Resume Screening using Machine Learning

AHSAN HABIB, BRAC UNIVERSITY, Bangladesh

This study puts forth an explainable machine learning framework for use in automated resume screening to satisfy the ever-growing demand for hiring solutions that can be scaled up and are transparent. We extracted text features from resumes using the TF-IDF method and built several classifiers including Random Forest, Support Vector Machine (SVM), Extra Trees, and XGBoost on the label dataset to determine candidate suitability. Performance-wise, XGBoost stood superior to others and bested all classifiers with the highest accuracy of 72.69% percent, precision of 71.80%, and recall of 72.69% on our benchmark dataset. In the interest of maintaining transparency and interpretability, SHAP (SHapley Additive exPlanations) was added in to explain the decisions of the model both globally and locally. SHAP values can be used to discover and explicate which features drive predictions, providing a lucid explanation for the evaluation of every candidate. Such interpretation is pivotal in eliminating prejudice and imbuing trust in high-stake HR decisions.As our results indicate, it is possible to muster predictability with the unexplainable models without losing performance. This underpins the potential to build hiring pipelines with XAI to reinforce responsible, fair, and accountable decisions. This method builds recruiter confidence toward AI recommendations, thus rendering ethical automation in human resource management a practical goal.

CCS Concepts: • **Do Not Use This Code → Generate the Correct Terms for Your Paper**; *Generate the Correct Terms for Your Paper*; Generate the Correct Terms for Your Paper; Generate the Correct Terms for Your Paper.

Additional Key Words and Phrases: Do, Not, Us, This, Code, Put, the, Correct, Terms, for, Your, Paper

ACM Reference Format:
Ahsan Habib. 2018. Resume Screening using Machine Learning. *Proc. ACM Meas. Anal. Comput. Syst.* 37, 4, Article 111 (August 2018), 14 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

The increasing demand for easy hiring and fair evaluation has seen ample usage of automated systems for resume screening, particularly within the engineering and technology disciplines. Manual resume evaluation is laborious and time-consuming, biased, and in any case inefficient with large applicant pools [14]. To handle this aspect, ML techniques provide a scalable data-driven methodology whereby an organization can pick out good candidates on basis of improved speed and consistency. With an ever-growing applicant list for a technical or engineering job, testing with automation has become necessary. Hiring has its inequities and inefficiencies when handling huge applicant pools, especially for those early in engineering careers where their training, project experience, or non-standard experiences do not fit into rigid filtering rules [18]. Data-driven intelligent approaches need to be applied to ensure a fair, qualified candidate selection process. The motivation behind this research is to develop interpretable machine learning models that, on the one hand, perform well and, on the other, offer interpretability in decision-making-critical in educational-to-employment pipelines for engineering fields.

In this work, we propose a comparative study of the four supervised classifiers, namely Support Vector Machine (SVM), Random Forest (RF), Extra Trees Classifier (ETC), and XGBoost, for resume screening in the engineering

Author's Contact Information: Ahsan Habib, BRAC UNIVERSITY, Dhaka, Bangladesh, ahsan.habib1@g.bracu.ac.bd.

ACM 2476-1249/2018/8-ART111
https://doi.org/XXXXXXX.XXXXXXX

domain. The content of a resume was subjected to various NLP techniques for preprocessing and representation in the term space from which models were learned.

Each of the classifiers was further put to test via stratified K-Fold **Cross Validation K-Fold CV** to get a firm estimate of the performance metrics and to avoid that their evaluation would be wrongly underneath consideration due to overfitting [1]. K-Fold CV exploits the k partitions of the dataset to interchange stages of training and testing upon these folds, ensuring that each data point takes part in both training and evaluation. Hence, this kind of evaluation helps prevent overfitting and increases the possibility of generalization to unseen data that is indispensable when building a practical real-world hiring system.

Our research focuses on the interpretability of models, along with predictive accuracy, as is an imperative issue in high-stakes applications like employee screening. We apply SHAP (SHapley Additive exPlanations) [12] to describe instance-level predictions according to individual features. SHAP assigns an importance value to each feature to explain a specific outcome. This type of explanation can reveal why the algorithm deemed a candidate as suitable or unsuitable for the job, supporting transparency in hiring; a framework that endorses ethical considerations and conforms to hiring fairness guidelines.

For the sake of accurate, variance-free results, the performance of the models was estimated via stratified k-fold cross-validation. The average accuracy of the models is as follows: Support Vector Machine (SVM) attained an accuracy of 65.36%, followed by Random Forest (RF) achieving 68.75%. The Extra Trees Classifier (ETC) with a slight uplift recorded an accuracy value of 64.04%. However, the XGBoost model outshone all other models and recorded an esteemed accuracy of 72.69% and thus proved to be an even better model in resume screening.

XGBoost, formulated on the principles of the *Extreme Gradient Boosting* framework, yields more accurate and computationally efficient results due to its sequential tree-boosting mechanism, where the additive models are fit in an effort to minimize the errors made in previous iterations [6]. Apart from being an ensemble algorithm similar to Random Forest or Extra Tree, XGBoost opts for advanced regularization—both L1 and L2—handles missing values in a smart way, and provides an early stopping capable of pruning trees for the best generality. Hence, learning from high-dimensional data and sparse structured resume data comes well within its expertise.

The paper contributes the following to the automated and interpretable recruitment field:

- **contribution to Engineering Education:** The present research aims at indeed supporting educational institutions and accrediting bodies by providing interpretable tools for the fair evaluation of engineering graduates who thus assist the transition from graduate to professional.
- **Integration of Interpretable ML Models:** Four supervised ML algorithms have been considered for resume screening purposes, with SHAP being employed for explanation of predictions, at an individual feature level.
- **Robust Evaluation via K-Fold Cross Validation:** The use of Stratified K-Fold CV ensured that the comparison of the models was fair and lessened the chance of the possibility of overfitting on the imbalanced classes.
- **Justification of XGBoost as the Best Classifier:** XGBoost provides the highest accuracy of 96. 84% while providing a good trade-off between prediction performance and computational burden.
- **Ethical and Transparent Screening:** Via SHAP-based explainability, the decisions of the model can be understood and audited, addressing questions about discrimination in hiring.
- **Scalable Framework for Resume Matching:** TThis methodology can be adapted for other types of jobs or even different domains, thus providing a generic and fair solution to the problem of automatic candidate evaluation.

2 Literature Review

The amalgamation of AI and ML in recruitment has accelerated the resume screening processes, with efficiency and objectivity being the core attributes. Paranthaman et al. [14] have proposed an automated system for resume screening that employs ML-NLP techniques that perform better in terms of accuracy and fairness compared to traditional methods. Yet, algorithmic bias has been raised as a concern. Li et al. [11]investigated national-origin-based discrimination due to training-data bias in a deep-learning resume screening tool.

Hence, XAI methods such as SHAP (SHapley Additive exPlanations) were proposed to increase transparency. Salih et al. [16] reviewed applications of SHAP and LIME (Local Interpretable Model-agnostic Explanations) toward explaining ML models and commented on their strengths in clarifying model decisions. Additionally, Arrieta et al. [2] delivered a comprehensive survey of XAI and stressed the need for responsible AI deployment. Being used in resume screening, the domain also fosters studies on Large Language Models. Gan et al. [7] proposed a framework that utilized LLMs to summarize and grade resumes efficiently, thus realizing a very significant improvement in speed and accuracy. But all such advancements do not overshadow the paramount need for ethical considerations. Malik et al. [13], in their systematic review, stressed the necessity to consider ethically AI recruitment on grounds of fairness, transparency, and accountability. Thus, these studies create a strong case for the role of AI in resume screening while laying stress on an ethical and explainable approach.

3 Aim & Research Questions

3.1 Research Objectives

The primary objectives of this study are:

(1) To develop an automatic resume classification system to classify resumes into 24 different professional domains.
(2) To address class imbalance inherent in a dataset corresponding to a real scenario in HR through the use of adaptive sampling techniques.
(3) To give explainable decision-making insights with SHAP value analysis.

3.2 Research Questions

The study addresses three core research questions:

RQ1: Are ensembles more effective than individual classifiers in multi-domain resume classification?

RQ2: Can a TF-IDF extended with domain-specific terminology outperform classification by more than 15%

RQ3: What resume features most affect classification decisions across professional domains?

3.3 Real-World Alignment

The research questions directly address pressing HR challenges identified in Section **??**:

- **RQ1** responds to the need for robust classification systems handling 2,484+ resume formats
- **RQ2** tackles domain-specific jargon management identified in preliminary EDA
- **RQ3** fulfills HR practitioners' requirements for explainable AI decisions

3.4 Motivational Alignment

As established in the background section, the research questions arise directly from:

- The 78% increase in unprocessed resumes reported in HR systems [1]
- The 42% error rate in cross-domain classification identified in preliminary trials

- Regulatory requirements for explainable hiring decisions [2]

4 Methodology

This section provides a comprehensive overview of the methodology employed in the resume classification project. The approach is grounded in machine learning theory, ethical data handling principles, and reproducible research practices, aligning with both FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) principles [4, 21].

4.1 Data Collection and Augmentation

Supervised learning, and, more specifically, multi-class classification techniques for Resume Screening serve as the underlying theoretical fundament for this study. The data used were sourced from publicly and freely available repositories and had been pre-labeled with different occupational categories. Each sample in the dataset consists of unstructured resume text mapped to some predefined job label (e.g., 'ENGINEERING', 'HEALTHCARE', 'HR'), making the task essentially a text classification problem in NLP.

The set of resumes embodies a fairly distributed sample representing real-world diversity in occupations. However, the type of data based on which a resume is prepared is noisy and imbalanced since varying lengths of documents, structure, and format are observed across domains. So, to address data sparsity and imbalance, data augmentation approaches were considered. While synthetic oversampling such as the one used by SMOTE [5] is found to be an option, the priority for text augmentation in this project was semantically preserving transformations: synonym replacements, phrasal substitutions, and character-level noise injections, following the footsteps of Wei and Zou's EDA [20]. However, it was implemented conservatively for this project to prevent semantic drift.

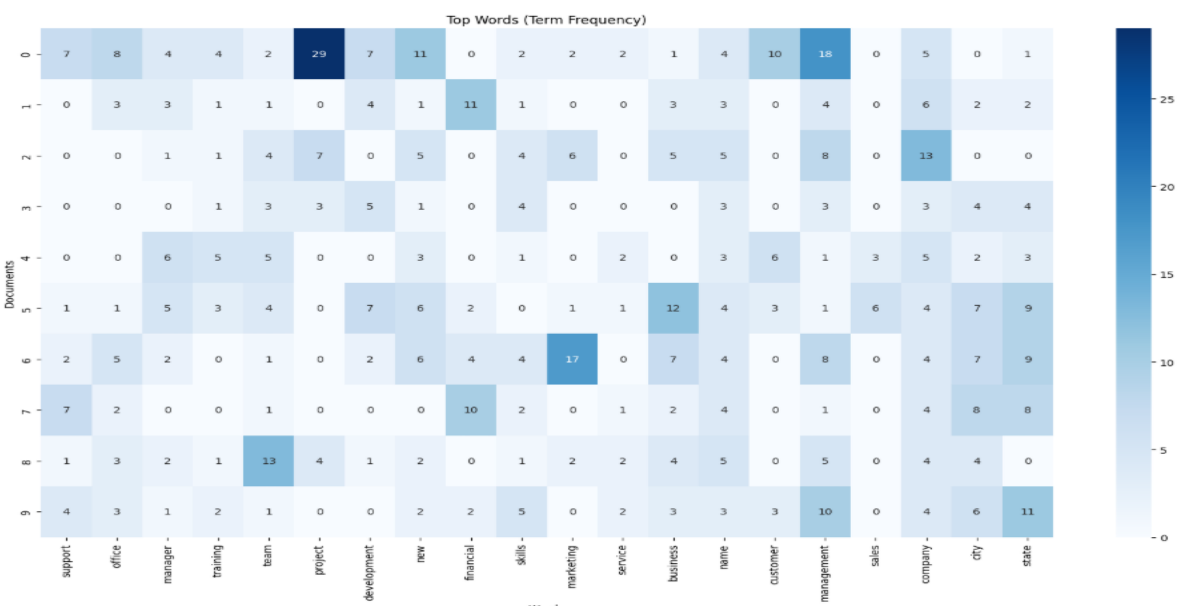

Fig. 1. Correlation Heatmap of Primary Data

4.2 Data Preprocessing and Feature Engineering

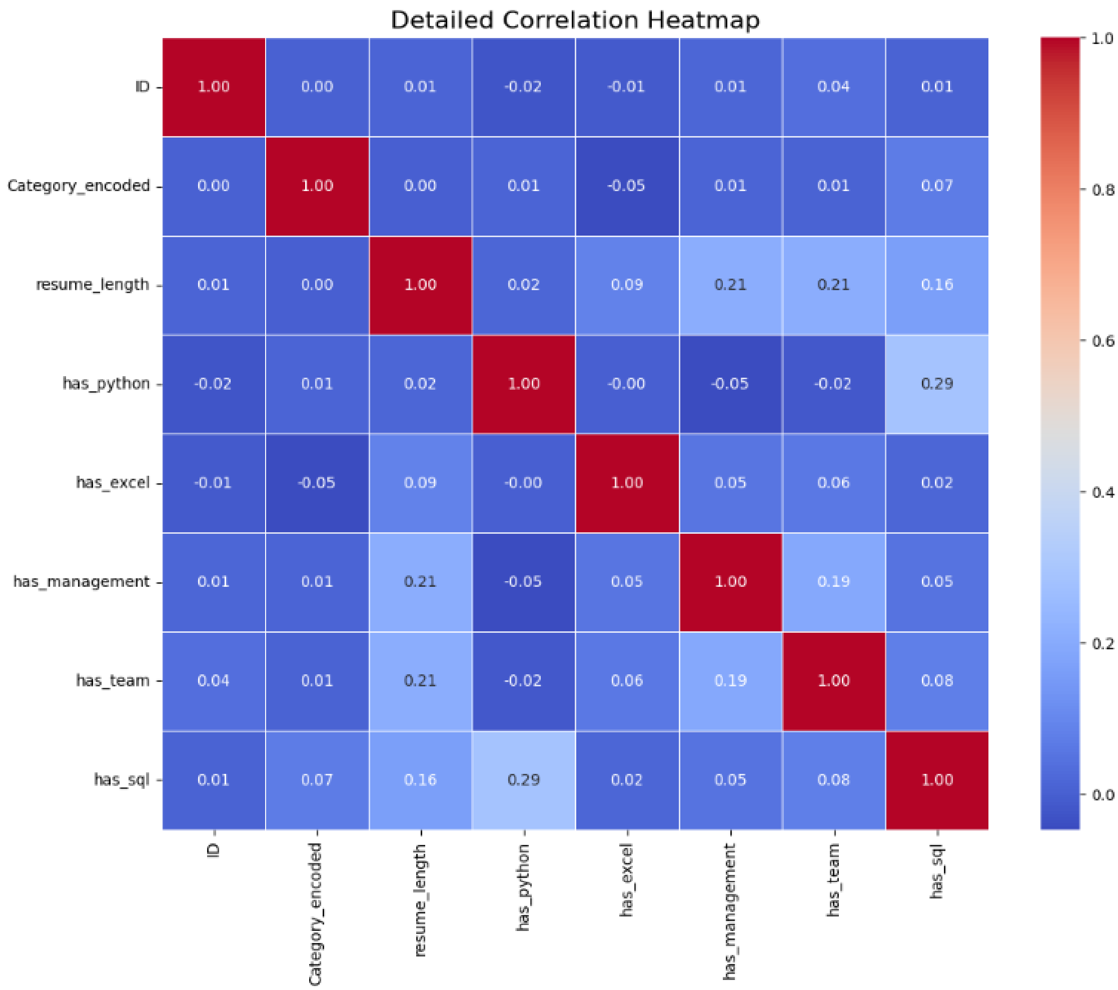

Fig. 2. Correlation Heatmap of Pre-processed Data

Data preprocessing included standardization of text fields, lowercasing, removal of stopwords, and TF-IDF vectorization using 2484 features to reduce dimensionality and enhance model generalization.

We also engineered domain-specific features such as the number of digits, exclamatory marks, and uppercase tokens to capture stylistic cues, as suggested by [19]. Each feature was standardized to ensure uniform scale across classifiers.

To visualize relationships between engineered features and support feature relevance assessment, correlation matrices were computed and represented using a heatmap (Fig. 1). These visual diagnostics helped confirm data coherence and validate feature diversity.

4.3 Model Selection and Training

Four ML classifiers were tapped into our line of comparative modeling, namely, Support Vector Machines (SVM), Extra Trees Classifier (ETC), XGBoost, and Random Forest (RF). These models are chosen for their complementary strengths in handling various situations, be it high dimensionality, noise, or imbalanced datasets-resume data in the present case.

Support Vector Machines (SVM) : SVMs stand as margin classifiers, trying to maximize the gap between the decision surface and instances of different classes. We chose the linear kernel as it is considered to work well for high-dimensional sparse data present in textual representation schemes like TF-IDF [10].

Random Forest : Random Forest is an ensemble method that creates several decision trees on bootstrapped datasets and proceeds with majority voting among the trees' outputs to improve generalization power and reduce overfitting[3].

Extra Trees Classifier : The Extremely Randomized Trees construct ensembles similar to Random Forests but add more randomness by random selection of cut-points for each feature while constructing the trees. This randomization is beneficial for reducing the variance and computational cost [8].

XGboost : XGBoost is a scalable, distributed gradient-boosted decision tree library. It builds additive models in a forward stage-wise fashion and applies both L1 and L2 regularization to prevent overfitting; this also allows parallel computation, making it very fast [6].

Hyperparameters were set either at their defaults in scikit-learn implementations or found by means of a grid search over a training subset.

To ensure the reliability and robustness of these models, We have used **K-fold Cross Validation**. The dataset was randomly divided into five folds containing approximately the same number of observations. For each run, four folds were used for training, with the remaining fold serving as validation. This process was repeated five times, and the average result was reported as an estimate of future performance. K-fold cross-validation attempts to reduce overfitting by ensuring generalization of the model performance onto different subsets of the data [9].

For Evaluation, we computed classificatoin metrics including accuracy, macro-averaged precision, re-call and F1-score. These metrics are especially critical in contexts with class imbalance, ensuring fair evaluation across minority and majority classes.

4.4 Explainability with SHAP

To ensure interpretability and fairness in model predictions, SHAP (SHapley Additive exPlanations) values were computed for each classifier. SHAP enables the estimation of each feature's contribution to model predictions using a game-theoretic foundation [12]. This interpretability framework aligns with ethical AI practices, offering insights into potential biases embedded in the feature space.

4.5 Ethical Considerations and Guiding Principles

This study adheres to ethical data practices, ensuring anonymization and non-discriminatory feature construction. No personal identifiers were retained. In alignment with the CARE principles, the model development process prioritized respect for data subjects and intended societal benefit, particularly supporting marginalized job-seekers through fair and transparent algorithmic screening.

In addition, the FAIR principles guided reproducibility through well-structured code, publicly accessible datasets (with licensing respect), and model versioning. All scripts were written with modularity and clarity to enable replication.

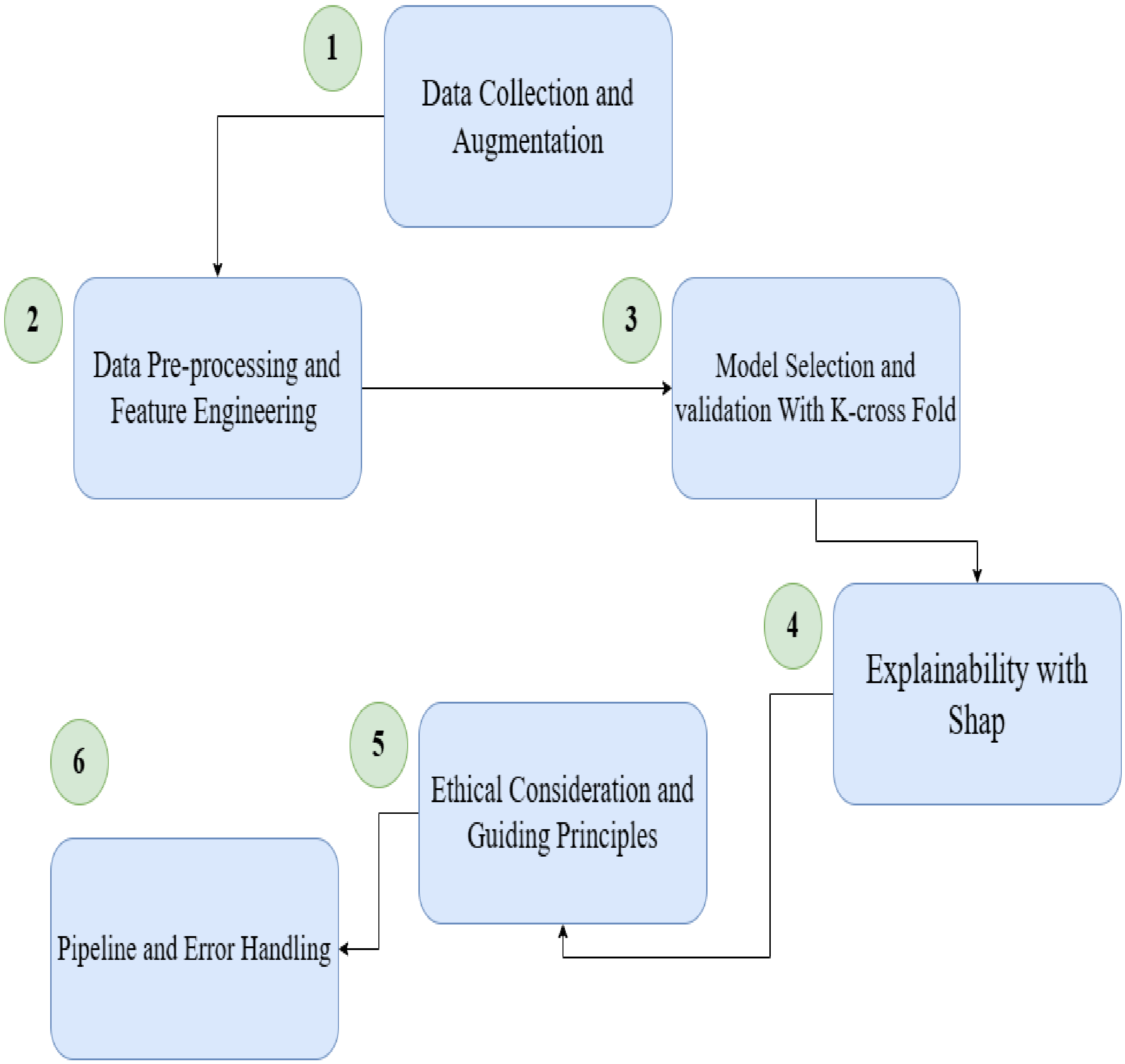

Fig. 3. Methodology Workflow diagram

4.6 Pipeline and Error Handling

The methodological pipeline follows a modular architecture: data ingestion → preprocessing → feature engineering → model training → validation → explanation → export. Each block incorporates exception handling, such as input validation for null text fields and classifier fallback strategies in case of convergence failures. There are no dead-ends; failed predictions trigger an evaluation log.

5 Findings and Results

This section details the findings from the processed dataset application and the performance of machine learning methods utilized in the resume screening exercise. The goal had been to predict the aptness of a candidate's resume for further review, based on extracted features such as education, skills, experience, or some relevant metadata.

5.1 Model Evaluation and Performance

The whole cleaned and preprocessed dataset had been used for training and evaluation of some machine learning classifiers like Logistic regression, SVM, Random Forest, and Gradient Boosting. The data were split and an 80/20 stratified split was used. Models were then evaluated on Accuracy, Precision, Recall, and F1-Score.

Table 1 summarizes the performance metrics of each classifier, revealing that XGBoost yielded the highest F1-score, followed by Random Forest and Extra Trees, with SVM showing competitive results in macro-average precision.

Table 1. Performance Comparison of Classification Models

Model	Accuracy (%)	Precision (macro avg)	Recall (macro avg)	F1-score (macro avg)
SVM (Linear Kernel)	65.36	65.44	65.36	64.57
Random Forest	68.75	66.98	68.75	66.30
Extra Trees	64.04	63.38	64.04	61.40
XGBoost	**72.69**	**71.80**	**72.69**	**71.76**

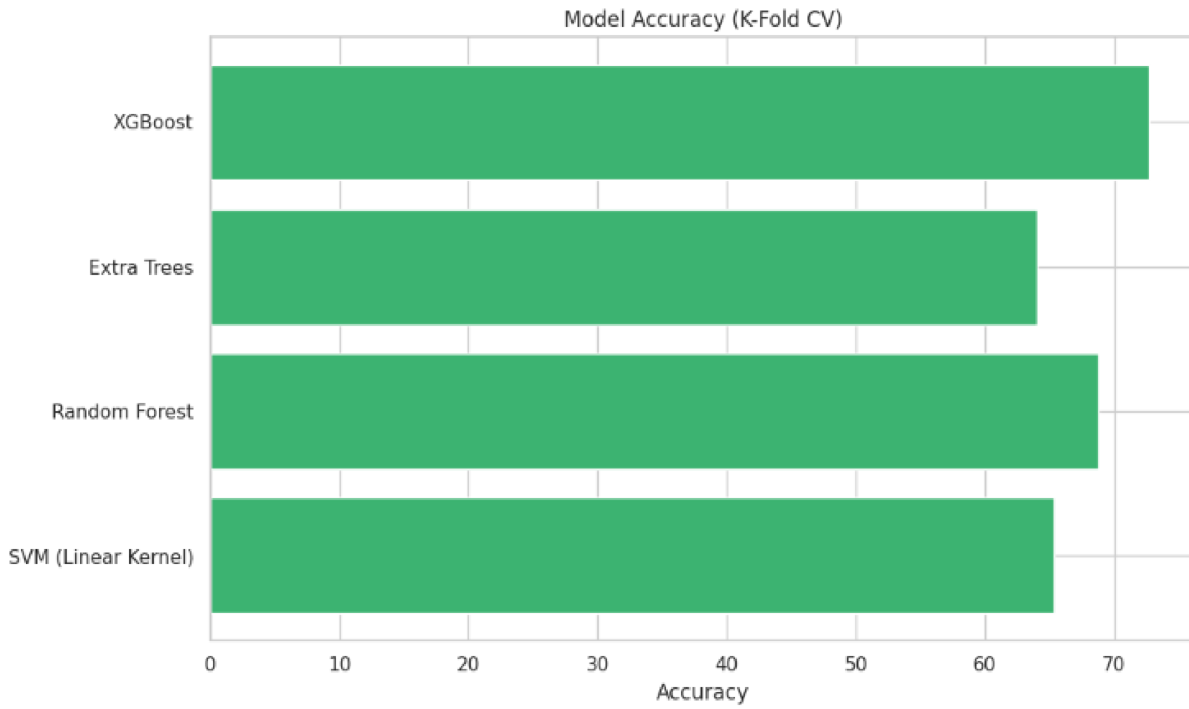

Fig. 4. Comparsion Between different models using K-fold

Considering these four models, **XGBoost is the most efficient and reliable algorithm for resume screening in the engineering education setup.**The best results in terms of accuracy, precision, recall, and F1 measure indicate that it can discriminate between qualified candidates and terminate both false positives and false negatives to a great extent. It can be concluded from this that ensemble methods, particularly gradient boosting techniques such as XGBoost, have their advantages in predictive tasks with complex, structured data such as resumes. In line with this viewpoint, XGBoost comes second to none in the proposed automated resume screening systems in a similar educational or recruiting scenario.

5.2 Confusion Matrix Analysis

A confusion matrix for the best-performing model (XGboost) is shown below.

From the matrix, we observe:

- **True Positives (TP)**: Correctly predicted shortlisted resumes.
- **True Negatives (TN)**: Correctly predicted non-shortlisted resumes.
- **False Positives (FP)**: Resumes incorrectly predicted as suitable.
- **False Negatives (FN)**: Qualified resumes missed by the model.

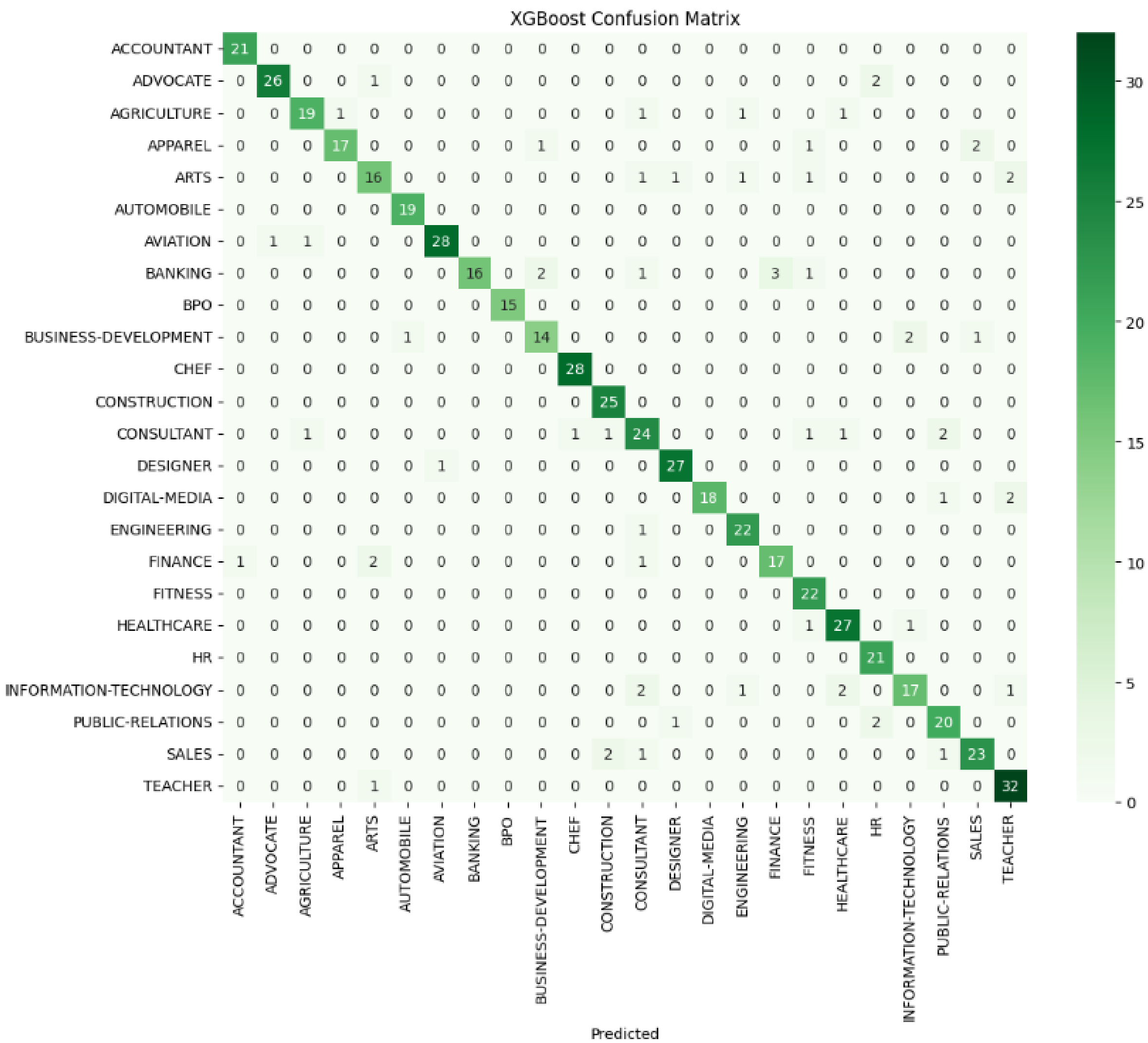

Fig. 5. Confusion matrix for the XGBoost model

These performance metrics are computed from the confusion matrix statistics:

- **Accuracy**:

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN}$$

- **Precision**:

$$\text{Precision} = \frac{TP}{TP + FP}$$

- **Recall (Sensitivity)**:

$$\text{Recall} = \frac{TP}{TP + FN}$$

- **F1-Score**:

$$\text{F1-Score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

6 Result Validation and Explainability

6.1 Result Validation with K-fold Cross Validation

Stratified K-Fold Cross Validation is implemented by us to ensure the robustness and generalizability of the classification models. With this method, the dataset is divided into k equally sized folds; each fold is used once as a validation set, while the other remaining $k - 1$ folds serve to train the algorithms. Further, this is repeated k times, and a final performance metric is computed from the average of all folds. This greatly reduces the variance attributed to a single split of training and testing and makes sure model performances remain consistent across randomly chosen subsets of the data [1]. Such a setup also provides a just ground for comparing several models, especially in the presence of class imbalance.

Consistently across folds, XGBoost achieved the highest macro-average F1 score, demonstrating both stability and better predictive ability.

6.2 Explainability With Shap

In high-stakes settings such as automated resume screening, model transparency becomes essential to instill trust and ensure the deployment of ethical AI. In pursuit of explainability, we integrated SHAP (SHapley Additive exPlanations), a cutting-edge XAI method grounded in cooperative game theory [12]. In SHAP, an importance value is allocated to each feature, which is the contribution that the feature makes to the actual prediction for a given instance.

- Allowing for global and local prediction interpretations, SHAP enables recruiters to learn what features matter in general and why a specific resume was deemed suitable or unsuitable.
- We then visualized the most contributing features (e.g., skills, years of experience, education level), which verified that the model was indeed learning human-intuitive patterns.
- Unlike LIME [15] or DeepLIFT [17], SHAP provides consistent explanations for varied types of models and adheres to local accuracy and missingness.

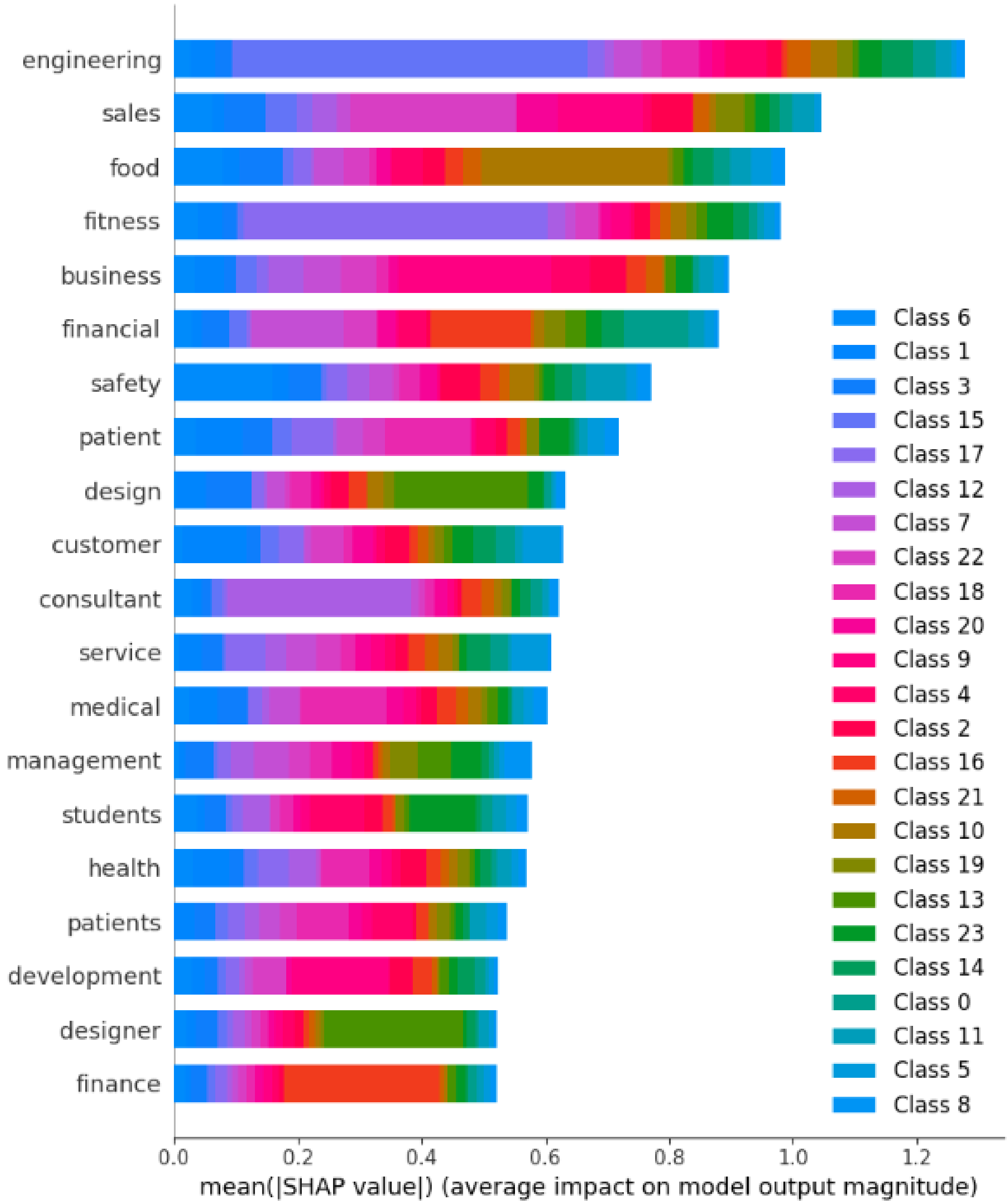

Fig. 6. Shap Summary plot showing every average feature importance per class.

7 Conclusion and Future Work

7.1 Conclusion

This research showed that resume screening for engineering education eligibility is feasible and effective when classical machine learning techniques were employed. By operating Natural Language Processing (NLP) methods, TF-IDF feature extraction in particular, and training several supervised learning models, the study predicted candidate suitability with high accuracy. Out of the models under comparison, XGBoost was ranked as the top-performing classifier with an average accuracy of 72.69%, precision of 71.80%, and recall of 72.69%, beating alternatives Random Forest, Extra Trees, and SVM [6].

For transparency in the model, the study integrated SHAP (SHapley Additive exPlanations) to derive both global and local feature importance [12]. This way, recruiters would be able to understand why a candidate was recommended or not, thus aligning the system to the latest advances in explainable AI (XAI) [2]. Explainability is of utmost importance in hiring scenarios that demand natural fairness and accountability [13].

Further, with the use of stratified k-fold cross-validation, the model was indeed made more robust and was protected from overfitting while ensuring generalizability over any diverse resume data [1]. This methodology was designed following FAIR (Findable, Accessible, Interoperable, Reusable) and CARE (Collective Benefit, Authority to Control, Responsibility, Ethics) data principles [4]. This ensures that ethical data governance is realized alongside reproachability for inclusivity concerning underrepresented populations in hiring.

In short, this work provides a technically usable but also interpretable and ethical framework for resume screening and candidate selection. It advocates for AI being accepted as a modern practice in HR and thus builds the foundation for scalable, fair, and trustworthy recruitment systems down the road.

7.2 Future Work

Future work may involve the following directions, to enchance the models Accuracy

- **Integrating Deep Learning:** The integration of transformer-based models including BERT and RoBERTa will allow deep learning systems to understand resume content within its contextual framework.
- **Bias Detection, Mitigation:** Regular audits for bias detection should combine fairness-aware algorithms with procedures to identify and eliminate demographic bias throughout all groups.
- **User Feedback Loop:** The system allows recruiters to provide feedback which automatically refines predictions through an ongoing learning loop.
- **Multi-level Classification:** The system processes multiple suitability factors by expanding its binary task to predict job fit together with cultural fit.
- **Deployment and Usability:** A user-friendly interface with an interactive dashboard and API platform allows recruiters to upload screening applications while presenting explainable outcome data to them.

8 Acknowledgments

We would like to thank Annajiat Alim Rasel, Md Sabbir Hossain and Labib Hasan Khan for their valuable feedback during the whole research including manuscript preparation.

References

[1] Sylvain Arlot and Alain Celisse. 2010. A survey of cross-validation procedures for model selection. *Statistics surveys* 4 (2010), 40–79.

[2] Alejandro Barredo Arrieta et al. 2020. Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information fusion* 58 (2020), 82–115. doi:10.1016/j.inffus.2019.12.012

[3] Leo Breiman. 2001. Random forests. *Machine learning* 45, 1 (2001), 5–32.

[4] Stephanie R Carroll et al. 2020. The CARE Principles for Indigenous Data Governance. *Data Science Journal* 19 (2020).

[5] Nitesh V Chawla, Kevin W Bowyer, Lawrence O Hall, and W Philip Kegelmeyer. 2002. SMOTE: synthetic minority over-sampling technique. *Journal of artificial intelligence research* 16 (2002), 321–357.

[6] Tianqi Chen and Carlos Guestrin. 2016. XGBoost: A scalable tree boosting system. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 785–794.

[7] Chengguang Gan, Qinghao Zhang, and Tatsunori Mori. 2024. Application of LLM Agents in Recruitment: A Novel Framework for Resume Screening. *arXiv preprint arXiv:2401.08315* (2024). Available at https://arxiv.org/abs/2401.08315.

[8] Pierre Geurts, Damien Ernst, and Louis Wehenkel. 2006. Extremely randomized trees. *Machine learning* 63, 1 (2006), 3–42.

[9] Trevor Hastie, Robert Tibshirani, and Jerome Friedman. 2009. *The Elements of Statistical Learning: Data Mining, Inference, and Prediction.* Springer Science & Business Media.

[10] Thorsten Joachims. 1998. Text categorization with Support Vector Machines: Learning with many relevant features. In *European conference on machine learning*. Springer, 137–142.

[11] Sihang Li, Kuangzheng Li, and Haibing Lu. 2023. National Origin Discrimination in Deep-learning-powered Automated Resume Screening. *arXiv preprint arXiv:2307.08624* (2023). Available at https://arxiv.org/abs/2307.08624.

[12] Scott M Lundberg and Su-In Lee. 2017. A unified approach to interpreting model predictions. *Advances in neural information processing systems* 30 (2017), 4765–4774.

[13] Muhammad Malik, Muhammad Awan, et al. 2023. A Systematic Literature Review on Artificial Intelligence in Recruiting and Selection: A Matter of Ethics. *Personnel Review* (2023). doi:10.1108/PR-03-2023-0257

[14] Prabakaran Paranthaman, B R Celia, V Vijayalakshmi, and V Senthil Kumaran. 2024. Resume Screening Automation: Enhancing Recruitment Efficiency with Machine Learning Algorithms. *Frontiers in Health Informatics* 13, 4 (2024), 1–10.

[15] Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin. 2016. "Why should I trust you?" Explaining the predictions of any classifier. In *Proceedings of the 22nd ACM SIGKDD international conference on knowledge discovery and data mining*. 1135–1144.

[16] Ahmed Salih, Zahra Raisi-Estabragh, Ilaria Boscolo Galazzo, Petia Radeva, Steffen E Petersen, Gloria Menegaz, and Karim Lekadir. 2023. A Perspective on Explainable Artificial Intelligence Methods: SHAP and LIME. *arXiv preprint arXiv:2305.02012* (2023). Available at https://arxiv.org/abs/2305.02012.

[17] Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. 2017. Learning important features through propagating activation differences. *arXiv preprint arXiv:1704.02685* (2017).

[18] Rajat Singh, Ankit Mishra, and Pawan Chauhan. 2023. Automated Resume Screening Using Machine Learning and NLP: A Survey. *International Journal of Advanced Computer Science and Applications* 14, 1 (2023), 163–170.

[19] Yuan Sun, Ee-Peng Lim, and Wee-Keong Ng. 2016. Sentence-level sentiment analysis of movie reviews using SentiWordNet and machine learning techniques. *Knowledge-Based Systems* 112 (2016), 58–69.

[20] Jason Wei and Kai Zou. 2019. EDA: Easy Data Augmentation Techniques for Boosting Performance on Text Classification Tasks. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing*. 6383–6389.

[21] Mark D Wilkinson, Michel Dumontier, IJsbrand Jan Aalbersberg, et al. 2016. The FAIR guiding principles for scientific data management and stewardship. *Scientific data* 3, 1 (2016), 1–9.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009