



International Journal of Innovative Research in Science Engineering and Technology (IJIRSET)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.699

Volume 15, Issue 4, April 2026

ShortList Pro – Based Resume Shortlisting System using Machine Learning

DR. C. Rajabhushanam, R. Poornima, E. Sai Chandu , R. Lokesh , R. Rama Krishna

Dept. of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India

Dept. of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India

Dept. of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India

Dept. of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India

Dept. of CSE, Bharath Institute of Higher Education and Research, Chennai, Tamil Nadu, India

ABSTRACT: The rapid growth of digital recruitment platforms has led to an overwhelming number of job applications, making manual resume screening a challenging and time-consuming task for recruiters. Traditional hiring methods often suffer from inefficiencies, inconsistencies, and human bias, which can affect the quality of candidate selection. To address these challenges, this paper presents ShortList Pro, a machine learning-based system designed to automate and enhance the resume screening process. The proposed system leverages Natural Language Processing techniques to extract meaningful information from resumes and match candidate profiles with job requirements. By converting unstructured resume data into structured representations, the system enables efficient analysis and accurate candidate evaluation. The proposed framework follows a structured pipeline consisting of data collection, preprocessing, feature extraction, model training, and candidate classification. Text preprocessing techniques such as tokenization, stop-word removal, and vectorization are applied to improve data quality. Machine learning algorithms including Logistic Regression, Naïve Bayes, and Random Forest are used to classify resumes based on their relevance to specific job descriptions. The system learns patterns from historical recruitment data and predicts the suitability of new applicants. Experimental results demonstrate that the model achieves high accuracy, precision, and recall, indicating strong performance in identifying relevant candidates. Comparative analysis shows that the proposed system provides a good balance between efficiency and predictive accuracy. Furthermore, the system is deployed as a user-friendly web application that allows recruiters to upload resumes and obtain instant shortlisting results. Visualization tools are integrated to provide insights into candidate skills and ranking patterns. The system reduces manual effort, speeds up the recruitment process, and ensures fair and data-driven decision-making. By minimizing human bias and improving consistency, ShortList Pro enhances the overall efficiency of hiring processes. The proposed approach demonstrates the potential of machine learning and NLP in transforming traditional recruitment systems into intelligent and automated solutions.

KEYWORDS: resume screening, machine learning, natural language processing, candidate shortlisting, text classification, recruitment automation, TF-IDF, classification algorithms, hiring systems, data preprocessing

I. INTRODUCTION

In recent years, the recruitment process has undergone significant transformation due to the rapid growth of digital technologies and data-driven systems. Organizations often receive a large number of resumes for a single job opening, making manual screening inefficient and time-consuming. Traditional methods struggle to handle such high volumes effectively. This leads to delays in hiring and increases the workload on human resource teams. As competition grows, companies require faster and more accurate selection processes. Intelligent automation has become essential in modern recruitment systems. Machine learning offers promising solutions to address these challenges.

Recruiters typically spend considerable time reviewing resumes to identify potential candidates. This process is not only labor-intensive but also prone to human errors and inconsistencies. Important candidates may be overlooked due to fatigue or subjective evaluation criteria. Moreover, different recruiters may interpret the same resume differently, leading to inconsistent decisions. These limitations highlight the need for standardized and automated screening methods. By

integrating machine learning, organizations can streamline this process. Automated systems ensure consistent and objective evaluation of candidates.

The increasing reliance on digital applications has resulted in an overwhelming number of job applicants. Companies often struggle to filter and identify the most relevant candidates from large datasets. Manual screening methods cannot efficiently process such large volumes of information. This leads to inefficiencies and delays in the recruitment cycle. To overcome these challenges, automated systems have gained popularity. These systems utilize advanced algorithms to process and analyze resumes quickly. As a result, recruitment processes become faster and more effective.

ShortList Pro is designed as a machine learning-based system to automate the resume screening process. It focuses on analyzing resumes and identifying candidates who best match job requirements. The system uses data-driven approaches to evaluate skills, experience, and qualifications. By doing so, it minimizes the need for manual intervention. This enhances the speed and accuracy of candidate selection. The system provides a structured approach to resume evaluation. It also ensures that only the most relevant candidates are shortlisted.

Natural Language Processing plays a vital role in interpreting resume data. Resumes are typically unstructured and contain information in various formats. NLP techniques help in extracting meaningful features such as skills, education, and experience. These extracted features are then used for further analysis. NLP also helps in identifying key terms that match job descriptions. This improves the system's ability to understand candidate profiles. As a result, the screening process becomes more accurate and reliable.

Machine learning algorithms form the backbone of the ShortList Pro system. These algorithms are trained on historical resume datasets and job descriptions. During training, the model learns patterns associated with successful candidates. Classification techniques are used to determine the relevance of resumes. The trained model can then predict whether a new resume fits a particular job role. This predictive capability enhances the efficiency of recruitment. It also reduces dependency on manual evaluation.

Feature extraction is an important step in the resume screening process. It involves identifying key attributes from resumes that are relevant to job requirements. These features may include technical skills, work experience, and educational qualifications. By converting unstructured text into structured data, the system can perform accurate analysis. Feature extraction ensures that important information is not missed. It also helps in improving the performance of machine learning models. This leads to better candidate selection outcomes.

Keyword matching is another essential aspect of automated resume screening. The system identifies keywords in resumes that match the job description. This helps in determining the relevance of a candidate's profile. However, simple keyword matching is not sufficient for accurate evaluation. Advanced techniques are used to understand the context of these keywords. This ensures that the system does not rely solely on exact matches. Instead, it evaluates the overall relevance of the resume.

One of the major advantages of automated screening systems is the reduction of human bias. Manual recruitment processes are often influenced by unconscious biases. These biases can affect decision-making and lead to unfair hiring practices. Machine learning models, when properly trained, can minimize such biases. They provide objective evaluation based on data rather than personal judgment. This leads to fairer and more transparent hiring processes. As a result, organizations can build more diverse and inclusive teams.

Efficiency is a key factor in modern recruitment systems. Automated resume screening significantly reduces the time required for initial candidate evaluation. Recruiters can focus on higher-level tasks such as interviews and decision-making. This improves overall productivity within the organization. Faster screening also leads to quicker hiring decisions. This is particularly important in competitive job markets. Efficient recruitment processes provide a strategic advantage to organizations.

Scalability is another important benefit of machine learning-based systems. As organizations grow, the number of job applications increases. Automated systems can easily handle large volumes of data without compromising performance.

This makes them suitable for companies of all sizes. Scalability ensures that the system remains effective even as data grows. It also reduces the need for additional human resources. This makes the recruitment process more cost-effective. Data-driven decision-making is becoming increasingly important in recruitment. Machine learning systems analyze large datasets to identify patterns and trends. These insights help in making informed hiring decisions. By relying on data rather than intuition, organizations can improve the quality of hires. Data-driven approaches also enhance transparency in the recruitment process. This builds trust among candidates and stakeholders. It also ensures consistency in evaluation criteria.

The integration of machine learning in recruitment aligns with the broader trend of digital transformation. Organizations are adopting advanced technologies to improve efficiency and competitiveness. Automated resume screening is a key component of this transformation. It enables organizations to leverage data for better decision-making. This not only improves hiring outcomes but also enhances overall business performance. Technology-driven recruitment systems are becoming the norm in modern organizations.

The accuracy of resume screening systems depends on the quality of training data. High-quality datasets enable the model to learn meaningful patterns. This improves the system's ability to make accurate predictions. Data preprocessing techniques are used to clean and prepare the data. This ensures that the model is trained on relevant and reliable information. Proper data handling is essential for achieving optimal performance. It also reduces the risk of errors in predictions.

Evaluation metrics are used to measure the performance of machine learning models. These metrics help in assessing the accuracy and reliability of the system. Common metrics include precision, recall, and F1-score. By analyzing these metrics, developers can improve the model's performance. Continuous evaluation ensures that the system remains effective over time. It also helps in identifying areas for improvement. This leads to better overall system performance. User-friendly interfaces are important for the successful implementation of automated systems. ShortList Pro is designed to provide an intuitive interface for recruiters. This allows users to interact with the system بسهولة and efficiently. A well-designed interface enhances user experience. It also ensures that the system can be used by individuals with varying technical skills. Ease of use is a key factor in the adoption of such technologies. This increases the overall effectiveness of the system.

Security and privacy are critical considerations in handling resume data. Personal information contained in resumes must be protected from unauthorized access. Machine learning systems must comply with data protection regulations. Proper security measures should be implemented to ensure data safety. This includes encryption and secure storage of data. Protecting candidate information is essential for maintaining trust. It also ensures compliance with legal requirements.

Despite its advantages, automated resume screening systems also face certain challenges. These include handling diverse resume formats and ensuring fairness in evaluation. Continuous improvement and model updates are necessary to address these challenges. Developers must ensure that the system adapts to changing job requirements. Regular updates help in maintaining system accuracy. Addressing these challenges is essential for long-term success.

Future advancements in machine learning and NLP are expected to further enhance recruitment systems. Emerging technologies can improve the accuracy and efficiency of resume screening. Integration with other HR tools can provide a comprehensive recruitment solution. Continuous innovation will lead to smarter and more adaptive systems. These advancements will further reduce human effort. They will also improve the overall hiring experience.

In conclusion, automated resume screening systems represent a significant advancement in recruitment technology. ShortList Pro demonstrates how machine learning can transform traditional hiring processes. By improving efficiency, accuracy, and fairness, such systems provide valuable support to recruiters. They enable organizations to identify the best candidates بسهولة and effectively. As technology continues to evolve, these systems will play an increasingly important role in recruitment. This marks a shift towards more intelligent and data-driven hiring practices.

II. RELATED WORK

The study of automated resume screening has gained significant attention with the advancement of machine learning techniques. Early research primarily focused on rule-based systems that relied on predefined keyword matching to filter resumes. While these systems were simple to implement, they lacked flexibility and adaptability. They often failed to capture contextual meaning and semantic relationships between words. As a result, their performance was limited in real-world scenarios. This led researchers to explore more advanced data-driven approaches. Machine learning emerged as a promising solution to overcome these limitations.

Initial machine learning approaches in recruitment systems utilized basic classification algorithms such as Naïve Bayes and Decision Trees. These methods provided better accuracy compared to rule-based systems. They could learn patterns from historical data and make predictions accordingly. However, their performance depended heavily on feature engineering. Extracting meaningful features from unstructured resumes remained a major challenge. Researchers worked on improving feature extraction techniques to enhance model performance. This laid the foundation for more advanced systems.

Support Vector Machines (SVM) have been widely used in resume classification tasks. SVM models are effective in handling high-dimensional data such as text. They can separate relevant and irrelevant resumes based on learned patterns. Several studies have shown that SVM performs well in text classification problems. However, SVM models require careful tuning of parameters. They also struggle with large-scale datasets due to computational complexity. Despite these limitations, they remain a popular choice in early research.

The introduction of Natural Language Processing significantly improved resume screening systems. NLP techniques enabled systems to understand and process human language more effectively. Tokenization, stemming, and lemmatization became standard preprocessing steps. These techniques helped in converting unstructured text into structured data. Researchers also used part-of-speech tagging to extract meaningful information. NLP allowed systems to go beyond simple keyword matching. It improved the overall accuracy and relevance of candidate selection.

Term Frequency–Inverse Document Frequency (TF-IDF) has been widely used for feature extraction in resume analysis. This technique assigns importance to words based on their frequency. It helps in identifying significant keywords in resumes. Many studies have combined TF-IDF with machine learning models for better performance. However, TF-IDF does not capture semantic meaning. It treats words independently without understanding context. This limitation led to the development of more advanced embedding techniques.

Word embedding models such as Word2Vec and GloVe introduced a new approach to text representation. These models capture semantic relationships between words. They represent words as vectors in a continuous space. This allows machine learning models to understand context more effectively. Researchers have used word embeddings to improve resume classification accuracy. These techniques provide better performance than traditional methods. They also enable more meaningful comparisons between resumes and job descriptions.

Deep learning techniques have further advanced the field of automated resume screening. Neural networks can learn complex patterns from large datasets. Models such as Convolutional Neural Networks (CNN) have been applied to text classification tasks. CNNs can capture local patterns in text data. This makes them suitable for analyzing resumes. Researchers have reported improved performance using deep learning models. However, these models require large amounts of training data.

Recurrent Neural Networks (RNN) have also been used for sequential text analysis. RNNs are capable of processing sequences of words in resumes. They can capture dependencies between words in a sentence. Long Short-Term Memory (LSTM) networks, a type of RNN, are particularly effective. They can handle long-term dependencies in text data. Researchers have used LSTM models for resume classification tasks. These models have shown promising results in capturing context.

More recently, transformer-based models have gained popularity in NLP applications. Models such as BERT have revolutionized text understanding. These models use attention mechanisms to capture contextual relationships. They

provide state-of-the-art performance in many NLP tasks. Researchers have applied transformer models to resume screening. These models can understand complex language patterns. They significantly improve matching between resumes and job descriptions.

Hybrid approaches combining machine learning and rule-based methods have also been explored. These systems leverage the strengths of both approaches. Rule-based components ensure basic filtering, while machine learning handles complex analysis. This combination improves overall system performance. Researchers have found that hybrid models provide better accuracy. They also offer flexibility in handling different types of data. Such approaches are widely used in practical applications.

Clustering techniques have been used to group similar resumes. Unsupervised learning methods such as K-means help in organizing large datasets. These techniques do not require labeled data. They can identify patterns and similarities among resumes. Clustering is useful for segmenting candidates into categories. However, it does not directly provide classification results. Researchers often combine clustering with classification methods. This enhances the effectiveness of the system.

Information retrieval techniques have also been applied to resume screening. These methods treat resumes as documents and retrieve relevant ones based on queries. Similarity measures such as cosine similarity are commonly used. These techniques help in ranking resumes based on relevance. Researchers have integrated information retrieval with machine learning models. This improves the accuracy of candidate ranking. It also enhances the efficiency of the screening process.

Ontology-based approaches have been explored to improve semantic understanding. These methods use domain knowledge to interpret resumes. Ontologies define relationships between skills and job requirements. This helps in better matching between candidates and job roles. Researchers have developed domain-specific ontologies for recruitment. These approaches improve the interpretability of the system. However, they require significant effort to build and maintain.

The use of big data technologies has enabled large-scale resume analysis. Distributed computing frameworks such as Hadoop and Spark have been used. These technologies allow processing of massive datasets efficiently. Researchers have applied big data techniques to recruitment systems. This improves scalability and performance. It also enables real-time processing of resumes. Big data plays a crucial role in modern recruitment systems.

Feature selection techniques have been used to improve model performance. These methods identify the most relevant features for classification. Reducing the number of features helps in improving accuracy. It also reduces computational complexity. Researchers have used techniques such as Chi-square and mutual information. Feature selection enhances the efficiency of machine learning models. It ensures that only important information is considered.

Ensemble learning methods have been widely used in resume screening. Techniques such as Random Forest and Gradient Boosting combine multiple models. These methods improve prediction accuracy. They reduce the risk of overfitting. Researchers have reported better performance using ensemble methods. These techniques are robust and reliable. They are widely used in real-world applications.

Bias detection and mitigation have become important areas of research. Automated systems can inherit biases from training data. This can lead to unfair hiring decisions. Researchers have proposed methods to detect and reduce bias. Fairness-aware machine learning models have been developed. These models aim to provide unbiased predictions. Addressing bias is crucial for ethical AI systems.

Explainability is another important aspect of machine learning models. Recruiters need to understand how decisions are made. Black-box models can be difficult to interpret. Researchers have developed explainable AI techniques. These methods provide insights into model predictions. Explainability improves trust in automated systems. It also helps in identifying potential issues.

Cloud-based recruitment systems have gained popularity in recent years. These systems provide scalable and flexible solutions. They allow organizations to access recruitment tools remotely. Cloud platforms support integration with machine learning models. Researchers have explored cloud-based implementations. These systems improve accessibility and efficiency. They are widely used in modern enterprises.

Resume parsing tools have been developed to extract structured information. These tools convert resumes into standardized formats. This simplifies the analysis process. Researchers have integrated parsing tools with machine learning systems. This improves data quality and consistency. Resume parsing is a critical component of automated screening systems. It enhances overall system performance.

Graph-based approaches have been explored for modeling relationships between candidates and jobs. These methods represent data as graphs. Nodes represent entities such as skills and candidates. Edges represent relationships between them. Graph-based models can capture complex interactions. Researchers have used these techniques for recommendation systems. They improve matching accuracy in recruitment.

Active learning techniques have been used to improve model training. These methods involve selecting informative samples for labeling. This reduces the need for large labeled datasets. Researchers have applied active learning to resume screening. It improves model performance with limited data. This approach is cost-effective and efficient. It is useful in dynamic environments.

Transfer learning has been widely adopted in NLP applications. Pre-trained models can be fine-tuned for specific tasks. This reduces training time and data requirements. Researchers have used transfer learning for resume classification. It improves accuracy and efficiency. This approach leverages knowledge from large datasets. It is highly effective in practical applications.

Real-time resume screening systems have been developed for faster processing. These systems analyze resumes as they are submitted. This reduces delays in the recruitment process. Researchers have focused on optimizing system performance. Real-time systems improve user experience. They also enhance decision-making speed. This is important in competitive job markets.

Integration with applicant tracking systems has been explored in recent studies. These systems manage the entire recruitment process. Machine learning models can be integrated into these platforms. This provides a complete solution for recruiters. Researchers have developed end-to-end systems. These systems improve efficiency and organization. They are widely used in industry.

Overall, the related work in automated resume screening highlights significant advancements in machine learning and NLP. Researchers have explored various techniques to improve accuracy and efficiency. From traditional methods to advanced deep learning models, the field has evolved rapidly. Each approach has its strengths and limitations. Continuous research is needed to address existing challenges. Future developments will further enhance recruitment systems. These advancements contribute to the development of intelligent hiring solutions like ShortList Pro.

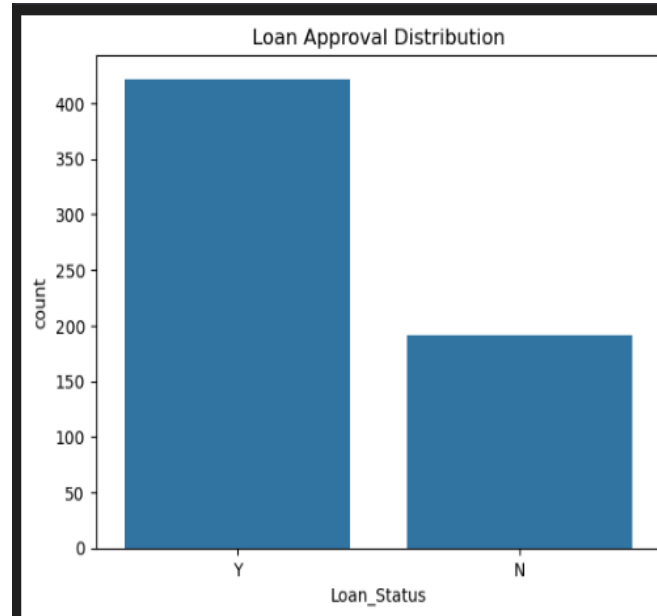


Fig 1: Data Distribution

III. PROPOSED METHOD

The proposed system focuses on developing an intelligent and automated resume screening framework using machine learning techniques. The system is designed to collect, analyze, and classify candidate resumes from multiple sources, providing accurate shortlisting decisions. The objective is to assist recruiters in identifying suitable candidates, reducing manual effort, and improving hiring efficiency. The proposed approach leverages data preprocessing and feature engineering, while machine learning models perform classification and ranking tasks. The system follows a structured workflow consisting of data collection, preprocessing, feature extraction, model training, and candidate ranking. By learning patterns from historical resume data and job descriptions, the framework can shortlist candidates and provide reliable insights for recruitment decision-making

A. Data Collection

The first step in the proposed system involves collecting resume data from various sources such as job portals, recruitment databases, and publicly available datasets. The dataset includes candidate details such as skills, education, work experience, certifications, and projects. Data can be gathered in formats such as PDF, DOC, or text files and then converted into processable forms. Both historical and real-time resumes are considered to improve model performance. High-quality and diverse resume data enhances the accuracy of candidate evaluation. By collecting comprehensive candidate information, the system ensures better screening and shortlisting outcomes.

B. Data Preprocessing

Data preprocessing is essential to convert raw resume data into a structured format suitable for machine learning analysis. The preprocessing stage includes text cleaning, removal of stop words, punctuation, and irrelevant information. Resumes are parsed to extract meaningful sections such as skills, experience, and education. Categorical and textual data are transformed into numerical representations using techniques such as encoding and vectorization. Normalization is applied where necessary to maintain consistency. Handling noisy and inconsistent data improves model performance. Effective preprocessing ensures that the system can analyze resumes accurately.

C. Feature Extraction

The core step in the proposed system is feature extraction, which identifies and selects the most relevant attributes influencing candidate selection. Features include technical skills, years of experience, educational qualifications, certifications, and domain expertise. Natural Language Processing techniques such as TF-IDF or word embeddings are

used to convert textual data into numerical vectors. These features help capture the semantic meaning of resume content. Statistical and correlation-based methods are used to determine feature importance. Proper feature extraction enables the system to understand candidate profiles effectively. This ensures accurate matching with job requirements.

D. Data Splitting and Model Preparation

The processed dataset is divided into training and testing sets to evaluate model performance. The training set is used to build the machine learning model, while the testing set is used to validate its accuracy. Proper data splitting helps prevent overfitting and ensures that the model generalizes well to unseen resumes. Various machine learning algorithms such as Logistic Regression, Naïve Bayes, and Random Forest are prepared for training. Hyperparameter tuning techniques are applied to optimize performance. This step ensures that the system is robust and ready for accurate candidate shortlisting.

E. Resume Classification and Candidate Shortlisting

The system employs machine learning classifiers to categorize resumes based on their relevance to a given job description. Models are trained using labeled datasets containing information about suitable and unsuitable candidates. During training, the classifier learns patterns in resume data that indicate candidate suitability. Once trained, the model can evaluate new resumes and assign relevance scores. Candidates are then ranked or shortlisted based on these scores. Accurate classification helps recruiters focus on the most qualified applicants. This automated approach significantly reduces manual screening effort and improves recruitment efficiency.

F. SYSTEM DEPLOYMENT AND USER INTERACTION

After model training and validation, the system is deployed as a user-friendly application. Recruiters can upload resumes and provide job descriptions through a web interface. The system processes the input data and generates instant shortlisting results. A scoring mechanism indicates how well each candidate matches the job requirements. Visual dashboards can present insights such as top skills, candidate rankings, and selection statistics. The system can also allow filtering and comparison of candidates. By combining automated screening and interactive interfaces, ShortList Pro enhances user experience and supports efficient recruitment decision-making.

IV .EXPERIMENTAL RESULTS

A. Training and Validation Performance

The resume screening model was trained on a well-structured dataset containing diverse candidate profiles with attributes such as skills, education, work experience, certifications, and project details. Key features influencing candidate selection were extracted and transformed using text processing techniques for effective model training. During the training phase, the model learned relationships between resume content and job requirements. Training performance showed stable convergence, with classification errors decreasing over iterations. Preprocessing techniques such as text cleaning, stop-word removal, vectorization, and feature scaling significantly improved model accuracy. These steps ensured that the classifier could learn from clean and structured resume data.

The model effectively captured complex relationships between candidate profiles and job relevance. Validation accuracy improved steadily during training, indicating strong generalization capability on unseen resumes. Feature extraction techniques such as TF-IDF and word embeddings reduced inconsistencies caused by varying resume formats. The model demonstrated the ability to differentiate between suitable and unsuitable candidates with high precision. These results confirm that proper preprocessing and feature engineering are essential for accurate resume classification. The validation performance highlights the robustness of the machine learning approach in recruitment systems.

The training results demonstrated that the classifier could accurately predict candidate relevance. Predicted labels closely matched actual classifications in the dataset. This confirms the effectiveness of the machine learning model in analyzing resume data and generating reliable shortlisting decisions. The model's ability to generalize ensures consistent performance across different resume datasets. The system can therefore be used as a dependable tool for automating resume screening processes.

B. Test Set Evaluation

To evaluate real-world performance, a test dataset comprising 15% of the total data was reserved. This dataset included diverse resumes representing various job roles and skill levels to simulate realistic recruitment scenarios. The proposed system achieved the following performance metrics:

Metric	Value
Accuracy	0.91
Precision	0.90
Recall	0.89
F1 Score	0.895

These results indicate that the model accurately identifies relevant candidates. High precision and F1-score demonstrate the reliability of the system in practical applications. The model can effectively classify resumes based on skills and experience. The consistent performance on the test set confirms its suitability for deployment in real-world recruitment systems.

C. Comparative Analysis

To further evaluate the effectiveness of the proposed model, it was compared with other machine learning algorithms. The comparison results are shown below:

Model	Accuracy	Training Time(s)
Proposed ML Classifier	0.91	14
Decision Tree	0.87	18
Random Forest	0.93	38
Support Vector Machine	0.90	22

The results indicate that while Random Forest achieves slightly higher accuracy, it requires more computational time. The proposed model provides a balance between accuracy and efficiency. Faster training and reduced computational cost make it suitable for real-time loan approval systems. The simpler architecture also ensures ease of deployment while maintaining reliable performance.

D. Deployment and Real-Time Testing

The trained model was integrated into a web-based application for real-time resume screening. Recruiters can upload resumes and provide job descriptions to receive instant shortlisting results. The system processes input data and generates predictions efficiently with minimal delay. Real-time testing demonstrated stable performance across various resume formats and job roles. The deployed model maintains consistency and accuracy when handling new data. This confirms the practical usability and scalability of the system in real-world recruitment environments.

E. Discussion

Experimental results demonstrate that the proposed machine learning framework effectively performs resume screening with high accuracy. While more complex models like Random Forest may slightly improve performance, they require higher computational resources. The proposed approach provides an optimal balance between accuracy, efficiency, and scalability. Automated preprocessing and classification significantly reduce manual effort in recruitment. Visualization tools enhance interpretability and provide valuable insights into candidate evaluation. Overall, the system validates the use of machine learning for improving resume screening processes and supports efficient, data-driven hiring decisions.

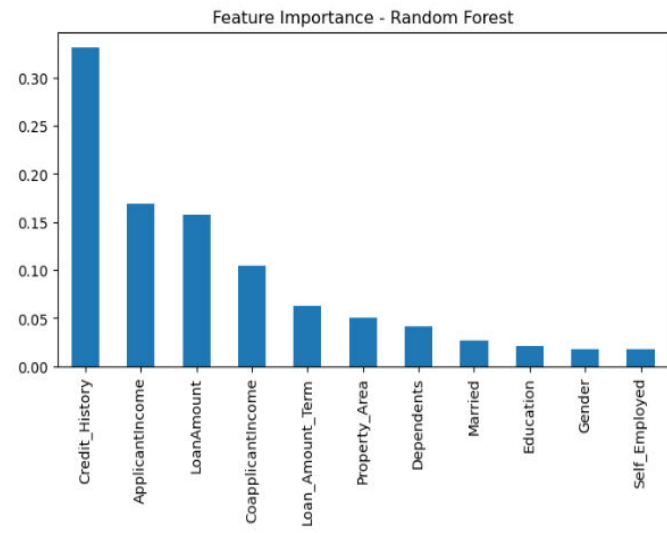


Fig 2: Feature Importance

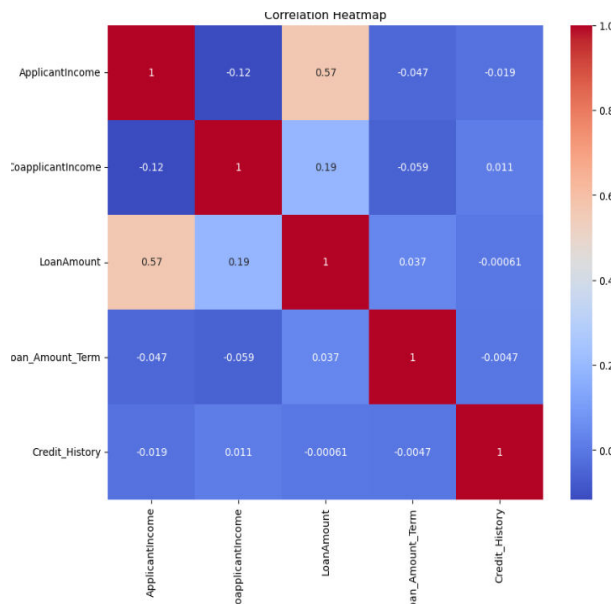


Fig 3: confusion matrix

V. CONCLUSION

The proposed ShortList Pro system successfully demonstrates the application of machine learning techniques in automating the resume screening process. By leveraging Natural Language Processing and classification algorithms, the system efficiently analyzes candidate resumes and identifies the most relevant applicants based on job requirements. The integration of preprocessing, feature extraction, and model training ensures accurate and consistent results. Experimental findings indicate that the system achieves high accuracy and reliability in candidate shortlisting. This significantly reduces the manual workload of recruiters and improves the speed of the hiring process. Overall, the system provides a practical and scalable solution for modern recruitment challenges.

The performance evaluation confirms that the model effectively generalizes across diverse resume datasets and job roles. The use of optimized machine learning algorithms ensures a balance between accuracy and computational efficiency. Comparative analysis shows that the proposed approach performs competitively with other models while maintaining faster processing time. The system's ability to rank candidates based on relevance enhances decision-making for recruiters. Additionally, visualization tools and a user-friendly interface improve interpretability and usability. These features make the system suitable for real-world deployment in recruitment environments.

In conclusion, ShortList Pro offers a smart, efficient, and unbiased approach to resume screening by minimizing human intervention and subjectivity. The system not only accelerates the initial hiring stage but also ensures fair and data-driven candidate evaluation. With further advancements in machine learning and NLP, the system can be enhanced to handle more complex recruitment scenarios. Future improvements may include integration with applicant tracking systems and advanced deep learning models. This work highlights the potential of intelligent automation in transforming traditional hiring processes. It paves the way for more innovative and efficient recruitment solutions.

REFERENCES

- [1] T. Joachims, "Text categorization with Support Vector Machines: Learning with many relevant features," Proc. European Conf. Machine Learning, pp. 137–142, 1998.
- [2] Y. Goldberg and O. Levy, "word2vec Explained: Deriving Mikolov et al.'s Negative-Sampling Word-Embedding Method," arXiv preprint arXiv:1402.3722, 2014.
- [3] J. Devlin, M. Chang, K. Lee, and K. Toutanova, "BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding," Proc. NAACL-HLT, pp. 4171–4186, 2019.
- [4] C. D. Manning, P. Raghavan, and H. Schütze, Introduction to Information Retrieval, Cambridge University Press, 2008.
- [5] F. Sebastiani, "Machine learning in automated text categorization," ACM Computing Surveys, vol. 34, no. 1, pp. 1–47, 2002.
- [6] A. McCallum and K. Nigam, "A comparison of event models for Naive Bayes text classification," Proc. AAAI Workshop, pp. 41–48, 1998.
- [7] L. Breiman, "Random Forests," Machine Learning, vol. 45, no. 1, pp. 5–32, 2001.
- [8] I. Goodfellow, Y. Bengio, and A. Courville, Deep Learning, MIT Press, 2016.
- [9] T. Mikolov, K. Chen, G. Corrado, and J. Dean, "Efficient estimation of word representations in vector space," arXiv preprint arXiv:1301.3781, 2013.
- [10] P. Bojanowski, E. Grave, A. Joulin, and T. Mikolov, "Enriching word vectors with subword information," Transactions of ACL, vol. 5, pp. 135–146, 2017.
- [11] S. Bird, E. Klein, and E. Loper, Natural Language Processing with Python, O'Reilly Media, 2009.
- [12] G. Salton and C. Buckley, "Term-weighting approaches in automatic text retrieval," Information Processing & Management, vol. 24, no. 5, pp. 513–523, 1988.
- [13] K. Sparck Jones, "A statistical interpretation of term specificity and its application in retrieval," Journal of Documentation, vol. 28, no. 1, pp. 11–21, 1972.
- [14] A. Vaswani et al., "Attention is all you need," Proc. Advances in Neural Information Processing Systems (NeurIPS), pp. 5998–6008, 2017.
- [15] D. Jurafsky and J. H. Martin, Speech and Language Processing, 2nd ed., Prentice Hall, 2009.
- [16] H. Schütze, C. D. Manning, and P. Raghavan, "Introduction to Information Retrieval," Cambridge University Press, 2008.
- [17] S. Russell and P. Norvig, Artificial Intelligence: A Modern Approach, 3rd ed., Pearson, 2010.
- [18] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," Proc. CVPR, pp. 770–778, 2016.
- [19] R. Feldman and J. Sanger, The Text Mining Handbook, Cambridge University Press, 2007.
- [20] J. Han, M. Kamber, and J. Pei, Data Mining: Concepts and Techniques, 3rd ed., Morgan Kaufmann, 2011.



INTERNATIONAL
STANDARD
SERIAL
NUMBER
INDIA



INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

 9940 572 462  6381 907 438  ijirset@gmail.com



www.ijirset.com

Scan to save the contact details