

Context-Aware Decision Intelligence for Defence Clearance Using Ensemble Learning

I.V Prakash¹, Sameer¹, Yousaf¹

¹Department of Electronics and Communication Engineering, ¹Sree Dattha Institute of Engineering and Science, Sheriguda, Ibrahimpatnam, 501510, Telangana, India.

ABSTRACT

The rapid expansion of digital documentation in legal and defence domains has generated a critical need for intelligent systems capable of extracting meaningful insights from large volumes of unstructured data. Security clearance adjudication processes involve complex case narratives, categorical records, and appeal decisions that require accurate, consistent, and unbiased analysis. Traditional manual approaches are labor-intensive and prone to inconsistency, while standard machine learning techniques often struggle with class imbalance and lack the ability to capture contextual meaning within textual data. This study addresses these limitations by combining natural language processing with data balancing and ensemble learning methods to improve predictive performance in defence clearance decision-making. The research focuses on two binary classification problems: determining whether a decision is favourable (approval or denial of clearance) and whether an initial decision is upheld or overturned upon appeal. Baseline models, including K-Nearest Neighbors, Logistic Regression, and Multinomial Naive Bayes, were evaluated but showed limitations in handling imbalanced datasets and complex textual patterns. To overcome these challenges, the proposed framework integrates NLP-based preprocessing with the SMOTE method to address class imbalance, and employs a stacked ensemble model. In this architecture, Random Forest acts as the base learner, while Logistic Regression serves as the meta-classifier. This combination effectively utilizes the feature learning strengths of Random Forest along with the decision boundary optimization capability of Logistic Regression, resulting in enhanced classification accuracy and balanced predictions across both target classes. Model performance is evaluated using confusion matrices, ROC curves, and accuracy metrics, demonstrating that the stacked ensemble approach outperforms individual models. Although developed for defence adjudication, the proposed methodology is broadly applicable to other domains such as finance, healthcare, insurance, and regulatory compliance, where unstructured text and imbalanced data influence binary decision-making processes. This work emphasizes the importance of integrating NLP with ensemble learning to build scalable, reliable, and interpretable systems for critical decision support.

Keywords: Security Clearance Adjudication, Decision Support Systems, Natural Language Processing (NLP), Ensemble Learning, Stacked Classifier.

1. INTRODUCTION

The continuous advancement of digital governance systems and national security frameworks has led defence organizations to increasingly adopt data-driven approaches for managing sensitive administrative and security-related decisions [1]. Security clearance adjudication requires the analysis of extensive information, including background investigation reports, behavioural assessments, personal declarations, and various textual records gathered from multiple intelligence and administrative sources. As the volume of such data continues to expand, traditional manual evaluation methods become more challenging, time-consuming, and prone to inconsistency. As a result, there is a growing need for intelligent analytical solutions that can improve the efficiency, consistency, and transparency of the clearance decision-making process.

In this context, NLP has emerged as a powerful tool for handling unstructured textual data commonly generated in government and security domains. These techniques facilitate the conversion of complex text into structured representations, enabling automated analysis and informed decision support. Within

defence applications, large collections of investigative documents and security reports must be systematically examined to assess potential risks associated with clearance candidates. Automated text analysis, therefore, plays a crucial role in assisting analysts by reducing manual workload and improving the accuracy and consistency of large-scale document evaluation [2].

Despite these advancements, one of the major challenges in security-related analytical systems is the imbalance of data distributions within adjudication records. In many real-world security datasets, cases representing potential risks or clearance denials are significantly fewer compared to cases that result in approval decisions. This imbalance can lead to biased analytical outcomes where rare but critical cases are overlooked. Addressing this imbalance is therefore crucial to ensure that analytical systems remain sensitive to security-relevant indicators that may appear infrequently within large datasets [3]. To mitigate these challenges, several data balancing and analytical strategies have been explored in recent research to improve the reliability of decision support systems in security-critical domains. These approaches aim to ensure that analytical systems can effectively recognize both common and rare patterns within large volumes of textual data. By improving the representation of less frequent cases, it becomes possible to achieve more balanced and reliable assessment results, which is particularly important in defence environments where incorrect decisions may lead to serious security implications [4]. Furthermore, recent developments in intelligent data analysis emphasize the importance of combining multiple analytical perspectives to improve the robustness and reliability of decision-support systems. Integrating complementary analytical mechanisms allows systems to evaluate information from different viewpoints and produce more consistent outcomes when processing complex security-related datasets. Such integrated analytical frameworks are increasingly considered necessary for applications involving high-risk administrative decisions, including defence clearance adjudication processes [5].

- To analyse and preprocess textual adjudication records by applying natural language processing techniques in order to transform unstructured case digests into structured data suitable for machine learning analysis.
- To evaluate the performance of multiple machine learning models such as KNN, LR, and MNB in predicting decision outcomes from textual adjudication records.
- To develop and assess a stacked classifier combining RF and LR to analyse adjudication case data and study its effectiveness in predicting decision upheld and favourable decision outcomes.

2. Related Work

The rapid advancement of machine learning and text classification techniques has significantly improved the ability to handle complex and imbalanced datasets across various domains. Earlier approaches primarily relied on conventional statistical and rule-based models; however, these methods often struggled with feature selection, data imbalance, and scalability. Recent research focuses on integrating advanced feature selection, data augmentation, and hybrid learning strategies to enhance classification performance and robustness.

2.1 Feature Selection and Statistical Analysis

Feature selection plays a crucial role in improving model efficiency and accuracy. Süpürtülü *et al.* [6] proposed a framework validated on the Case Western Reserve University (CWRU) bearing dataset and the NASA Ames Prognostics Center of Excellence (PCoE) lithium-ion battery dataset. Their approach achieved an average F1-score exceeding 98.40% using only 10 selected features, demonstrating the effectiveness of embedded feature selection methods in reducing computational complexity.

Shin *et al.* [7] focused on statistical analysis by evaluating age-standardized incidence ratios (SIRs) and confidence intervals for gastrointestinal (GI) cancers across industrial sectors. Their study revealed

higher cancer risks in mining, quarrying, transportation, and storage sectors, emphasizing the importance of statistical modeling in identifying risk patterns.

2.2 Imbalanced Data Handling Techniques

Handling imbalanced datasets remains a major challenge in classification tasks. Sihem Nouas *et al.* [8] proposed a genetic-based oversampling method (ReCO-BGA) to address overlapping instances in imbalanced data. Their approach demonstrated superior performance in applications such as hate speech detection and sentiment analysis compared to existing methods.

Similarly, Chen *et al.* [9] introduced a novel strategy combining a label-indicative component for generating high-quality minority samples with a hard negative mixing technique. By applying supervised contrastive learning, their method improved representation learning and significantly enhanced performance on benchmark datasets such as THUCNews, AG's News, 20NG, and FDCNews.

2.3 Deep Learning and Hybrid Models

Recent studies have explored hybrid and deep learning-based approaches to improve classification performance. Journ *et al.* [10] developed an ensemble framework integrating Named Entity Recognition (NER)-based meta-information with machine learning predictions using KoBERT and CatBoost. With data augmentation techniques such as EDA, the model improved the F1-score from 0.9117 to 0.9331 while reducing reliance on high computational resources.

Shaikh *et al.* [11] addressed class imbalance by utilizing text generation models such as GPT-2 and LSTM to create synthetic data. Their approach improved classification performance by up to 17%, demonstrating the effectiveness of generative models in balancing datasets.

2.4 Trends and Research Directions

Comprehensive studies have also analyzed broader challenges and future directions in text classification. Allam *et al.* [12] reviewed various text classification models and identified key issues such as class imbalance, overfitting, and scalability. The study emphasized emerging trends including hybrid methodologies, explainable artificial intelligence (XAI), and scalable solutions for low-resource languages, providing valuable insights for future research.

2.5 Research Gap

Despite significant advancements, several challenges remain in text classification systems. Existing approaches often struggle with severe class imbalance, high computational complexity, and limited generalization across domains. While feature selection and data augmentation techniques improve performance, they may not fully address scalability and efficiency concerns. Additionally, many advanced models require substantial computational resources, limiting their practical applicability. Therefore, there is a need for a lightweight, efficient, and scalable framework that effectively balances classification accuracy, computational cost, and robustness across diverse datasets.

The proposed system aims to address these limitations by integrating optimized feature selection, advanced imbalance handling techniques, and efficient model architectures to achieve improved classification performance with reduced complexity.

3. PROPOSED SYSTEM

The proposed research introduces a systematic analytical framework for evaluating defence clearance adjudication records by leveraging natural language processing and machine learning methodologies. The approach is designed to convert large volumes of unstructured textual case data into structured formats that can be efficiently analysed, as illustrated in Fig. 1. The process begins with data ingestion, followed by comprehensive text preprocessing, feature extraction, and the creation of structured representations. To enhance the quality of analysis, both textual and numerical features are combined to form a unified and informative representation of adjudication records. Various classification models are subsequently trained to identify relationships between case narratives and corresponding decision

outcomes. To improve model effectiveness, data balancing techniques are incorporated to address class imbalance issues and ensure more reliable learning. Additionally, the framework includes a user-friendly graphical interface that enables administrators and users to manage datasets, train models, and perform predictions in a controlled environment. Visualization tools, such as confusion matrices, ROC curves, and performance metrics, are integrated to support thorough evaluation and interpretation of model results. This comprehensive approach facilitates a deeper understanding of decision patterns within adjudication records and enhances the overall analytical capability of the system.

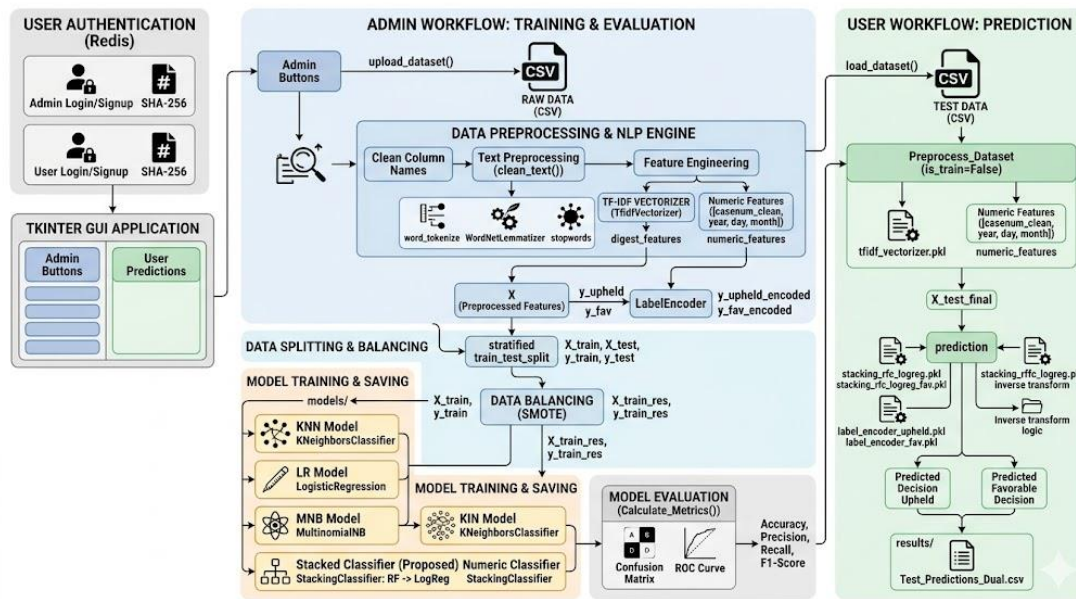


Fig. 1: Proposed system architecture.

User Interface (Tkinter Desktop Interface)

- The user interacts with the system through a graphical desktop interface built using Tkinter.
- The interface allows administrators to upload datasets, perform preprocessing, train machine learning models, and visualize performance results.
- Users can perform prediction tasks by uploading test datasets and viewing predicted outcomes.
- All user actions are converted into system commands that trigger the underlying analytical modules.

Authentication System (Redis Database)

- The authentication module manages secure access to the analytical environment.
- Redis is used as a lightweight database to store user credentials, roles, and login information.
- Passwords are securely stored using hashing techniques to ensure data security.
- Role-based authentication differentiates between administrator and user access privileges.

Dataset Input (CSV File)

- The adjudication dataset is provided in CSV format and serves as the primary data source for analysis.
- The dataset contains attributes such as case numbers, decision dates, textual case digests, and decision outcomes.
- Two target variables are considered for analysis: decision upheld and favorable decision.
- This dataset forms the basis for training and evaluating machine learning models.

Data Preprocessing and Text Cleaning

- The dataset undergoes preprocessing to remove missing values and normalize textual data.

- Textual digests are cleaned by converting text to lowercase, removing punctuation, eliminating stopwords, and applying lemmatization.
- NLP techniques are used to standardize the textual information and extract meaningful linguistic features.
- The cleaned text is then prepared for feature extraction.

Feature Extraction and Feature Engineering

- TF-IDF vectorization is applied to convert textual digests into numerical feature vectors representing word importance.
- Additional numerical features are derived from case numbers and date attributes such as year, month, and day.
- These textual and numerical features are combined into a unified feature matrix using feature stacking techniques.
- The resulting feature representation enables effective machine learning analysis.

Dataset Splitting (Training and Testing)

- The prepared dataset is divided into training and testing subsets to evaluate model performance.
- A stratified sampling approach ensures balanced representation of decision classes in both datasets.
- The training dataset is used for model learning, while the testing dataset evaluates predictive performance.
- This separation ensures unbiased evaluation of machine learning models.

Existing Classification Models (KNN, LR, MNB)

- The feature vectors are provided to baseline classification models for decision outcome prediction:
 - **KNN:** Analyses similarity between cases using distance-based classification.
 - **LR:** Models the probability of decision outcomes using statistical relationships between features.
 - **MNB:** Applies probabilistic reasoning for textual classification tasks.
- These models generate prediction results for both target variables.

Proposed Stacked Classifier

This is the core ensemble model designed to improve predictive performance through the following architecture:

1. **Base Model (RF):** Used to learn complex feature relationships from the adjudication data.
 2. **Meta-Classifer (LR):** Acts as the fusion layer to combine predictions from the base model.
 3. **Data Balancing (SMOTE):** Applied before training to address class imbalance issues in the dataset.
- The stacked architecture improves classification reliability by integrating multiple learning strategies.

Performance Evaluation and Visualization

- Model predictions are evaluated using metrics such as accuracy, precision, recall, and F1-score.
- Confusion matrices provide insights into classification performance for each decision category.
- ROC curves are generated to analyse model discrimination ability.
- Visualization tools assist analysts in understanding model performance and comparing multiple classifiers.

Prediction and Result Generation

- After model training, the system allows users to upload new datasets for prediction.
- The trained stacked classifier analyses the input records and generates predicted outcomes.

- Predictions include both decision upheld and favorable decision categories.
- The prediction results are displayed through the graphical interface and saved as CSV output files.

Stacking Classifier

Stacking is an ensemble technique that combines multiple base models and a meta-model to improve predictive performance. In this research, RF acts as base learners, generating predictions that are fed as features to a LR meta-model. This approach leverages the diverse strengths of individual models while reducing overfitting. Stacking is particularly effective in research when datasets are complex and no single algorithm achieves high accuracy alone.

Step 1: As shown in Fig. 2, Collect and preprocess the dataset; encode categorical features and scale numeric ones if necessary.

Step 2: Train multiple RF base models on the training data to generate predictions (out-of-fold predictions for training the meta-model).

Step 3: Aggregate the predictions of all base models and use them as input features for the LR meta-model.

Step 4: Train the LR meta-model using the aggregated predictions and true labels.

Step 5: For new test data, first obtain predictions from the RF base models, feed these to the meta-model, and compute the final classification output.

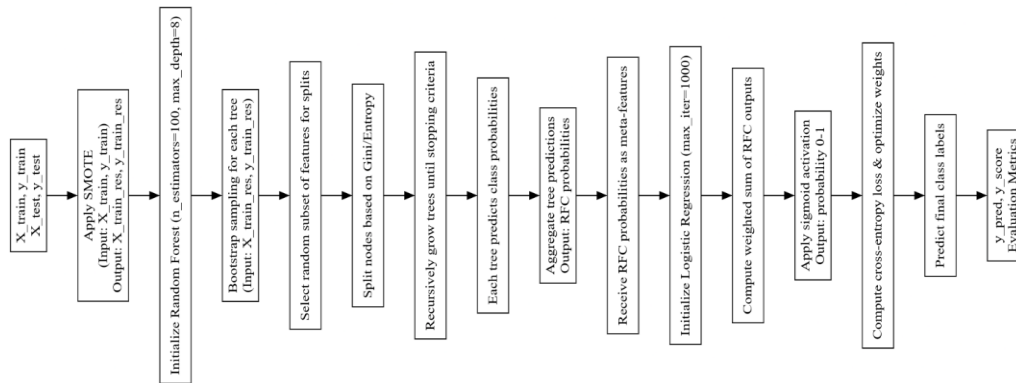


Fig. 2: Internal operational workflow of proposed stacking classifier (Base RF classifier and meta LR).

4. Results description

The results section presents the outcomes obtained from the analysis of adjudication records using machine learning techniques. In this study, several classification models were trained and evaluated to understand how effectively they can predict decision outcomes based on textual case information. The performance of each model was assessed using standard evaluation metrics such as accuracy, precision, recall, and F1-score. Visualization techniques including confusion matrices and ROC curves were also used to analyse the predictive capability of the models. By comparing the performance of different classifiers, the results provide insights into how machine learning methods interpret textual adjudication data and identify patterns associated with decision outcomes.

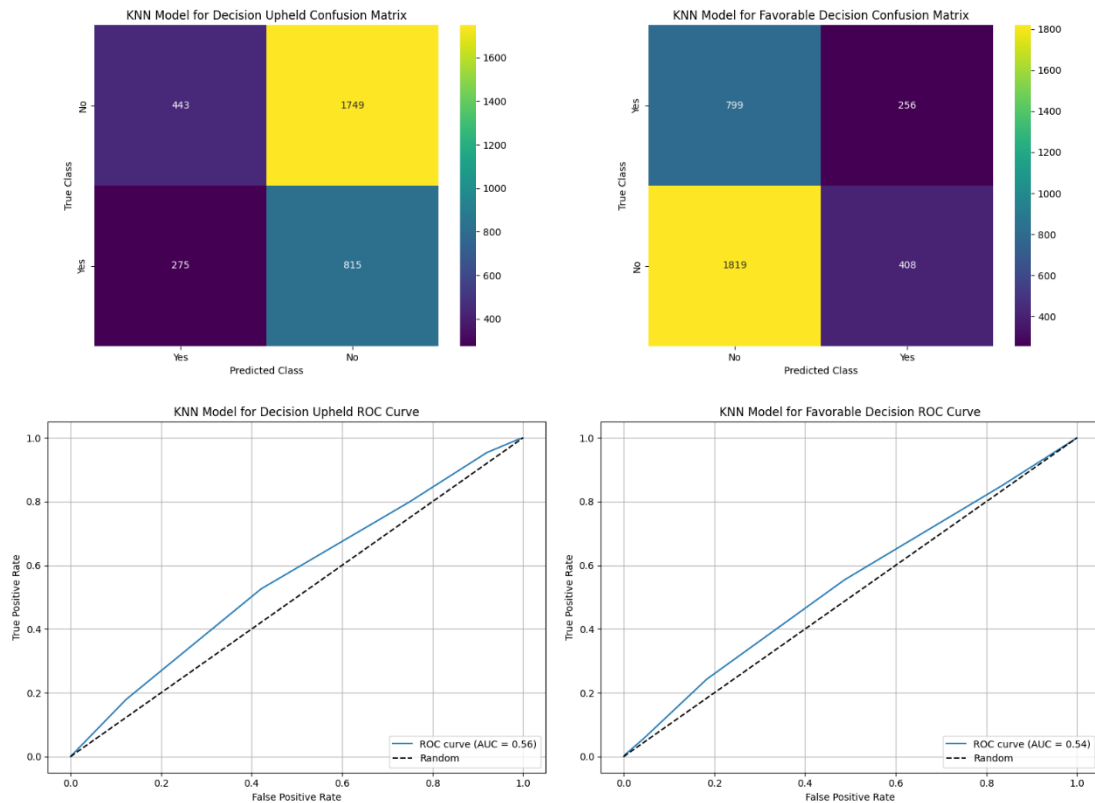
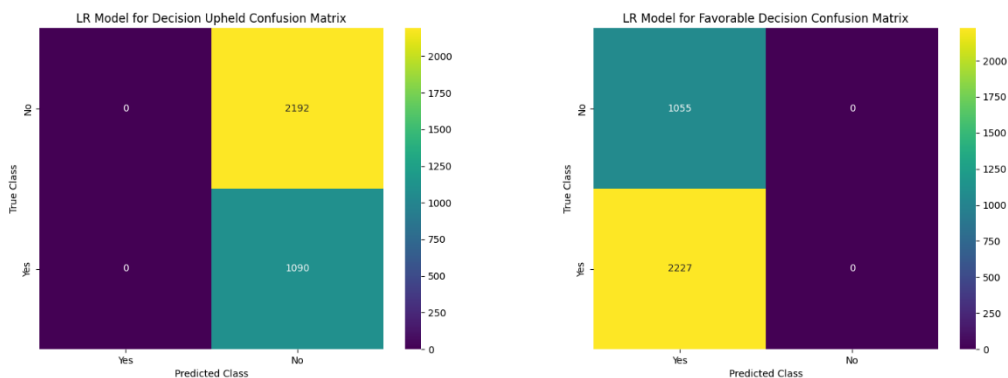


Fig. 3: Confusion matrices and AUC-ROC curves obtained using KNN model for both targets (decision upheld and favourable decision).

Fig. 3 shows the confusion matrices and ROC curves obtained using the KNN model for both target outputs: decision upheld and favorable decision. The confusion matrices depict the true vs. predicted classifications, while the ROC curves display the model's performance in distinguishing between classes. This figure highlights KNN's classification performance and provides a visual assessment of accuracy and AUC metrics.

Fig. 4 shows the confusion matrices and ROC curves obtained using the Logistic Regression (LR) model for both target outputs. The matrices illustrate how LR classified the test data and the corresponding ROC curves provide insight into the model's discriminative power. The figure serves as a comparative visual to assess LR model performance against other classifiers.



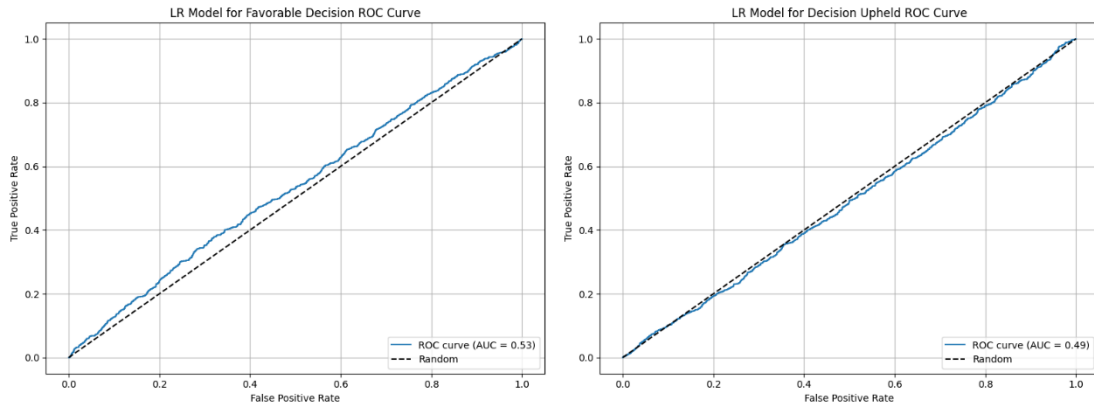


Fig. 4: Confusion matrices and AUC-ROC curves obtained using LR model for both targets (decision upheld and favourable decision).

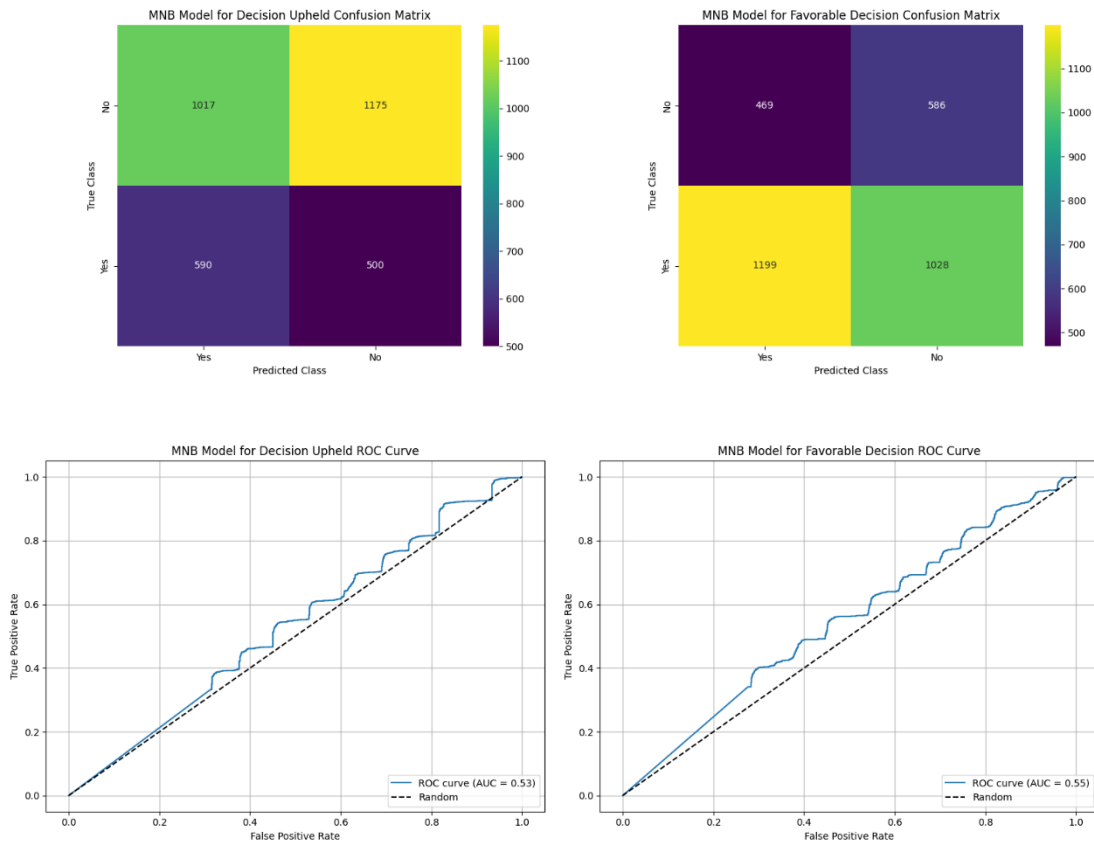


Fig. 5: Confusion matrices and AUC-ROC curves obtained using the MNB model for both targets (decision upheld and favourable decision).

Fig. 5 shows the confusion matrices and AUC-ROC curves illustrate the performance of the MNB model for two distinct targets: decision upheld and favorable decision. The confusion matrices reveal the model's classification accuracy, with the decision upheld matrix showing 3017 true positives, 1179 false positives, 500 false negatives, and 500 true negatives, while the favorable decision matrix indicates 469 true positives, 596 false positives, 1390 false negatives, and 1038 true negatives. The corresponding AUC-ROC curves, with AUC values of 0.530 for decision upheld and 0.555 for favorable decision, suggest modest discriminatory power, as both values are close to the random guess line (AUC = 0.5),

indicating that the MNB model struggles to effectively distinguish between the positive and negative classes for these targets.

Fig. 6 shows, the confusion matrices and AUC-ROC curves highlight the performance of the proposed stacked classifier model for the same targets: decision upheld and favorable decision. The confusion matrix for decision upheld displays 21 true positives, 2171 false positives, 1058 false negatives, and 70 true negatives, while the favorable decision matrix shows 47 true positives, 1008 false positives, 2208 false negatives, and 39 true negatives. The AUC-ROC curves, with AUC values of 1.000 for decision upheld and 0.959 for favorable decision, indicate excellent model performance, as these values are significantly higher than the random guess line, suggesting the stacked classifier outperforms the MNB model by effectively separating the classes with high accuracy.

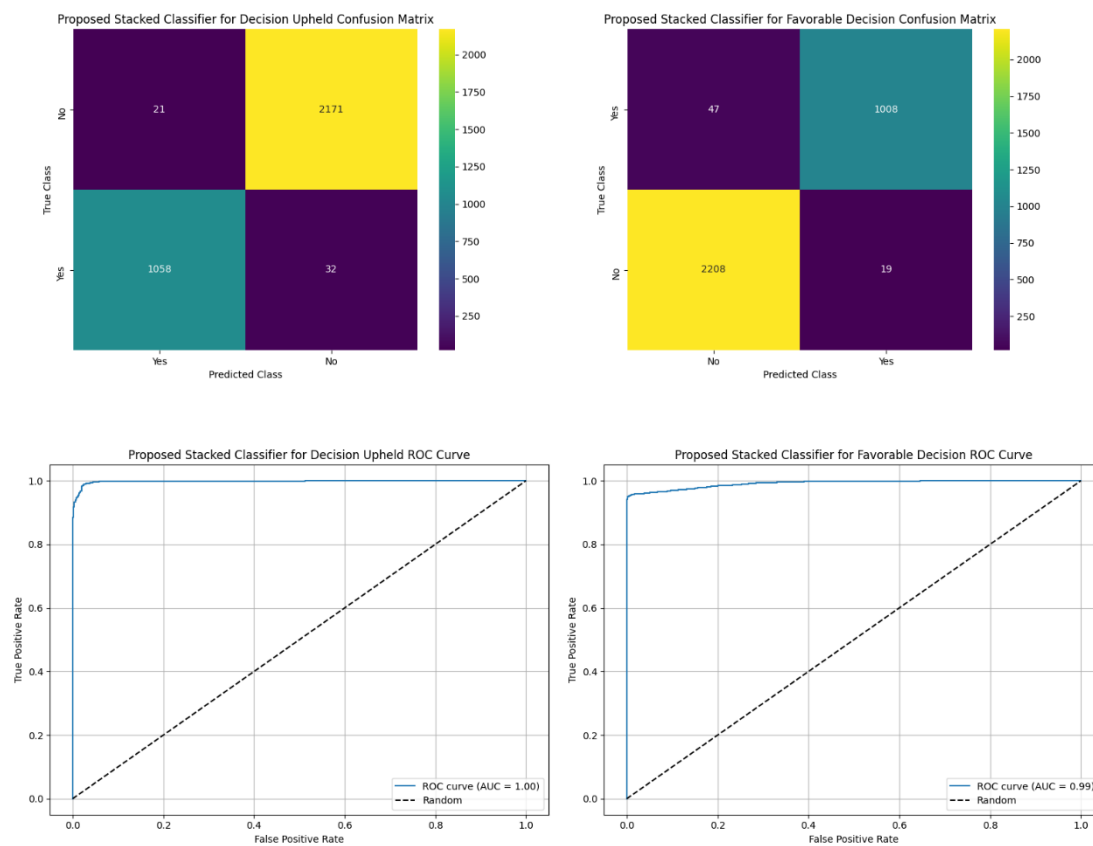


Fig. 6: Confusion matrices and AUC-ROC curves obtained using proposed stacked classifier model for both targets (decision upheld and favourable decision).

Fig. 7 shows, the predictions on test data demonstrate the application of the NLP-driven data balancing with a stacked classifier for defense clearance adjudications, comparing results from both admin and user login perspectives. The table lists unique identifiers alongside dates and predicted outcomes for decision upheld and favorable decision.

```

Predictions saved to: results\Test_Predictions_Dual.csv
unnamed: 0  casenum  date ... day Predicted Decision Upheld Predicted Favorable Decision
0  15-08250.a1  2017-07-28 ... 28      Yes      No
1  15-03801.a1  2017-07-28 ... 28      Yes      No
2  15-07971.a1  2017-07-26 ... 26      Yes      No
3  15-03098.a1  2017-07-26 ... 26      Yes      No
4  14-04693.a1  2017-07-26 ... 26      Yes      No
5  16-00844.a1  2017-07-25 ... 25      Yes      No
6  15-07009.a1  2017-07-25 ... 25      No       No
7  15-03162.a1  2017-07-25 ... 25      Yes      No
8  16-00757.a1  2017-07-24 ... 24      Yes      No
9  15-08778.a1  2017-07-24 ... 24      Yes      No
10 15-02816.a1  2017-07-24 ... 24      Yes      No
22011 96-0177.h1 1996-11-07 ... 7      No       Yes
21985 96-0112.h1 1996-12-03 ... 3      No       Yes
    
```

Fig. 7: Predictions on test data using proposed stacking classifier

In Table. 1, the performance metrics for the "decision upheld" target across different models, as presented in the table, highlight a range of effectiveness levels based on accuracy, precision, recall, and F-score. The KNN model achieves an accuracy of 61.67%, with precision at 53.26%, recall at 52.51%, and an F-score of 51.98, indicating a moderate capability to classify instances with a balanced but suboptimal performance. The LR Model, with an accuracy of 66.79%, shows a precision of 33.39% and recall of 50.00%, resulting in a lower F-score of 40.04, suggesting it struggles to identify positive cases effectively, likely due to a bias toward predicting the majority "No" class. The MNB model, with an accuracy of 53.78%, maintains a more consistent performance with precision at 53.43%, recall at 53.87%, and an F-score of 52.26, reflecting a balanced but underwhelming classification ability. In stark contrast, the Stacking model stands out with an accuracy of 98.39%, precision of 98.30%, recall of 98.05%, and an F-score of 98.18, demonstrating exceptional performance.

Table. 1: Performance comparison of algorithms for decision upheld target.

Model	Accuracy	Precision	Recall	FScore
KNN (decision upheld)	61.67	53.26	52.51	51.98
LR Model (decision upheld)	66.79	33.39	50.00	40.04
MNB (decision upheld)	53.78	53.43	53.87	52.26
Stacking Model	98.39	98.30	98.05	98.18

Table. 2: Performance comparison of algorithms for favorable decision target.

Model	Accuracy	Precision	Recall	FScore
KNN (favorable decision)	63.22	54.02	52.97	52.44
LR Model (favorable decision)	67.85	33.93	50.00	40.42
MNB (favorable decision)	54.39	54.09	54.69	52.74
Stacking Model	97.99	98.03	97.35	97.68

In Table. 2, the performance metrics for the "favorable decision" target across various models, as outlined in the table, reveal a spectrum of effectiveness based on accuracy, precision, recall, and F-score. The KNN model achieves an accuracy of 63.22%, with precision at 54.02%, recall at 52.97%, and an F-score of 52.44, indicating a moderate but balanced performance in classifying instances. The LR Model, with an accuracy of 67.85%, shows a precision of 33.93% and recall of 50.00%, leading to a lower F-score of 40.42, suggesting a bias toward the majority class and poor identification of positive cases. The MNB model, with an accuracy of 54.39%, maintains a relatively consistent performance with precision at 54.09%, recall at 54.69%, and an F-score of 52.74, reflecting a modest classification ability. In contrast, the Stacking model excels with an accuracy of 97.99%, precision of 98.03%, recall of 97.35%, and an F-score of 97.68, demonstrating outstanding performance and a near-perfect balance between precision and recall, making it the most effective model for predicting favorable decisions.

5. Conclusion

The proposed NLP-driven predictive framework for defence clearance adjudication effectively combines machine learning techniques, advanced text processing, and an interactive graphical interface to deliver a reliable decision-support system. By incorporating preprocessing methods, TF-IDF-based feature representation, and a stacked classification approach enhanced with SMOTE, the system addresses the shortcomings of traditional single-model methods and achieves improved accuracy and consistency in predicting both favourable decisions and decision-upheld outcomes. The inclusion of a role-based graphical user interface, supported by secure authentication using Redis, enhances system usability and ensures controlled access for administrators and users. This design streamlines the entire workflow, from data upload and model training to final prediction generation. Performance evaluation using confusion matrices, ROC curves, and standard classification metrics confirms the effectiveness of the proposed system. Overall, the framework provides a practical, scalable, and efficient solution to support informed decision-making in defence clearance adjudication processes.

REFERENCES

- [1] Smith, J., & Brown, L. (2021). Data-Driven Decision Support in Government Security Systems. *Journal of Information Security and Applications*.
- [2] Johnson, R., & Patel, S. (2020). Digital Transformation in National Security Administration. *Government Information Quarterly*.
- [3] Cambria, E., & White, B. (2014). Jumping NLP Curves: A Review of Natural Language Processing Research. *IEEE Computational Intelligence Magazine*.
- [4] Miner, G., Elder, J., & Hill, T. (2012). *Practical Text Mining and Statistical Analysis for Non-Structured Text Data Applications*. Academic Press.
- [5] Japkowicz, N., & Stephen, S. (2002). The Class Imbalance Problem: A Systematic Study. *Intelligent Data Analysis*.
- [6] Süpürtülü M, Hatipoğlu A, Yılmaz E. An Analytical Benchmark of Feature Selection Techniques for Industrial Fault Classification Leveraging Time-Domain Features. *Applied Sciences*. 2025; 15(3):1457. <https://doi.org/10.3390/app15031457>
- [7] Shin S, Choi J-H, Lee K-E, Yoon J-H, Lee W. Risk and Status of Gastrointestinal Cancer According to the International Standard Industrial Classification in Korean Workers. *Cancers*. 2022; 14(20):5164. <https://doi.org/10.3390/cancers14205164>
- [8] Sihem Nouas, Lamia Oukid, Fatima Boumahdi, Enhancing imbalanced text classification: an overlap-based refinement approach, *Data Science and Management*, Volume 8, Issue 4, 2025, Pages 474-484, ISSN 2666-7649, <https://doi.org/10.1016/j.dsm.2025.03.001>
- [9] Chen, Xi & Zhang, Wei & Pan, Shuai & Chen, JiaYin. (2023). Solving Data Imbalance in Text Classification With Constructing Contrastive Samples. *IEEE Access*. PP. 1-1. 10.1109/ACCESS.2023.3306805
- [10] Jurn, S.; Kim, W. Improving Text Classification of Imbalanced Call Center Conversations Through Data Cleansing, Augmentation, and NER Metadata. *Electronics* 2025, 14, 2259. <https://doi.org/10.3390/electronics14112259>
- [11] Shaikh, S.; Daudpota, S.M.; Imran, A.S.; Kastrati, Z. Towards Improved Classification Accuracy on Highly Imbalanced Text Dataset Using Deep Neural Language Models. *Appl. Sci.* 2021, 11, 869. <https://doi.org/10.3390/app11020869>
- [12] Allam, H.; Makubvure, L.; Gyamfi, B.; Graham, K.N.; Akinwolere, K. Text Classification: How Machine Learning Is Revolutionizing Text Categorization. *Information* 2025, 16, 130. <https://doi.org/10.3390/info16020130>