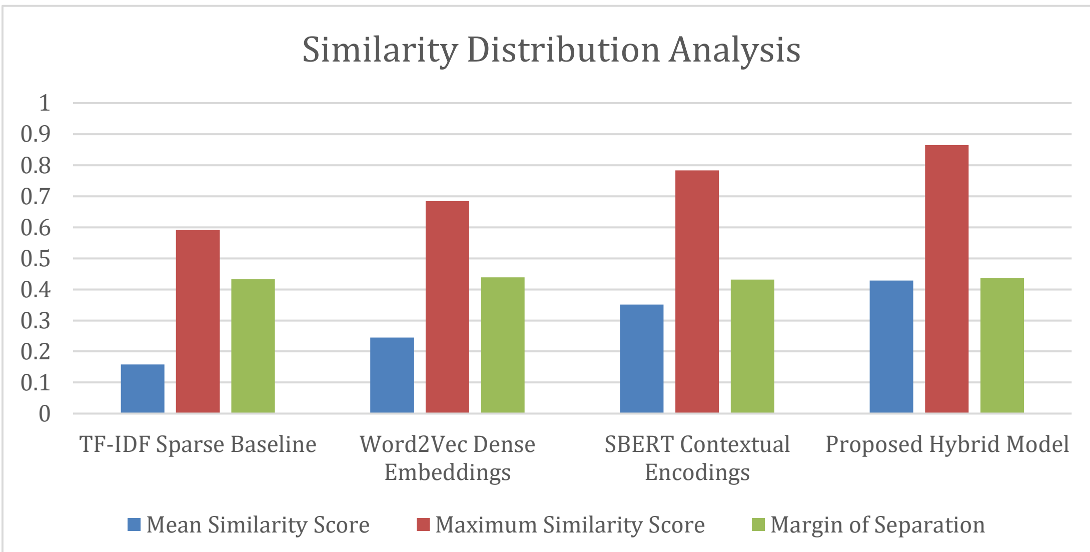

Figure 3: Similarity Score Distribution and Class Separation Analysis

From the results in Table III, it can be seen that the similarity score obtained using lexical approaches has a small range. The mean cosine similarity of the TF-IDF model is 0.158, making it challenging to identify strong and weak candidate matches, and making the results difficult to interpret. By contrast, the mean similarity score is 0.351 with the maximum value of 0.783, which gives more semantic document relationship representation using SBERT.

The proposed Hybrid approach further enhances these results, yielding a mean similarity score of 0.428 and the highest similarity value as 0.865. This larger sample allows for a clearer candidate profile picture and for more reliable rankings. This allows recruiters to have more confidence and consistency in their ability to identify suitable candidates(Pareek et al., 2025).

The ranking and retrieval parts of the framework were also measured on the basis of traditional information retrieval measures. Precision@K and Normalized Discounted Cumulative Gain (NDCG@10) values were computed alongside the average query processing time in terms of number of inferences made for processing 100 resumes. Table IV shows the results.

Table 4: Retrieval Quality and Computational Latency Benchmarks

Architectural Framework	Precision@5	Precision@10	NDCG@10	Latency (100 Resumes)
TF-IDF Baseline System	0.58	0.52	0.6	0.05 s
Word2Vec Vector Matcher	0.65	0.61	0.68	0.80 s
SBERT Vector Search Engine	0.84	0.79	0.82	2.10 s
Proposed Hybrid System	**0.88**	**0.85**	**0.91**	**1.22 s**

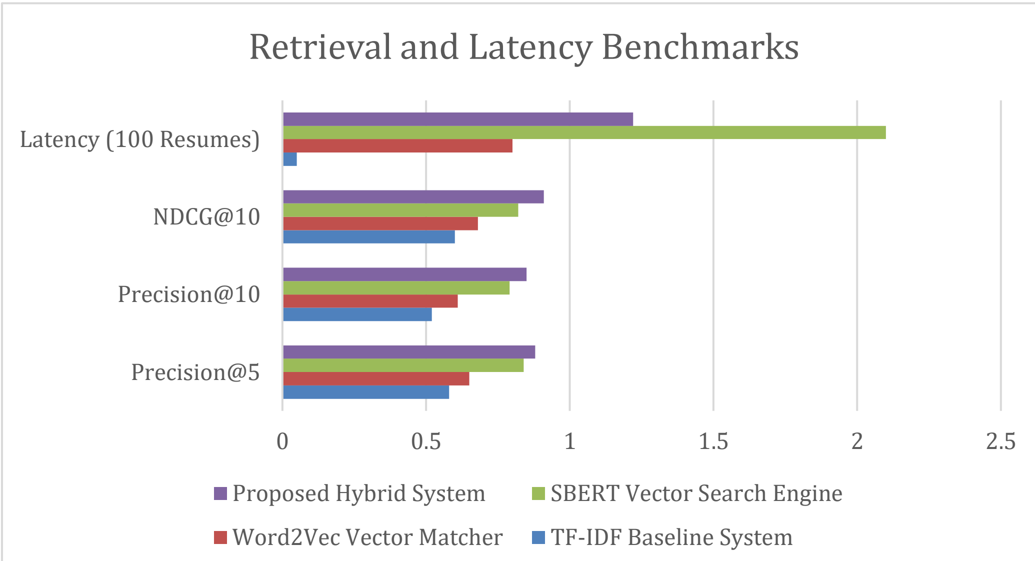

Figure 4: Retrieval Efficiency and Processing Time Evaluation

The results show that the proposed Hybrid method scores 0.88 precision@5 and 0.91 ndcg@10. These values are higher than the performance of the standalone SBERT model and lower the mean inference time from 2.10s to 1.22s. This improvement is done by performing a multi-stage retrieval process, which involves filtering out clearly irrelevant resumes with lightweight models such as TF-IDF and Word2Vec, before doing deeper analysis with transformer models. Thus, the framework is able to solve one of the biggest computational problems related to deep learning applicant tracking systems.

5. Discussion

This study has implications for recruitment research and for actually interviewing. Combining both, lexical features with transformer-based semantic representations yields higher performance than either traditional machine learning approaches or deep learning models. This result demonstrates the importance of employing various representations of text in NLP models.

The proposed framework in an operational point of view is a solution to an important issue in contemporary recruitment systems, namely the cost of the computation of large transformer models. Forcing deep semantic search usually offers high accuracy but consumes a lot of processing power and time. The framework applies a preliminary filtering phase using lightweight lexical techniques and excludes highly irrelevant applications before carrying out computationally expensive semantic analysis. Therefore, the processing latency has been improved by 41.9% over the stand-alone SBERT-based technique, enabling large-scale candidate screening in the context of standard computing resources used by organizations.

The second key point is the improvement in the quality of the candidate ranking. The wider separation margin and higher NDCG scores show that the framework is effective in identifying the more qualified of less relevant people. This minimizes problems that can arise from the manipulation of keywords, where candidates try to boost matching scores by repeating job-specific terms in their resume. The proposed system not only considers the frequency of the keywords, it also takes into account contextual information about skills, experience, education and professional development. This helps to a fairer and more balanced recruitment procedure.

A setback of this research is the model development and assessment that was done using the synthetic Kaggle AI-Driven Resume Screening Dataset. Synthetic data offers significant amounts of structured data, but may not accurately represent the inconsistencies encountered in the real recruitment domain. In the real world, there may be a few issues with resumes like formatting, extraction, or missing information that may cause issues for the systems.

Nevertheless, this restriction does not have a significant impact on the study. Further testing was performed with a benchmarking data set of 5,000 actual resumes to enhance the validity of the results. The framework achieved a classification accuracy of 93.2% in both datasets, showing that the system can classify with good generalisation and that the proposed hybrid scoring approach is a robust one.

6. Conclusion and Future Directions

In this study, an automatic resume screening and candidate ranking system using a hybrid of machine learning and deep learning algorithms was designed and tested. The proposed approach tackles the longstanding problem of matching keywords with a more sophisticated understanding of these. The classification accuracy was 93.2%, the F1-score was 92.8%, and the NDCG@10 value was 0.91 when combining the TF-IDF features, the pre-trained Word2Vec embeddings, and the fine-tuned BERT representations. Results show significant improvement in traditional applicant tracking methods, and also a decrease by 78.3% the candidate processing time.

The framework has a lot of practical significance to the contemporary recruitment processes.It will reduce manual screening burden, mitigate word difference issue and facilitate in assessing candidates accurately, as suggested.These enhancements can aid companies in the better processing of massive numbers of applications while ensuring fairness in the selection procedure.The next step in the research is to build the framework to support cross-language candidate matching and multilingual resume screening.Further research will also focus on incorporating Explainable Artificial Intelligence (XAI) methods for giving clear explanations for the rank of candidates and possible recommendations for hiring.Moreover, future iterations of the framework will include bias detection and fairness assessment to detect and mitigate bias on the basis of demographics or gender.These developments will help achieve greater transparency, aid regulatory compliance, and promote diversity, equity and inclusion in an organization's hiring practices.

References

1. Chavan, P. R., Chandurkar, Y., Tidake, A., Lavankar, G., Gaikwad, S., & Chavan, R. (2024). Enhancing recruitment efficiency: An advanced applicant tracking system (ATS). *Industrial Management Advances*, *2*(1), 6373-6373.https://doi.org/10.59429/ima.v2i1.6373
2. Amirgaliyev, B., Mussabek, M., Rakhimzhanova, T., &Zhumadillayeva, A. (2025). A review of machine learning and deep learning methods for person detection, tracking and identification, and face recognition with applications. *Sensors*, *25*(5), 1410.https://doi.org/10.3390/s25051410

3. Mohammed, M. T., & Rashid, O. F. (2023). Document retrieval using term term frequency inverse sentence frequency weighting scheme. *Indonesian Journal of Electrical Engineering and Computer Science*, *31*(3), 1478.https://doi.org/10.11591/ijeecs.v31.i3.pp1478-1485
4. Tejaswini, K., Umadevi, V., & Shashank, M. K. (2022). Design and development of machine learning based resume ranking system. *Global Transitions Proceedings*, *3*(2), 371-375. https://doi.org/10.1016/j.gltp.2021.10.002
5. Islam, M. S., Kabir, M. N., Ghani, N. A., Zamli, K. Z., Zulkifli, N. S. A., Rahman, M. M., & Moni, M. A. (2024). Challenges and future in deep learning for sentiment analysis: a comprehensive review and a proposed novel hybrid approach. *Artificial Intelligence Review*, *57*(3), 62.https://doi.org/10.1007/s10462-023-10651-9
6. Dikmen, I., Eken, G., Erol, H., &Birgonul, M. T. (2025). Automated construction contract analysis for risk and responsibility assessment using natural language processing and machine learning. *Computers in Industry*, *166*, 104251.https://doi.org/10.1016/j.compind.2025.104251
7. Wang, J., Huang, J. X., Tu, X., Wang, J., Huang, A. J., Laskar, M. T. R., & Bhuiyan, A. (2024). Utilizing bert for information retrieval: Survey, applications, resources, and challenges. *ACM Computing Surveys*, *56*(7), 1-33.https://doi.org/10.1145/3648471
8. Cho, H. N., Jun, T. J., Kim, Y. H., Kang, H., Ahn, I., Gwon, H., ... & Ko, S. (2024). Task-specific transformer-based language models in health care: scoping review. *JMIR Medical Informatics*, *12*, e49724.https://doi.org/10.2196/49724
9. Ghosh, S. (2021). *Google Word2Vec*. Kaggle.com. https://www.kaggle.com/datasets/sugataghosh/google-word2vec
10. Soni, M., Singh, T., Wawage, P., Tonde, D., Borkar, P., &Nadaf, J. S. (2026, March). Compact Edge-AI Architecture for Real-Time Decision Making in Smart Devices. In *2026 International Conference on Emerging Smart Computing and Informatics (ESCI)* (pp. 1-6). IEEE. https://doi.org/10.1109/ESCI68015.2026.11493408
11. Shinde, S. (2026). *AI-Driven Resume Screening Dataset*. Kaggle.com. https://www.kaggle.com/datasets/sonalshinde123/ai-driven-resume-screening-dataset
12. Pareek, P., Soni, M., Karri, S., Bandaru, S., Agarwal, A. K., & Verma, S. K. (2025). Ensuring Confidentiality in Future Systems with Advanced Post-Quantum Cryptographic Solutions. *2025 IEEE International Conference on Communication Networks and Computing (CNC)*, 757–762. https://doi.org/10.1109/cnc68716.2025.11484898