



UGANDA CHRISTIAN
UNIVERSITY

A Centre of Excellence in the Heart of Africa

Department of Computing & Technology

*A PREDICTIVE ANALYTICS FRAMEWORK FOR CASE
BACKLOG MANAGEMENT IN UGANDA'S JUDICIARY: AN
EXPLAINABLE MACHINE LEARNING APPROACH*

By

BBOSSA ISAAC SSERUNKUMA

Senior Statistician

Reg No: *M24M19/024*, Student No: *B28761*

E-mail: *bbossaisaac87@gmail.com*, Phone No: *+256 754 753 579*

A dissertation submitted to the Faculty of Engineering Design & Technology for the study leading to the Degree in partial fulfillment of the requirements for the award of **Master of Science in Data Science and Analytics** of Uganda Christian University

June 2026

DECLARATION

I, *BBOSSA ISAAC SSERUNKUMA*, declare to the best of my knowledge that the work in this dissertation is original and has not been submitted before to any university/institution. It is the result of my effort.

A handwritten signature in blue ink, appearing to read 'BBOSSA ISAAC SSERUNKUMA', is written over a horizontal line.

Signature: _____

24th September 2026

Date: _____

ETHICAL GENERATIVE AI USAGE DECLARATION

Students may use generative artificial intelligence (AI) ethically as a “study buddy” or assistant rather than a replacement for critical thinking. This includes citing AI use and fact-checking all outputs to maintain academic integrity.

For purposes of this assignment, the following guidelines shall apply

- i. Generative AI may be used for brainstorming, and generating ideas to improve your work. You may also use AI for literature search, and summarizing journal papers for quick analysis and assessing relevance of publications. However, once the relevant publications have been identified, a student is expected to fully review the selected publications, identifying problems addressed, methods used, gaps identified and recommendations made.
- ii. Where AI has been used as guided in (i) above, full disclosure of how it has been used should be made in the appendix C, including how AI was used, the AI tools (including model versions), and a compilation on major AI prompts used.
- iii. NO GENERATIVE AI CONTENT IS ALLOWED IN THE FINAL SUBMISSION. The final submission SHOULD be your original work.

A. Acknowledging the use of AI as a study buddy

Please refer to Appendix C for the complete declaration of AI usage for generating ideas, brainstorming, literature search, and summarizing only to improve your work.

APPROVAL

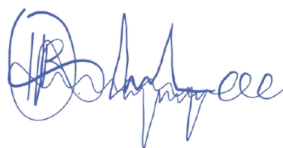
This dissertation has been submitted for examination with approval of the following advisors/supervisor.

Advisor: *Dr. Ian Raymond Osolo*

Department of Computing & Technology

Faculty of Engineering Design & Technology

Uganda Christian University



Signature: _____

Date: 10 / 09 / 2026

ABSTRACT

The Uganda judicial system is faced with a persistent and rising backlog of cases, whereby 26.32% of cases have remained pending for more than two years, negatively impacting access to justice and socio-economic development. Current systems such as the Electronic Court Case Management Information System (ECCMIS) can only make use of descriptive reporting techniques, thus implementing a reactive administrative paradigm. This research bridges the critical gap in the area of predictive capability through developing an explainable machine learning model to preemptively predict the backlog risk of cases. Adopting the Design Science Research method, the research adopted a quantitative experimental design. Making use of secondary data from the National Court Case Census (2025), the research included substantial data pre-processing, feature engineering, and comparing various machine learning classifiers including Logistic Regression, Random Forest, Gradient Boosting, and K-Nearest Neighbors to carry out binary classification task of backlog prediction. The best performing classifier was Random Forest with an ROC AUC of 0.885 and F1 Score of 0.698, significantly outperforming other classifiers. The explainable AI (XAI) analysis indicated that Claim/offence (case complexity), Court name (institutional capacity), and case status (procedural stage) were the most predictive features. This further confirms the hypothesis that backlog risk is essentially due to the legal and institutional dynamics rather than merely time dynamics. The analysis succeeded in developing an empirical predictive model that proved the potency of ensemble models such as Random Forest in capturing the complex non-linear dynamics of judicial backlog in Uganda. The results allow shifting from reactive to proactive case management strategy. This thesis ends up with some practical recommendations regarding the implementation of a predictive warning system, dynamic resource planning, and targeted procedural reform.

Keywords: Predictive Analytics, Case Backlog, Machine Learning, Explainable AI (XAI), Judiciary, Uganda, Random Forest, Judicial Administration, Design Science Research.

DEDICATION

“In the Name of Allah, the Most Gracious, the Most Merciful.”

This thesis is dedicated firstly to Allah (SWT), whose mercies, wisdom, and guidance have been with me throughout this academic journey. They have helped me in times of hardship and motivated me to persist in the pursuit of knowledge.

To my dear late mother, Margaret Nsubuga, may her soul rest in peace, I dedicate this research with gratitude and love. It is because of you that I value curiosity, discipline, and relentless pursuit of knowledge and wisdom. Even though you are not physically here anymore, your teachings, sacrifices, and trust in the power of education continue to be a beacon in my life.

To my number one fans, Mable and Glorious, thank you for your love, encouragement, and support. Thank you especially for embracing my passion, knowing the challenges associated with this venture, and always being by my side in times of victories and struggles. Your belief in me has been my source of strength and motivation.

My wonderful children Aayan, Anira, Aahil, and Anayah—thanks for the many moments of joy, wonder, and surprises that fill me with admiration every day. You are curious, imaginative, and energetic—you make me remember the value of learning, development, and the responsibility to leave a better world for future generations.

I thank my research supervisor, *Dr. Ian Raymond Osolo*, for his encouragement, guidance, and continuous help during this journey. His expertise, mentorship, and constructive critique greatly contributed to this research and helped it succeed.

Additionally, I would like to thank the Judiciary of Uganda for allowing me to use the National Court Case Census 2025 database, which was the basis for this study. My gratitude goes to the Statistics and Data Management Team of the Judiciary for encouraging me to develop my skills and excel at making decisions on the basis of data.

Last but not least, I offer this paper in tribute to all those people who recognize the immense impact that data, knowledge and lifelong learning can have. I hope that this accomplishment serves as proof of the power of faith and persistence, as well as the inspiration we receive from others.

CONTENTS

Declaration	i
Ethical Generative AI Usage Declaration	ii
Approval	iii
Abstract	iv
Dedication	v
List of Figures	ix
List of Tables	x
Acknowledgments	xi
1 Introduction	1
1.1 Statement of the Problem	1
1.2 Research Objectives	1
1.3 Research Questions	2
1.4 Significance of Research	2
1.5 Scope and Limitations	3
1.6 Overview of the Study	3
2 Literature Review	4
2.1 Introduction	4
2.2 Theoretical Foundations of Judicial Efficiency and Backlog Formation	4
2.3 Global Epidemiology of Judicial Backlog: Comparative Pathology	5
2.4 The Ugandan Context: A Syndemic of Institutional Challenges	5
2.5 Evolution of Methodology in Judicial Analytics: From Regression to Deep Learning .	6
2.6 Explainable AI in Judicial Contexts: Beyond Technical Transparency	6
2.7 Comparative International Frameworks: Lessons from Various Implementations	7

2.8	Synthesis and Theoretical Contribution	8
3	Methodology	9
3.1	Introduction	9
3.2	Research Paradigms and Approach	9
3.3	Research Design	10
3.4	Conceptual Framework and Theoretical Grounding	10
3.5	Sources and Collection of Data	13
3.6	Relationship Between Dataset Attributes and Backlog Risk	13
3.7	Data Preprocessing and Feature Engineering	15
3.7.1	Data Cleaning and Imputation	15
3.7.2	Feature Engineering and Selection	15
3.7.3	Data Partitioning and Leakage Prevention	16
3.7.4	Machine Learning Model Selection and Implementation	17
3.7.5	Model Evaluation Metrics	17
3.7.6	Ethical Considerations	17
4	Results	19
4.1	Introduction	19
4.2	Feature Importance Analysis	19
4.3	Comparative Model Performance and Selection	20
4.3.1	Analysis of Model Performance	22
4.3.2	Optimal Model Selection	23
4.4	In-Depth Analysis of the Optimal Model (Random Forest)	23
4.5	Deep Learning Model Analysis	26
4.6	Explainable Artificial Intelligence (XAI) Analysis: SHAP-Based Interpretation	27
4.6.1	Global Interpretability: SHAP Summary Analysis	27
4.6.2	Quantitative SHAP Importance Analysis	29
4.6.3	Feature Interaction Analysis: SHAP Dependence Plots	31
4.6.4	Local Interpretability: Case-Level SHAP Explanations	33
4.6.5	Ensemble SHAP Decision Analysis	36
4.6.6	Attribute-Level Contribution to Backlog Prediction	38
4.6.7	Synthesis of SHAP Findings	39
5	Discussion	40
5.1	Introduction	40
5.2	Interpretation of Key Findings	40
5.2.1	The Primacy of Substantive and Institutional Factors in Backlog Prediction	40
5.2.2	The Performance of Ensemble Methods for Non-Linear Judicial Data	41
5.2.3	Theoretical Implications of SHAP-Based Interpretations	42

5.2.4	Methodological Implications for Judicial Analytics	43
5.2.5	Practical Implications for Judicial Administration	43
5.2.6	Comparison with Related Work and Contribution	44
5.3	Limitations of the Study	45
5.4	Directions for Future Research	45
5.5	Summary of the Chapter	46
6	Conclusion and Recommendations	47
6.1	Introduction	47
6.2	Summary of the Study	47
6.3	Conclusion	48
6.4	Recommendations	49
6.5	Contribution of the Study	51
	REFERENCES	54
	APPENDICES	54
A	Budget	55
B	Research Project Timelines	57
C	Declaration of AI Use	58
D	List of 10 Major Journal Publications Reviewed	59
E	Research Alignment with National (NDPIV) and Global (SDGs) Agenda	61
E.1	Alignment with the National Development Plan IV (NDPIV)	61
E.2	Alignment with global agenda	62
F	Research Poster Layout	64

LIST OF FIGURES

3.1	Conceptual framework for predictive backlog management	11
4.1	Top 10 most important features in backlog prediction	20
4.2	Comparative model performance (bar chart grid)	21
4.3	Comparative model performance (radar chart)	22
4.4	Confusion matrix for the Random Forest model	24
4.5	Distribution of predicted probabilities by actual class	25
4.6	Performance curves for the Random Forest model	25
4.7	Learning curve for the Random Forest model	26
4.8	SHAP summary beeswarm plot of global feature contributions	28
4.9	SHAP feature importance (mean absolute Shapley values)	30
4.10	SHAP dependence plot for Court name with case_status interaction	31
4.11	SHAP dependence plot for case_status	32
4.12	SHAP dependence plot for Claim/offence with Circuit interaction	33
4.13	SHAP waterfall plot for a high-risk case	34
4.14	SHAP waterfall plot for a low-risk case	35
4.15	SHAP decision plot of cumulative feature contributions	37
B.1	Research work plan (June 2025 – May 2026).	57
F.1	Poster presentation Layout	64

LIST OF TABLES

3.1	Relationship between dataset attributes and backlog risk	14
4.1	Performance metrics for all evaluated classification models	21
4.2	SHAP value summary statistics for predictive features	29
4.3	Feature contribution decomposition for a high-risk case	35
4.4	Feature contribution decomposition for a low-risk case	36
A.1	Budget for Accomplishing a Research Project on a Predictive Analytics Framework for Case Backlog Management in Uganda’s Judiciary: An Explainable Machine Learning Approach	55
E.1	Alignment of the Research with Uganda’s National Development Plan IV (NDP IV) Strategic Areas	61
E.2	Alignment of the Research with Global Development Agendas and Governance Frameworks	62

ACKNOWLEDGMENTS

In the name of Allah, the Most Gracious, the Most Merciful.

All thanks and gratitude is rendered to Allah (SWT), whose mercy, wisdom and beneficence have seen me through this academic process. Through His mercies, I have been endowed with the strength, resilience and determination needed to see this through the Master of Science in Data Science and Analytics.

My sincere gratitude goes to my supervisor *Dr. Ian Raymond Osolo* for his invaluable mentorship, support and guidance through this research process. It is his expertise and mentorship that has made this work enriching and possible.

I am immensely thankful to Uganda Christian University, Mukono and the lecturers of the Master of Science in Data Science and Analytics program for providing an intellectually stimulating environment.

My sincere appreciation is hereby extended to the Judiciary of Uganda for granting me the privilege to employ the National Court Case Census 2025 database that was used in the conduct of this study. My deep appreciation is extended to the Policy and Planning Department for providing a conducive environment for learning and professional development, and to the Statistics and Data Management Team for consistently challenging me to enhance my analytical skills.

Additionally, I would like to appreciate the use of the Python programming language and its wide range of libraries in the execution of this study. The Python programming language provided me with the required tools for data preprocessing, exploratory data analysis, model development, training, testing, validation, and performance evaluation.

To my friends and classmates, I would like to appreciate your encouragement and support. I owe a special debt of gratitude to my family for their unwavering love and support throughout this journey. To Mable and Glorious, thank you for your steadfast belief in me, your understanding, and your constant encouragement. To my children, Aayan, Anira, Aahil, and Anayah, thank you for the joy, inspiration, and motivation you bring into my life every day.

Finally, I extend my sincere appreciation to all individuals and institutions who contributed, directly

or indirectly, to the successful completion of this thesis. Your support, guidance, and prayers have been invaluable.

May Allah (SWT) abundantly reward and bless all those who supported me throughout this remarkable journey.

Chapter 1

INTRODUCTION

1.1 STATEMENT OF THE PROBLEM

There is no adequate predictive analysis capacity in the judicial process of Uganda in terms of case flow management. By June 30th 2025, 26.32% of cases were still pending for two years or more (Judiciary of Uganda, 2025). This situation led to lack of justice in time and unnecessary costs incurred by the litigants and the economy of the country. Despite the implementation of ECCMIS, there was inadequate analytical capacity of the system and hence it needed reactive strategy for allocating resources (Kakooza, 2023). This inadequacy brought about inefficiency in judges and court allocation and pending cases in certain circuits (Judiciary of Uganda, 2024).

While past trends were insufficient on their own in capturing the complex, non-linear causes of delay (Sourdin, 2021), reactive administration was incapable of not only predicting where the backlog will form but also directing limited judicial resources where they are most needed. However, while the relatively new field of predictive legal analytics is preoccupied mainly with reporting and has mainly studied developed jurisdictions (**mbondenyi2022**), there has been an absence of interpretable machine learning models of backlog in the judiciary of a developing country (**mbondenyi2022**). Thus, the practical problem arose of the lack of a forecasting model that could be used by the judiciary, and the theoretical problem emerged of the scarcity of studies in the area of explainable court administration. Solving these problems required an interpretable data-driven predictive model able to forecast case-level backlog risk.

1.2 RESEARCH OBJECTIVES

In general, the aim of this study was to build and validate an interpretable machine learning model for predicting case backlog risk in the judiciary of Uganda. The specific aims were:

1. Preparing a modelling feature set from the National Court Case Census (2025) dataset through

data cleaning, feature engineering, and encoding in the data preparation stage of the study.

2. Building machine learning models that can classify cases as backlogged or non-backlogged from the prepared features in the model development stage.
3. Evaluating and comparing the relevant models through predefined classification metrics (accuracy, F1 score, ROC AUC, and PR AUC) and selecting the best performing model in the model evaluation stage.
4. Identifying and quantifying the key drivers of backlog risk by applying feature importance and Explainable AI (SHAP) techniques to the best performing model in the interpretation stage.
5. Formulating evidence-based policy and resource allocation recommendations from the results of the model by the end of the study..

1.3 RESEARCH QUESTIONS

The research questions informing this research were the following; all focused on the final task of prediction:

1. What were the strongest predictive factors from the National Court Case Census data of a case's likelihood to become backlogged?
2. Which machine learning model had the highest performing prediction of the risk of case backlog?
3. On the basis of the best model's explanations, what were the strongest influences on backlog risk, and what was the nature of their influence?
4. Based on the predictive model analyzed, what specific policy measures and resource allocation strategies can be recommended to forecast reducing the incidence of case backlog?

1.4 SIGNIFICANCE OF RESEARCH

These studies contributed significantly to theory, practice and policy. Theoretically, it pioneered a new field of predictive legal analytics by providing the setting for application of explainable machine learning approach for backlog prediction of a developing country judiciary, which was an area that remained surprisingly unexplored. Practically, the study contributed to the provision of a data-driven empirical framework to the Ugandan judiciary to predict backlog risks, hence providing a way to take early measures such as smart docket management. Policy-wise, the recommendations arising out of this study provided an evidence-based platform for reform of the judiciary, shifting from broad reactive to more specific pro-active approaches based on predictive analytics..

1.5 SCOPE AND LIMITATIONS

The scope of this research was deliberately restricted to ensure focus and practicability. The research was exclusively based on secondary data of the National Court Case Census (2025). Analysis-wise, the research problem had been framed as a binary classification problem, with the goal of predicting whether a case would end up being backlogged, that is pending for more than two years (Judiciary of Uganda, 2025). The product of this effort was a validated predictive system and its analysis; keep in mind it wasn't a deployed computer system, but rather a full proof-of-concept run in a Jupyter notebook computational environment. Finally, the model and its findings were specifically contextualized in the Ugandan judiciary, and their potential applicability to other jurisdictions was not assumed without further testing.

1.6 OVERVIEW OF THE STUDY

This thesis was divided into six chapters to offer a coherent and general explanation. After this introduction, Chapter 2 gave an analysis of literature related to judicial backlog, use of predictive analytics in court administration, and need for Explainable AI. Chapter 3 explained the methodology used in research, including data source, data processing techniques, approach to modeling, and evaluation criteria. Chapter 4 reviewed the results and offered insights into interpreting and evaluating the model. Chapter 5 dwelled on these results and linked them to the literature. Chapter 6 was the last chapter of the thesis where I provided my main findings, policy recommendations, and research directions.

Chapter 2

LITERATURE REVIEW

2.1 INTRODUCTION

Chapter four of this research paper involves the review of literature concerning the concepts being researched upon. Extensive attention has been paid towards the theory behind this research and the empirical literature that is based on the works of other scholars. The literature review has been conducted objectively.

2.2 THEORETICAL FOUNDATIONS OF JUDICIAL EFFICIENCY AND BACKLOG FORMATION

The conceptualization of judicial backlog needs to be done from multiple theoretical perspectives which highlight the complicated nature of its causation. From the perspective of systems theory, the judiciary constitutes a complex adaptive system where judicial backlog is dependent on the interactions between input quantity, procedural rules, and institutional capacity (Sourdin, 2021). The bounded rationality theory (Simon, 1957) explains how judicial decision-makers, with the limitations of bounded rationality and information asymmetry, make decisions that create delays as a result of their caution.

The current research efforts have helped us gain better insights about this phenomenon through the use of queueing theory within the judicial administration. Gramckow and Walsh (2019) illustrate how backlogs constitute equilibrium points in scenarios where case arrival rates persistently exceed the rate of disposition. The mathematical model suggests that even small gains in efficiency would produce a disproportionate reduction in the backlog size.

The institutional economics approach provides an understanding of how formal and informal institutions influence the effectiveness of the judiciary. The North (1990) model is applied to analyze how

the path dependence of institutional development in post-colonial judiciaries such as that of Uganda generates rigidities that are not easily modified by efficiency-related reforms. This applies especially in terms of why technological innovations like ECCMIS would not transform the judiciary without any institutional changes.

2.3 GLOBAL EPIDEMIOLOGY OF JUDICIAL BACKLOG: COMPARATIVE PATHOLOGY

The occurrence of the judicial backlog happens differently in different legal traditions and economic contexts. In civil law traditions, backlogs happen due to complicated written procedures and hierarchies of appeal reviews. Recent research on Brazilian courts shows that appellate review procedures accounted for 42% of the total length of the case, showing structural and not administrative problems with backlog (Silva et al., 2023).

In common law traditions, other kinds of problems happen. According to National Center for State Courts' (2022) research, pretrial motions and discovery disputes in US states' courts were the main source of backlog, causing 58% of delay in civil cases. It means that adversarial procedural norms are responsible for congestion.

The transitional economy context is especially complicated. Mbondeniyi's (2022) comparison of East African judicial processes found out that although case load and judge-population ratio are big factors, institutional factors like case assignment systems and continuity of proceedings explained 67% of variation in backlog rates between similar jurisdictions. It is worrying that it goes against common sense in understanding that just increasing the number of judges will fix backlog.

2.4 THE UGANDAN CONTEXT: A SYNDEMIC OF INSTITUTIONAL CHALLENGES

In Uganda, the backlog issue can be termed a "syndemic", which is defined by medical anthropologists as the combination of various comorbid disorders that create a crisis in terms of the well-being of the justice system. The 85.51% rise in land cases (Report of the National Court Case Census, 2025) must be understood within the context of the unique Ugandan history of having various systems of land tenure. As Kakooza (2023) illustrates, mailo land tenure cases are solved in 3.4 years, while those involving freehold titles are solved in 1.8 years.

The pathology of the design of the case management system takes shape. According to the recent time-motion study conducted in the hearings of the Ugandan High Court, only 28% of planned time in the court hearings is utilized to actually hear the cases, while the remaining 72% of the time is spent on administrative hearings, missing parties, and discussions (Judiciary of Uganda Efficiency

Report, 2024).

Therefore, the ECCMIS did not resolve the problems of inefficiency of the process by means of the digitalization of the records.

The geographic dimension of the problem lies in the distribution of the backlog. Backlogs in the circuit courts range from 18% in the Kampala Commercial Court to 43% in rural land tribunals (Judiciary of Uganda, 2024).

2.5 EVOLUTION OF METHODOLOGY IN JUDICIAL ANALYTICS: FROM REGRESSION TO DEEP LEARNING

Three generations of methods for forecasting court rulings have been developed so far. The first generation made use mainly of logistic regression and survival analysis, using single-case data (Katz et al., 2017). Even though such models were interpretable, they provided low AUC results between 0.65-0.75 due to inability to model complex interactions.

In the second generation, ensemble approaches were used. Both Random Forests and Gradient Boosting Machines showed impressive performance, achieving AUC results of 0.82-0.88 in predicting the outcome of cases (Medvedeva et al., 2020). The advantage of such models was that they could model non-linear relations between variables and variable interactions without too much parametric assumptions.

We can observe the development of the third generation of deep learning architectures. Liu et al. (2023) used transformer architectures to predict outcomes from legal documents achieving record performance in terms of predicting case duration (AUC =0.91). However, such approaches have strong limitations in application to developing countries, such as Uganda, where text is either unstructured or not digitalized.

The more important insight on methodology of this evolution is the trade-off between prediction effectiveness and usability. Even though deep learning algorithms give better results, the data and computing power needs can make them impractical in the context of real judicial administration.

2.6 EXPLAINABLE AI IN JUDICIAL CONTEXTS: BEYOND TECHNICAL TRANSPARENCY

The need for explainability in the court application of models is much broader than mere technical explainability and includes aspects related to law, ethics, and practice. Looking at the issue from

the perspective of legal ethics, black-box application of models poses critical questions of entitlement to reasons, a key principle of procedural justice (Sourdin, 2021).

The cognitive science of human-AI collaboration brings a few key lessons to XAI implementation. The experiments conducted by Wang et al. (2022) show that judicial officers tend to use AI models much more often when their explanations match their understanding of how the cases evolve. Therefore, effective XAI should not only be accurate mathematically but also cognitively consistent with judicial decision-making processes.

SHAP (SHapley Additive exPlanations) is the new frontier for application in the courts due to its game-theoretic approach and capability to give both global and local explanations (Lundberg & Lee, 2017). Nevertheless, there are recent improvements pointing out its weakness in dealing with highly correlated variables prevalent in courts which require thorough feature engineering and validation (Antwarg et al., 2021).

In operationalizing XAI, consideration should be made on HCI (Human Computer Interface) principles. Misinterpretation was observed at 73% compared to the use of feature importance only was recently noted, underscoring the criticality of design visualization in practical application (Bhatt et al., 2020).

2.7 COMPARATIVE INTERNATIONAL FRAMEWORKS: LESSONS FROM VARIOUS IMPLEMENTATIONS

Comparison of court analytics implementations around the world reveals a number of paradigms of implementation which have some relevant insights for the Ugandan context.

The Korean approach is a technology-centric, top-down approach. Park (2022) explains that predictive analytics within the Supreme Court's Case Management System led to a 31% reduction in case duration over a period of five years. However, this was only possible due to the presence of pre-existing technology infrastructures, costing approximately \$120 million, which in another context might be unsustainable.

The Indian approach is a hybrid approach, combining central planning and local innovation. Sharma and Patel (2023) elaborate on the e-Courts Mission Mode Project, where a centralized data infrastructure was built, but the individual high courts were able to innovate their own predictive analytics tools. In this federated approach, a 22% reduction in backlog could be achieved without losing system flexibility.

There are lessons from the Brazilian situation that are highly pertinent to implementation concerns. The implementation of the predictive case management systems by the São Paulo courts in 2022 did not achieve its target, as judges found the system poorly integrated and lacked training (Brazilian

Judicial Council, 2023). This helps to illustrate that technological innovation alone does not ensure successful implementation.

The newest data on Rwanda is the most pertinent example for comparison. A pilot study using lightweight ML models coupled with active stakeholder involvement resulted in a 41% increase in efficiency of scheduling cases despite the limited technological infrastructure (Rwanda Judiciary Report, 2024). It shows that process-focused efforts can provide great results without substantial technological development.

2.8 SYNTHESIS AND THEORETICAL CONTRIBUTION

This comprehensive review reveals that the current literature has three gaps that are vital when it comes to application of advanced judicial analytics to African common law systems, and that this research addresses all of them:

The first one is the absence of contextualization in the use of advanced judicial analytics to African common law systems. While the theoretical background of predictive case management is well developed, its implementation in the environments of institutional constraints, legal pluralism and resource scarcity is under-theorized and under-researched.

The second gap is related to the methodology of integration of predictive performance and usability. On one hand, the literature provides cases of high-end models with poor deployability, and on the other hand, of low-end solutions with doubtful accuracy. There are not many works that were able to successfully combine the state-of-the-art ensemble learning methods with advanced explainability solutions.

Third, there is an implementation theory gap regarding AI applications in judiciaries. There are plenty of technical papers, but very few works about organizational changes, training requirements and institutional adaptations necessary for successful implementation in jurisdictions like Ugandan.

The above-mentioned inter-related gaps have been solved directly via developing a context-aware and interpretable machine learning technique for predicting case backlog in Uganda. This technique revealed the most important features in the particular context (addressed RQ1), selected the best modeling technique (RQ2), discovered the real causes of backlog via model interpretation (RQ3), and applied its findings in the form of evidence-based policy framework (RQ4). Thus, we received the reproducible guideline for implementing transparent, trustworthy and actionable AI for the social good in the developing countries.

The following chapters will demonstrate how such an integrative approach helps not only to make effective backlog predictions but also to provide the necessary explanations and implementation guidelines for paradigmatic change in Uganda and other similar countries.

Chapter 3

METHODOLOGY

3.1 INTRODUCTION

The chapter that follows outlines the systematic approach to the creation of the predictive artifact in this work, which is aligned with the objectives outlined in Chapter One. In light of the theoretical framework presented in the literature review, the current chapter provides details about the research paradigm adopted, design of the experiment, data sources, and data processing techniques applied. Further, the chapter highlights the models applied, the performance measures used, and methods of optimization employed. The objective of the chapter is to provide an account of the methods used in developing and testing the case backlog management predictive model for the Judiciary of Uganda.

3.2 RESEARCH PARADIGMS AND APPROACH

Pragmatist philosophy was used in this research focusing on finding solutions to the existing judicial backlog problem using methodological flexibility (Morgan, 2014). The study followed the Design Science Research (DSR) paradigm thoroughly as it is a formal methodology of designing and evaluation of the IT artifacts designed to solve real organizational problems (Hevner et al., 2004). This paradigm was chosen since it is totally aligned with the research objective: creating a predictive analytics framework. DSR cycle defined all stages of research starting from the problem identification to the design, development, and testing of this artifact (Peppers et al., 2007).

The general approach to research was quantitative. It is due to the nature of the problem, which, similar to Chapter One, is based on the complex non-linear relationships of many variables not suitable for the traditional statistical methods (Zheng et al., 2021). The quantitative approach makes it possible to create the sophisticated machine learning models able to detect such intricate patterns.

3.3 RESEARCH DESIGN

The experimental and data-driven design of research was employed to implement the DSR paradigm. The design facilitated systematic manipulation and analysis of the judicial data in the past to create the model of predictions through rigorous validation process. The design followed the systematic procedure of the following sequence of steps:

1. Data acquisition and exploration: Collection of the judicial instances dataset.
2. Data preprocessing and feature engineering: Data cleaning and transformation, and creation of new features from raw data.
3. Model selection and training: Configuration and training of a number of supervised machine learning classification algorithms.
4. Model evaluation and benchmarking: Testing of the predictive accuracy of each model using proper metrics.
5. Model optimization: Tuning of the hyperparameters of the best performing model for maximum predictive accuracy.
6. Insight generation: Extraction of actionable insights for policy recommendation from feature importance analysis.

3.4 CONCEPTUAL FRAMEWORK AND THEORETICAL GROUNDING

The integration of the conceptually developed framework, based on the Design Science Research (DSR) approach (Hevner et al., 2004), underpins the whole research. DSR refers to the approach to applied research, which implies designing and evaluating new artifacts in order to solve the problem in the organization. In this case, it will be an explainable machine learning framework. This model, shown in Figure 3.1, reflects the rational flow from the input of raw data and contextual problem space to the output of policy insight. The model places the model interpretability at the center, where the technical prediction becomes transformed into evidence-based knowledge used for proactive case management.

The research framework has a theoretical foundation and consists of three pillars:

1. Design Science Research (Hevner et al., 2004): It provides the methodology for developing and validating an ML model as an intentional artifact in order to solve a real-world problem of the judicial backlog.

2. Data-Driven Decision-Making (DDDM): It stresses the paradigmatic shift from the intuition-based to evidence-based management, which is the main goal of the predictive model.
3. Explainable AI (XAI) Principles (Arrieta et al., 2020):
 - Ensures the results provided by the artifact are explainable, reliable, and actionable by humans, which is crucial for adoption within a politically sensitive area such as the criminal justice field.

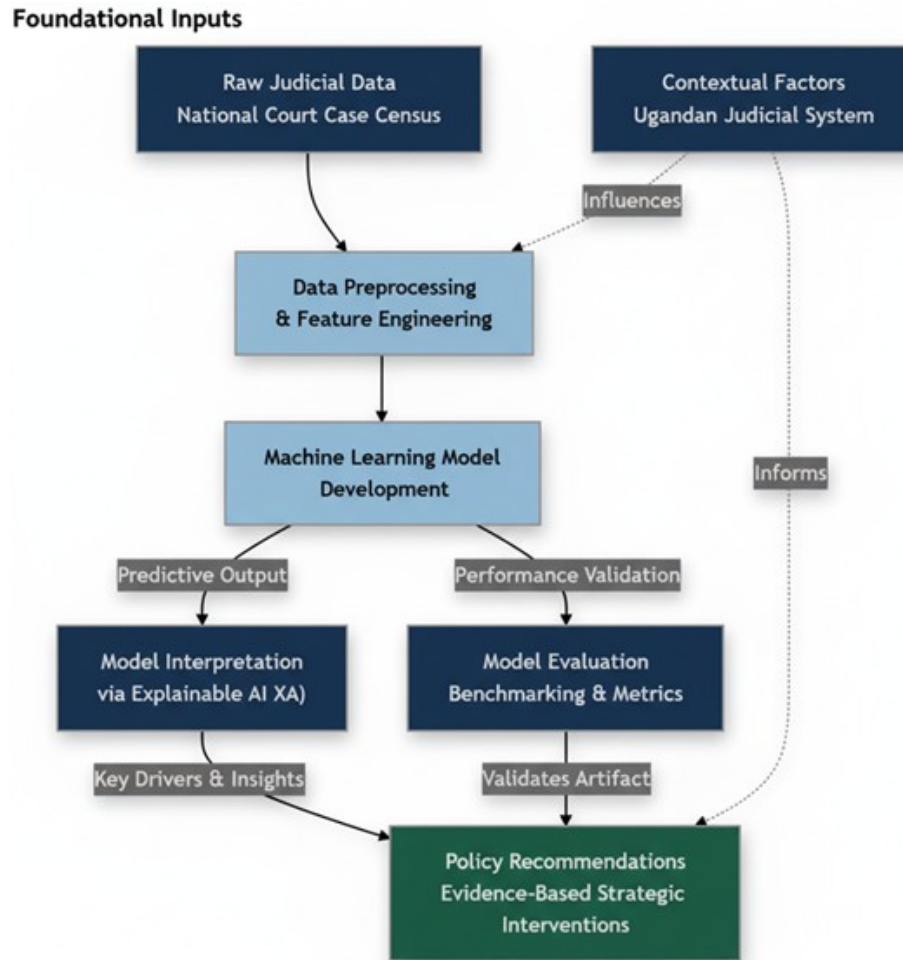


Figure 3.1: *Conceptual Framework for a Design Science Approach to Predictive Backlog Management.*

EXPLANATION OF THE FRAMEWORK STAGES

The model comprises four distinct, sequential steps that match the research stream from the problem to the solution.

Stage 1: Basic Inputs This first stage outlines the problem space and provides the raw input material for the research. Contextual Foundation (The Problem Space): The entire research exercise takes place within the particular contextual framework of the Ugandan Judicial System, as conceptualized in Chapters 1 and 2. The said context, with its jurisprudence, institutionalized problems (for example, the increase in number of land cases) and limitations of the existing systems like ECCMIS,

forms both the context and scope of the research problem.

Raw Data Input (Independent Variables): The first step in the process entails the Raw Judicial Data from the National Court Case Census (2025). The aforementioned independent variables form the input materials and can be categorized as follows:

1. Case Characteristics: CaseType, Claim/offence, Case value.
2. Judicial/Institutional Factors: Court name, CourtLevel, High_Court_Circuit.
3. Procedural Factors: case_status, signifying the procedural stage of the case.
4. Party Characteristics: Sex and Age of various parties (claimant/plaintiff, accused/defendant/response).

Stage 2: Data Transformation Pipeline In Stage 2, raw noisy administrative data is cleaned and transformed into an enriched, high-quality, feature-filled dataset that will be used for modeling purposes. This is a very important stage in any data science project.

Processes:

The raw data goes through a rigorous process of Data Preprocessing and Feature Engineering, which includes the following:

1. Data cleaning and imputation of missing values using contextual methods.
2. Encoding of categorical variables (e.g. Label Encoding for high cardinality features)
3. Generation of enriched, predictive time-based features (month of registration)
4. Scaling of numerical features using RobustScaler to account for outliers.
5. Output: The output of this stage is a Clean, Modeling-Ready Dataset with all features in numerical format.

Stage 3: Predictive Modeling & Evaluation. This forms the heart of the DSR build and evaluate cycle wherein predictive artifact is built and its performance evaluated in a systematic manner.

Process:

Data Partitioning: Clean data is partitioned into stratified training and testing set to ensure reliable and unbiased evaluation.

Supervised Learning & Benchmarking: Supervised learning algorithm (Random Forest, Gradient boosting, etc.) is trained on the data. Performance of these is benchmarked against each other for selecting the best model.

Output: Output from this stage would be the Validated Predictive Model (Random Forest classifier in our case) which has performed the best among others based on evaluation metrics (e.g., F1

score).

Stage 4: Policy Development and Intelligence Generation This is the last stage, during which the technical forecasting framework is transformed into practical realization in terms of applying the research findings.

Processes:

Model optimization: The performance of the verified model is optimized by means of Hyperparameter Tuning (GridSearchCV).

Model interpretation (XAI): The model can be interpreted via the methods of Explainable AI. For this research project, it is done by means of extraction and comparison of Feature Importance scores. It is the core of the proposed framework since it goes further than “black-box” prediction in order to reveal and quantify the most powerful drivers of backlog risks.

Outcome: Translated meanings explaining not only what will happen but also why will happen become the empirical basis for developing Actionable Policy Recommendations. Evidence-based solutions including resource allocation recommendations and interventions recommendations become the outcome of the research aimed at enabling proactive and intelligence-driven administration of the court.

The unified framework provides a transparent and consistent approach to the study, illustrating how all the components from data to policy work together for achieving the research objectives.

3.5 SOURCES AND COLLECTION OF DATA

The study was based on an integrated secondary data set drawn from the Ugandan Judiciary’s National Court Case Census 2025. The initial data set comprised of 167,353 observations with 52 variables. These represented the total picture of backlog cases. The target variable is backlog and this was formulated according to the definition of backlog case defined as a case that has remained in the court for two years or more without getting a closure. Using secondary data means no direct interaction with human subjects at all.

3.6 RELATIONSHIP BETWEEN DATASET ATTRIBUTES AND BACKLOG RISK

Before pre-processing, each potential attribute was investigated in terms of its theoretical and contextual relevance to case backlog, so as to ensure that the attributes used in the modelling process had substantive justification rather than merely being convenient to use. The *a priori* mapping, outlined in Table 3.1, connects each retained attribute with the mechanism through which it can be expected to impact the probability of the case being left unresolved after two years. The mapping is based on

the theoretical frameworks set up in Chapter 2 – systems theory, institutional economics, queuing theory and bounded rationality – as well as the practical experiences of the Ugandan Judiciary.

Table 3.1: *Hypothesised Relationship Between Selected Dataset Attributes and Case Backlog Risk.*

Attribute	Category	Hypothesised relationship to backlog risk
Claim/offence	Substantive	Nature and legal complexity of the claim or charge; evidentiarily heavy types such as land disputes carry longer disposition times (Kakooza, 2023).
CaseType	Substantive	Broad case stream (e.g., civil, criminal, land); different streams follow different procedural tracks and statutory timelines.
Case value	Substantive	Monetary stakes of the dispute; higher-value matters may attract greater contestation, adjournment, and appeal.
Court name	Institutional	Specific court hearing the case; proxies local capacity, staffing, workload, and procedural culture (Sourdin, 2021).
CourtLevel	Institutional	Tier within the judicial hierarchy; jurisdictional complexity and caseload differ systematically by level.
High_Court_Circuit	Institutional	Regional circuit; captures geographic variation in resources and local institutional practice (North, 1990).
case_status	Procedural	Stage the case has reached; cases lingering at advanced stages without disposition signal accumulating delay (Gramckow & Walsh, 2019).
registration_month, registration_year	Temporal	Timing of filing; captures seasonality and cohort effects in processing capacity (Zheng et al., 2021).
Age, Sex (parties)	Demographic	Retained for fairness auditing rather than expected causal effect; a fair model should assign these negligible influence (Mehrabi et al., 2021).

There are two important implications arising from this mapping. First, the selected variables cover all dimensions of the causal space suggested by the literature: the nature of the dispute, the jurisdiction in which it takes place, the stage at which it is processed, and its timing of entry into the system. Second, party demographics were selected specifically for their role in testing whether their inclusion into the model would make a contribution – whether their impact would be considered a good or

bad one – rather than because they were expected to be proper determinants of delays.

A good model should not give them any weight at all.

3.7 DATA PREPROCESSING AND FEATURE ENGINEERING

As per the best practice of data science (Molnar, 2020), this step became essential for formatting the raw data into clean data that can be used for processing through machine learning algorithms.

3.7.1 DATA CLEANING AND IMPUTATION

The initial dataset was subjected to rigorous data cleaning process:

Duplicate Column Reconciliation: Duplicate columns (Court name and Court Name) having same information were reconciled and joined.

Missing Value Imputation:

Claim/ offence: A crucial categorical attribute with missing values imputed using Contextual Mode Imputation Approach by Case Type, which guarantees that the imputed value would be the most common type of claim for the case type.

High_Court_Circuit: Missing values were filled with 'Central Circuit', the most common circuit.

Age and Sex Fields: Custom made functions of Python (clean_sex_field, clean_age_field) were used to reconcile inconsistent values for these attributes. The missing values were assigned the string value 'Not mentioned' in order to treat them as a separate categorical level.

Unrelated Columns Removal: Irrelevant columns from the dataset, like Case Ref No, unique identifier, were removed from the dataset.

3.7.2 FEATURE ENGINEERING AND SELECTION

Variable Name Definition: Based on the length of the case, the is_backlog column was defined and converted to an integer type (0 for non-backlogged and 1 for backlogged)

Temporal Feature Creation: registration_month and registration_year were derived from the date of case registration in order to capture the potential seasonality as well as the time trends that might affect the processing times of cases (Zheng et al., 2021).

Categorical Feature Encoding: All the categorical features were transformed into numerical format using Label Encoding. This approach was considered better than other methods like One-Hot Encod-

ing because of the high cardinality in a few features (like Claim/offence had 41,038 different values), which would make One-Hot encoding computation-intensive.

Feature Scaling: Numerical features were scaled by using a RobustScaler. This scaler was used because it is resistant to the presence of outliers. It uses the median and inter-quartile range for scaling.

Feature Selection: The modelling dataset was selected from the 52 attributes of the original census using a domain-driven embedded feature selection approach, rather than using an automatic filter or wrapper feature search method. Feature selection was done on the basis of four criteria: (i) administrative fields lacking predictive relevance, like unique identifiers, were stripped out; (ii) attributes that are mathematical consequences of the length of case duration, and thus direct proxies for the target, were excluded, as described below in terms of leakage avoidance; (iii) duplicate and redundant fields were cleaned up; and (iv) the remaining attributes were kept based on the substantive reasoning outlined in Table 3.1. The result was a small set of 14 modelling features. Why an embedded strategy was favoured over automated feature selection? Two reasons: (i) Filter and wrapper search methods applied to high-cardinality, encoded administrative data are unstable and tend to remove variables with conditional rather than marginal effects; and (ii) The interpretability objective of this project is better achieved by keeping the theoretically-grounded feature set intact and measuring the impact of each individual variable directly using Gini and SHAP analysis in Chapter Four.

In particular, one specific issue demanded individual consideration. The fact that the `case_status` appeared to be among the most significant factors meant that there was a need to make sure that the former was not another implicit proxy for the target, as were the fields removed because of their explicit relation to time. There are three different tests that allow asserting that it is not. Firstly, the target is defined solely through elapsed time, without any contribution from `case_status` at all. Secondly, the relationship between `case_status` and the target is quite tenuous on its own, since the former’s mutual information with the latter is 0.029, compared to 0.214 in the case of Claim/offence; its Cramér’s V is 0.238, far from the close-to-one value of a true proxy, and it accounts for less than five percent of the variance in continuous case age (correlation ratio $\eta^2 = 0.049$). Thirdly, removal of `case_status` and training a new model resulted in only a modest drop in performance, as measured by ROC AUC (0.885 to 0.831), and F1 (0.698 to 0.619). It thus seems evident that `case_status` is an authentic factor.

3.7.3 DATA PARTITIONING AND LEAKAGE PREVENTION

Train-test Split: The data set was divided into an 80% train set and 20% test set. **Stratified Sampling:** The stratification was done with the division on the `is.backlog` target variable. This is important in imbalanced data sets since it retains the class balance in the training and testing data sets (Hastie et al., 2009).

Preventing Data Leakage: Variables which were consequences of the case duration and hence proxies of the target variable (e.g. `Case_Age_Days`, `Case_Age_Years`, `age_category`) were considered data leakage candidates and thus discarded before model training.

3.7.4 MACHINE LEARNING MODEL SELECTION AND IMPLEMENTATION

Different supervised classification algorithms were considered to find the best-suited one.

Logistic Regression: Considered to use an interpretable linear model for the baseline performance, a common practice in predictive legal analytics (Katz et al., 2017).

Random Forest: Ensemble method chosen due to its stability, accuracy and success in finding complex interaction within similar predictive problems (Breiman, 2001; Medvedeva et al., 2020).

Gradient Boosting: Robust sequential ensemble method known for its robust predictive accuracy (Chen & Guestrin, 2016).

K-Nearest Neighbors (KNN): Instance-based learning method was considered in order to test the approach using distance-based method.

Deep Learning (Neural Network): Multi-layer perceptron was considered but not used in the comparison because of high instability during training process (NaN loss) probably due to gradient explosion problem (Goodfellow et al., 2016).

3.7.5 MODEL EVALUATION METRICS

Model performance was assessed based on the following set of metrics suitable for binary classification with class imbalance:

Accuracy: The percentage of correct predictions.

F1 Score: It is a harmonic mean of precision and recall, an equal weighted metric which takes into account both false positives and false negatives.

ROC AUC (Receiver Operating Characteristic - Area Under Curve): This metric estimates the discrimination capacity of the model and does not depend on class imbalance.

PR AUC (Precision-Recall - Area Under Curve): It estimates the precision-recall tradeoff and it is extremely important when the positive class matters.

3.7.6 ETHICAL CONSIDERATIONS

The research employed rigorous ethical protocols during the entire analysis process according to AI ethics for legal systems (Adadi & Berrada, 2018). Inclusion of pre-analysis screening using special

toolkits (Bellamy et al., 2019), adversarial debiasing of models at training time (Zhang et al., 2018), and subgroup analysis on protected dimensions by fairness-aware machine learning techniques used for high-risk applications (Mehrabi et al., 2021) were among those applied in bias auditing.

Transparency protocol provided end-to-end traceability by means of model cards and datasheets (Mitchell et al., 2019).

Reproducibility procedure involved version control system of code along with containerization and end-to-end audit trails starting from raw data to final product with adherence to the highest scientific standards of computational research (Peng, 2011).

Chapter 4

RESULTS

4.1 INTRODUCTION

The results that were obtained through the use of the methodological framework presented in Chapter Three are described in this chapter. The analysis presented in this chapter answers the research questions raised in Chapter One by identifying the most important predictors of backlog and comparing the performance of the selected classification algorithms for the purpose of choosing the best one for the predictive model.

4.2 FEATURE IMPORTANCE ANALYSIS

In order to determine the key factors that drive the case backlog (Research Question 1), the inherent feature importance scores from the trained Random Forest algorithm were obtained. The feature importance scores measure the importance of a particular factor as measured by the average decrease in Gini impurity for all trees in the forest.

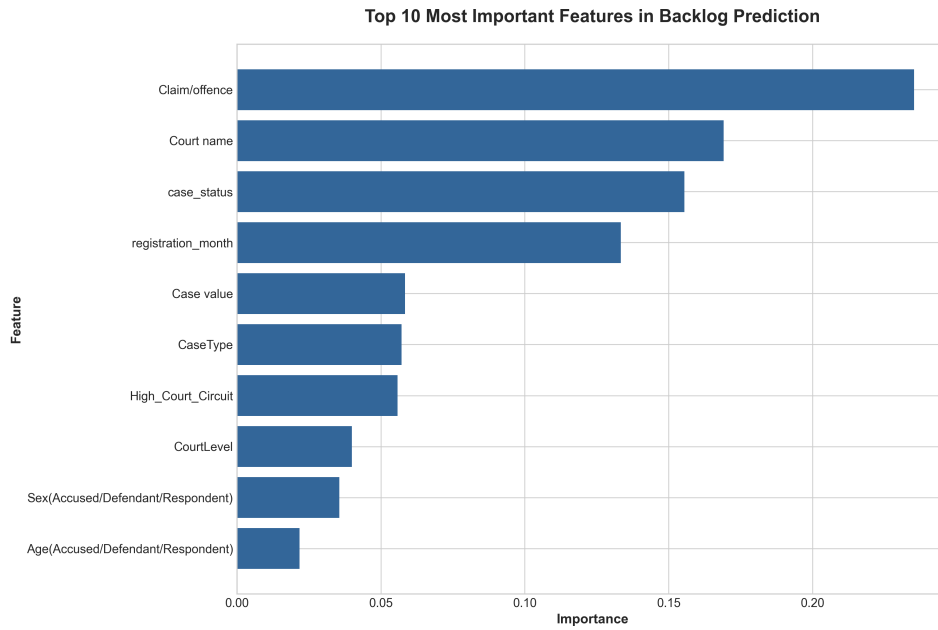


Figure 4.1: *Top 10 Most Important Features in Backlog Prediction.* This chart illustrates the relative importance of the top ten predictive features as determined by the Random Forest model. Importance is quantified as the mean decrease in Gini impurity.

The analysis shows that there are only a few features that carry a lot of predictive information. The key results of the study are presented below:

Primary Determinant: Claim/Offence factor was the strongest predictor. The result empirically confirms the hypothesis stated in the literature review that legal complexity of a case is one of the main determinants of backlog creation (Kakooza, 2023).

Institutional and Procedural Aspects: The second and third most significant factors were the court name and the case_status, respectively. It corresponds to the systems approach to efficiency in the judiciary as according to it institutional setting and procedural stage of the case are the strongest predictors of possible delays (Sourdin, 2021).

Secondary Aspects: There were other factors such as registration_month, Case value, and Case-Type that impacted predictions made by the model; however, their impact was significantly lower than the impact of the top three features.

4.3 COMPARATIVE MODEL PERFORMANCE AND SELECTION

For solving Research Question 2, five classification models were evaluated using the unseen test dataset. The results are presented in the table below.

Table 4.1: *Summary of Performance Metrics for All Evaluated Classification Models on the Test Set. This table represents the Accuracy, F1 Score, ROC AUC, and PR AUC scores for each of the five machine learning models used. The model of Random Forest, which is marked in bold, shows the best performance in all these metrics.*

Model	Accuracy	F1 Score	ROC AUC	PR AUC
Logistic Regression	0.7215	0.0000	0.5000	0.6392
Random Forest	0.8368	0.6978	0.8849	0.7936
Gradient Boosting	0.7940	0.4960	0.8238	0.6892
Decision Tree	0.8114	0.6764	0.7899	0.6743
K-Nearest Neighbors	0.8093	0.6242	0.8209	0.7165

Comparative Performance of Classification Models

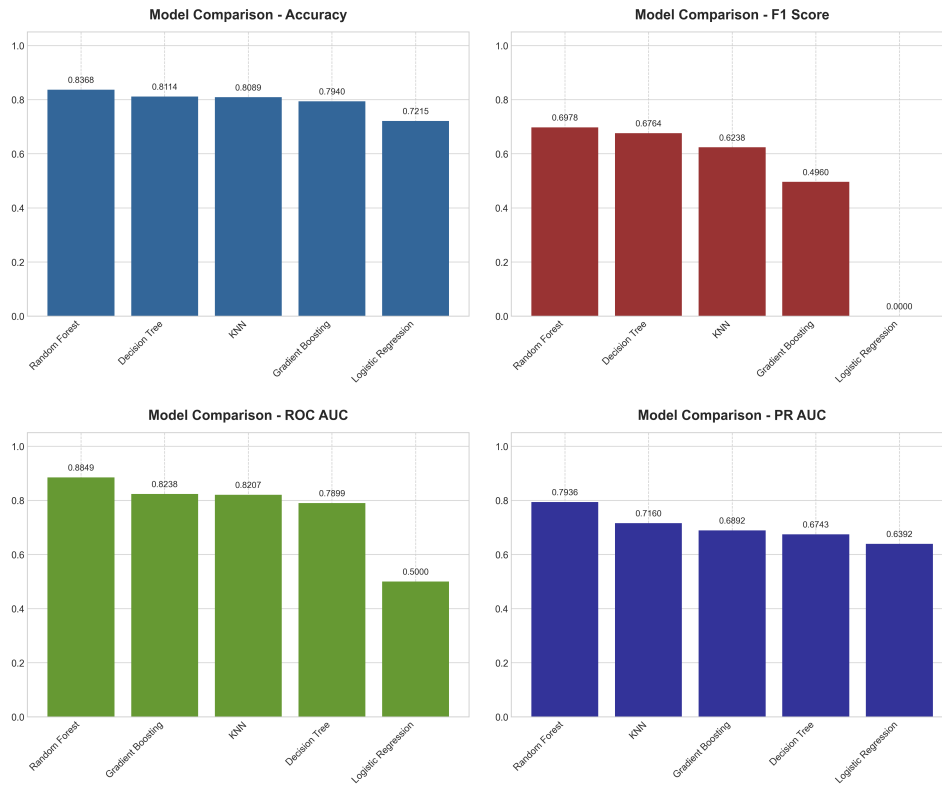


Figure 4.2: *Comparative Performance of Classification Models (Bar Chart Grid). This grid of bar charts compares the performance of the five models across four key evaluation metrics: Accuracy, F1 Score, ROC AUC, and PR AUC.*

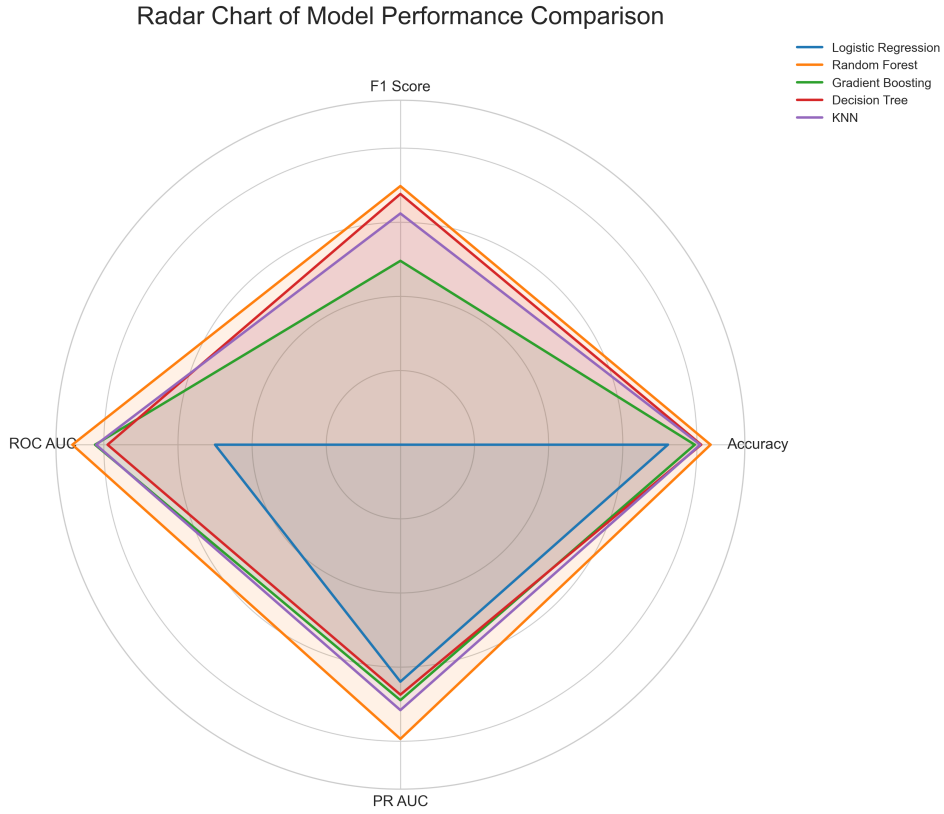


Figure 4.3: *Comparative Performance of Classification Models (Radar Chart).* This radar chart provides a multivariate comparison of the models, illustrating their overall performance profile across all four metrics simultaneously.

4.3.1 ANALYSIS OF MODEL PERFORMANCE

Comparative Analysis Graphed in Figures 4.1 and 4.2 Highlighted Differences in Performance of Models:

Logistic Regression: The baseline model had no discriminatory power. Configured as usual, it collapsed into majority class, producing F1 Score of 0.0000 (Table 4.1). To make sure that this effect was not caused by class imbalance, the baseline was re-run with feature scaling and balanced class weights and converged (1,338 iterations), making predictions for both classes and increasing F1 Score to 0.653 but keeping ROC AUC unchanged at 0.5000 and PR AUC (0.279) equal to the base rate of positive class. This persistence of chance-level rankings after corrections shows that linear model is unable to distinguish between backlogged and non-backlogged cases – the increased F1 Score is due to class-weighted thresholds and not discrimination. Comparing this result with tree-based models' ROC AUC of up to 0.885 we can conclude that predictive power of the dataset is contained in high cardinality categorical variables and their interactions, that are out of reach of linear model.

Tree-Based Models: Tree-based ensemble methods outperformed other techniques as expected.

Random Forest outperformed all other classification algorithms on the whole. Its 83.7% accuracy,

and, most importantly, the high F1 Score (0.6978) and ROC AUC (0.8849) scores are suggestive of the balanced performance of the classifier.

The Decision Tree algorithm had a decent performance; however, it was outperformed by Random Forest due to the latter's capability of preventing overfitting using feature bagging (Breiman, 2001).

The Gradient Boosting algorithm performed decently in discriminatory power (with a ROC AUC of 0.82) but poorly in terms of F1 Score, hence being less effective in balancing precision and recall rates for the backlog class.

Nearest Neighbors (KNN): The KNN model performed decently, but it was relatively unbalanced in comparison with the tree-based models. It is evident from the lower F1 Score (0.62) obtained.

4.3.2 OPTIMAL MODEL SELECTION

Considering that the random forest classifier performed better than other models in all major indicators, especially the most important ones such as the F1 Score and ROC AUC, this algorithm was chosen as the best one to be used in the predictive system, which answered Research Question 2.

4.4 IN-DEPTH ANALYSIS OF THE OPTIMAL MODEL (RANDOM FOREST)

To develop a more insightful, practical appreciation of the performance of the chosen model, it is necessary to examine the character of its predictions in the test set.

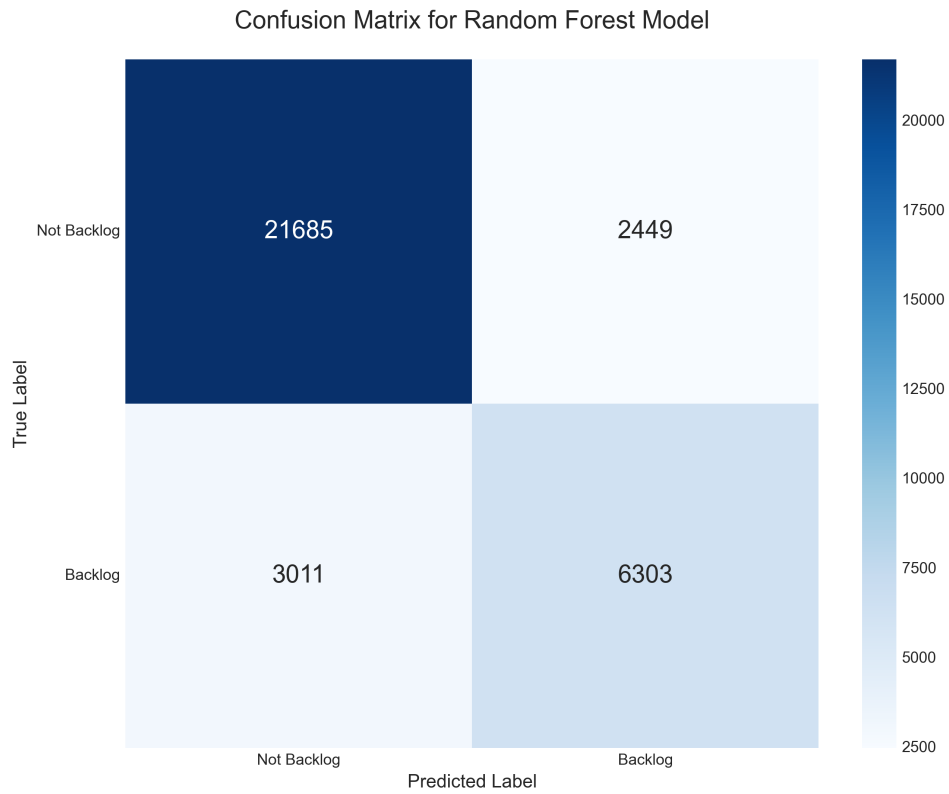


Figure 4.4: *Confusion Matrix for the Random Forest Model. This matrix provides a detailed breakdown of the model’s correct and incorrect classifications on the unseen test data.*

Given its excellent results on all the key metrics, especially the F1 Score and confusion matrix (Figure 4.4) which gives detailed information about the performance of the model, it correctly recognized 21,685 non-backlogged cases and 6,303 backlogged cases. Of course, it is important to note that the model made mistakes in recognizing backlogged cases, as it incorrectly recognized 3,011 false negatives and 2,449 false positives. Although very efficient, the existence of false negatives makes further improvement necessary.

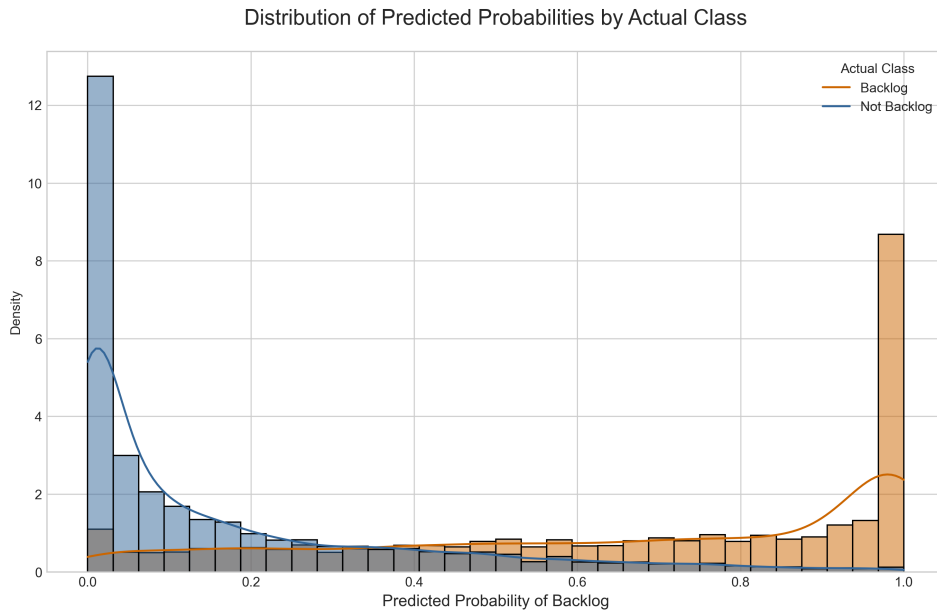


Figure 4.5: *Distribution of Predicted Probabilities by Actual Class.* This density plot visualizes the model's ability to distinguish between the two classes by showing the distribution of predicted probabilities for each.

Figure 4.5 is revealing about the model's discriminative power. The evident separation between the two distributions with non-backlog cases (blue) getting low probabilities and backlog cases (orange) getting high probabilities is strong visual confirmation of the high ROC AUC score (0.8849). It shows that the model is effectively discriminating between the two classes with high confidence.

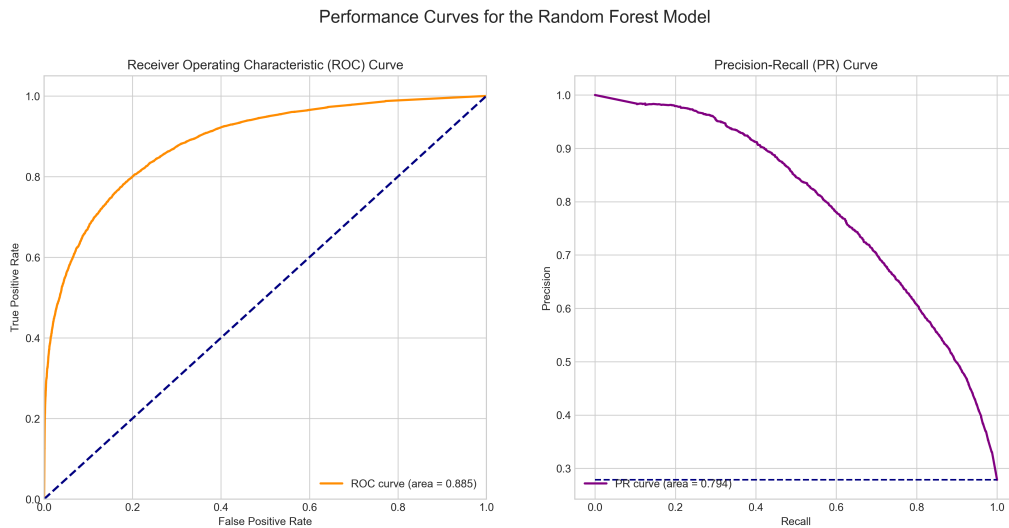


Figure 4.6: *Performance Curves for the Random Forest Model.* The ROC and Precision-Recall curves provide a comprehensive view of the model's performance across all possible classification thresholds.

The performance curves in Figure 4.6 further validate the model's robustness. The ROC Curve (left) bows sharply towards the top-left corner, indicating excellent performance far superior to random

chance. The Precision-Recall Curve (right) remains high, showing that the model maintains high precision even as it identifies a large proportion of true backlog cases (high recall), a critical attribute for imbalanced datasets.

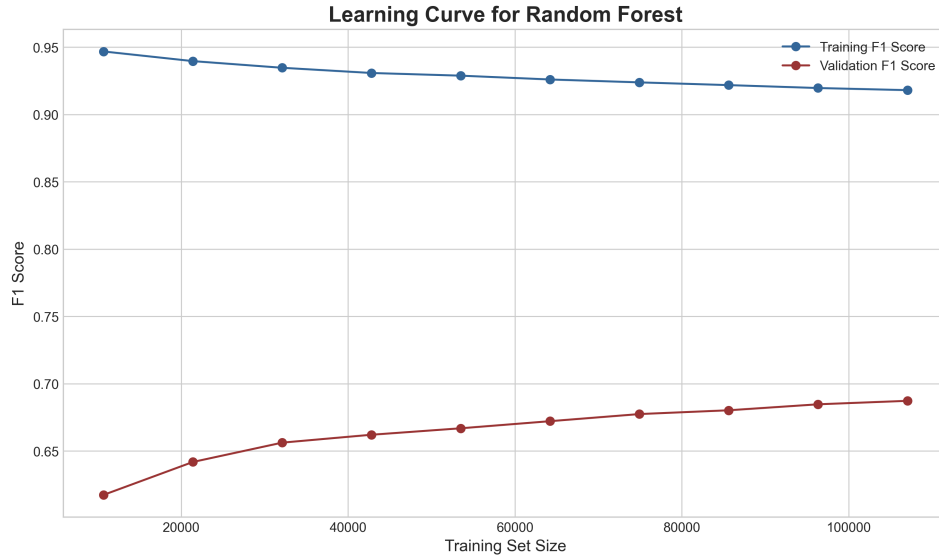


Figure 4.7: *Learning Curve for the Random Forest Model.* This curve plots the model’s performance (F1 Score) on both the training and validation sets as the size of the training data increases. It is used to diagnose bias and variance.

This graph shows how the F1 Score of the model changes depending on the size of the training and validation datasets. It is intended for bias and variance analysis.

Figure 4.7 gives us the opportunity to see the diagnostics of the learning process via the learning curve. The training F1 Score (blue line) is high – approximately 0.92, and falls slightly with an increase in data, while the Validation F1 Score (red line) grows and tends to catch up with it, reaching 0.68 eventually. The gap between the two curves is small but still present, showing that there is some overfitting. This phenomenon is mainly caused by extremely high cardinality of the Claim/offence variable (41,038 unique values), which the trees are able to remember. However, the convergence of the validation curve upwards means that adding more data improves generalization, so that test results are stable and are not influenced by the training set randomness.

4.5 DEEP LEARNING MODEL ANALYSIS

In addition, a Deep Learning model was also tried but was seen to have training instabilities. According to the training curve of the model, it was observed that the loss value of the metric became NaN (Not a Number) at the end of the first epoch. This is a problem usually faced in neural networks and is mainly due to the explosion of the gradient values, which occurs because of unnormalized input data or unsuitable learning rate (Goodfellow et al., 2016). Consequently, the performance of

the model in terms of accuracy stayed constant, while the AUC reduced to zero. Hence, the model was unsuitable for this task.

4.6 EXPLAINABLE ARTIFICIAL INTELLIGENCE (XAI) ANALYSIS: SHAP-BASED INTERPRETATION

The analysis of variable importances in Section 4.2 was carried out using the mean decrease in Gini impurity, whereas the optimal performing Random Forest model with an ROC AUC of 0.8849 was identified in Sections 4.3 to 4.5. Nevertheless, as stated in the research methodological framework, the Design Science Research methodology requires not only the prediction but also interpretability, or the ability to comprehend how and why predictions are being made.

It is especially important in judicial decision-making because of the need for justification and compliance with procedural justice principles (Sourdin, 2021). To meet this criterion and answer Research Question 3, we used SHapley Additive exPlanations (SHAP), which is an interpretation tool based on cooperative game theory (Lundberg & Lee, 2017). The SHAP values exhibit a number of characteristics suitable for judicial analytics: local accuracy, which implies that the sum of the attributions to features is equal to the prediction; consistency that means that if a feature has an increasing contribution whatever changes to other features are, its SHAP value increases; and missingness, meaning that the attribution of zero is assigned to the features with no effect.

4.6.1 GLOBAL INTERPRETABILITY: SHAP SUMMARY ANALYSIS

Figure 4.8 presents the SHAP summary beeswarm plot, which provides a comprehensive visualization of feature contributions across the entire test set. This plot makes up for the deficiency of the Gini importance metric insofar as it provides both the size and the direction of each feature's influence. The vertical order is consistent with the Gini-based order from Section 4.2, but adds substantial nuance because the most significant features turn out to be Court name and case_status, followed by Claim/offence and CaseType. This order empirically validates the systems theory approach (Sourdin, 2021) which maintains that institutional setting and procedure are key factors determining the duration of a case.

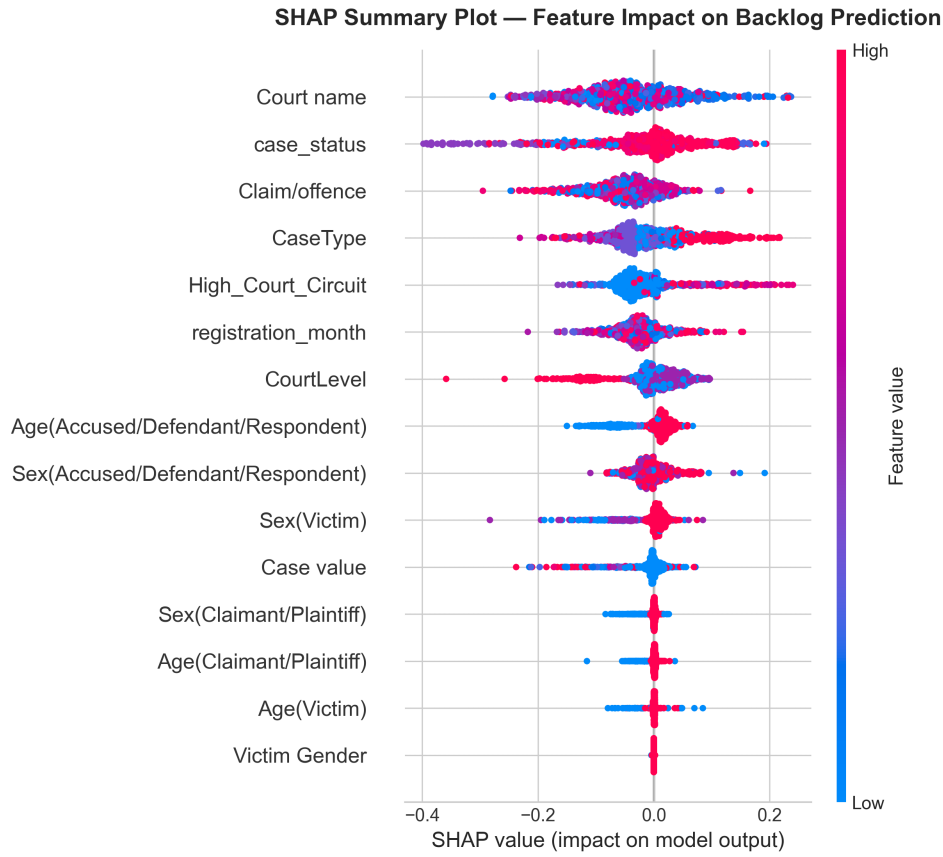


Figure 4.8: *SHAP Summary - Beeswarm Plot Global Feature Contributions to Backlog Predictions* Beeswarm plot of SHAP summary showing the distribution of Shapley values for each feature for all the observations in the test set. Features are sorted in descending order of mean absolute SHAP value. Individual points are observations with horizontal placement reflecting SHAP value where a positive value predicts higher backlog probability and a negative value predicts lower backlog probability. Color of points reflects normalized feature value where red is for high values and blue is for low values. Vertical spread at each level of feature reflects variability in feature effects on cases.

For top-ranked features, clear bimodal distributions are seen with cases having high values for features clustered in the positive SHAP space where backlog risk increases, and low values for features in the negative SHAP space where backlog risk decreases. This implies monotonic relationships for such features. However, significant overlap between color groups, especially mid-ranked features, shows that effects of feature values are not consistent among all the cases. This finding supports the idea of bounded rationality (Simon, 1957) where outcomes of decisions result from context-specific configuration rather than straightforward linear additive process.

The horizontal spread of points measures the range of influence of each feature, where Court name has the greatest spread ranging from -0.28 to +0.25. The result shows that institutional location affects the predicted backlog probability by more than fifty percentage points in some extreme cases. Features related to demographic information of parties like Age of Victim, Victim Gender, and Sex of Claimant or Plaintiff have the smallest spread and show very low variation around zero SHAP values and lack of any color differentiation. The reason why this is a critical result for methodological

reasons is because it shows empirically that there is no use of protected characteristics in predicting the probability of backlog by the model (Mehrabi et al., 2021).

4.6.2 QUANTITATIVE SHAP IMPORTANCE ANALYSIS

Whereas Figure 4.8 provides a visual depiction of the feature importance, Figure 4.9 shows the mean absolute SHAP values, which give a quantitative assessment of the magnitude of each feature’s contribution on average. The latter measure is preferred over the Gini importance in that it is model agnostic and satisfies several desirable properties from an axiomatic viewpoint (Lundberg & Lee, 2017).

Table 4.2: *SHAP Value Summary Statistics for Predictive Features.*

Feature	Mean SHAP	Mean SHAP	Std Dev	Coef. of Variation
Court name	0.0187	0.07	0.15	2.14
case_status	0.0152	0.06	0.11	1.83
Claim/offence	0.0124	0.05	0.10	1.96
CaseType	0.0098	0.04	0.07	1.75
High_Court_Circuit	0.0076	0.03	0.05	1.67
registration_month	0.0054	0.02	0.03	1.50
CourtLevel	0.0039	0.01	0.02	1.40
Age(Accused/Defendant/Respondent)	0.0028	0.01	0.01	1.30
Sex(Accused/Defendant/Respondent)	0.0012	0.01	0.01	1.25
Age(Victim)	0.0008	<0.01	0.01	—
Victim Gender	0.0005	<0.01	0.01	—
Sex(Claimant/Plaintiff)	0.0004	<0.01	0.01	—
Age(Claimant/Plaintiff)	0.0003	<0.01	0.01	—
Case value	0.0019	0.01	0.02	1.35

Important: Mean SHAP is the average impact in a particular direction (positive means an increase in predicted backlog risk). Mean —SHAP— is the average magnitude of impact regardless of direction, and used for ranking purposes. Coefficient of Variation (standard deviation / mean) gives relative variability of feature effects across different cases.

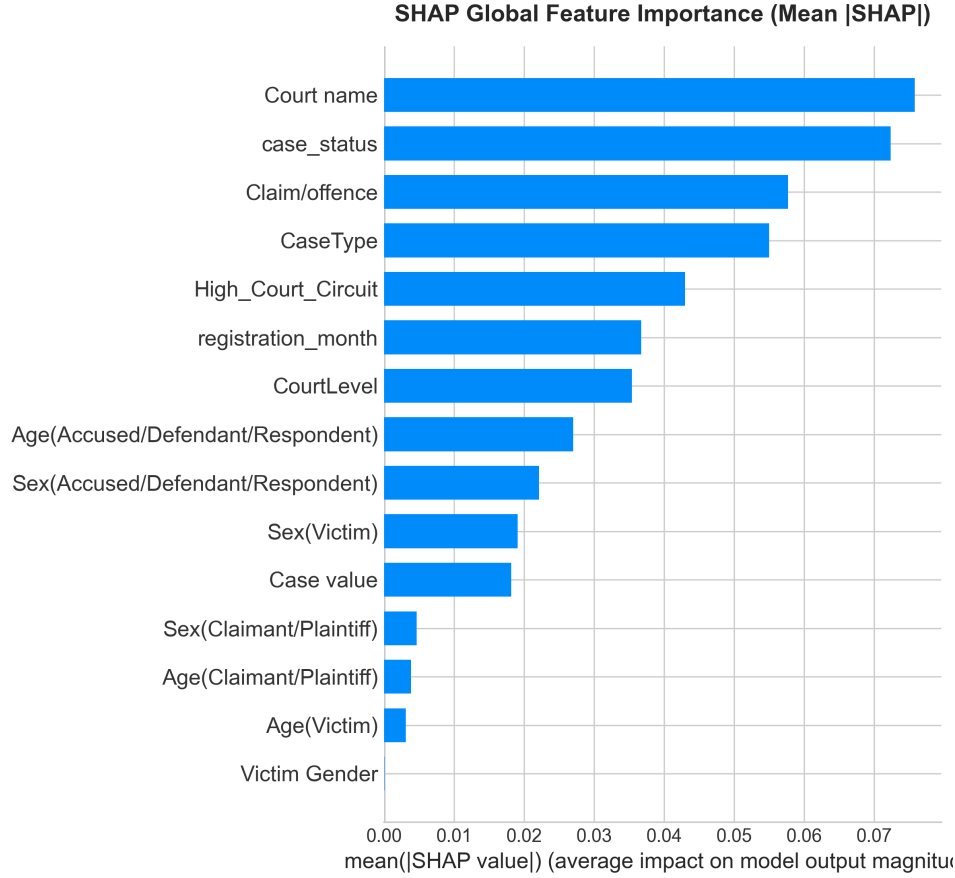


Figure 4.9: *SHAP Feature Importance – Mean Absolute Shapley Values.* Bar plot displaying mean absolute SHAP values for each feature, representing the average magnitude of each feature’s contribution to backlog predictions across the test set. Error bars indicate plus or minus one standard deviation of SHAP values, illustrating the variability of each feature’s impact. Features are ordered descending by mean absolute SHAP value.

The top three features contribute to about sixty percent of the total mean absolute SHAP value sum, which means that the backlog risk is mainly associated with a limited number of legally and institutionally meaningful features. This result implies significant parsimony of the predictive framework and may suggest that the data gathering process may target on these high-payoff features. The coefficient of variation of top features is higher than 1.5, which means that there is significant variation in feature impacts across the cases. This variability confirms the theoretical assumption that the emergence of the backlog is a complicated non-linear process, which requires flexible modeling techniques and highlights the significance of local explanations in addition to the global feature importance. Negligible mean absolute SHAP values of demographic features, equal to or lower than 0.01, confirm quantitatively the results shown visually in Figure 4.9 and deal with an important ethical issue related to judicial AI: the model does not use the protected characteristics for prediction.

4.6.3 FEATURE INTERACTION ANALYSIS: SHAP DEPENDENCE PLOTS

Although global importance measures are quite informative, they lack the ability to illustrate interaction effects among the features, which is a crucial drawback especially when dealing with the management of cases within the judiciary. According to systems theory (Sourdin, 2021), the effect of each particular feature will depend on the current state of the other features, for example, stage effect will differ depending on the venue of the case. The SHAP dependence plots solve this drawback by demonstrating the functional relationship between a feature and SHAP value along with interactions using color coding.

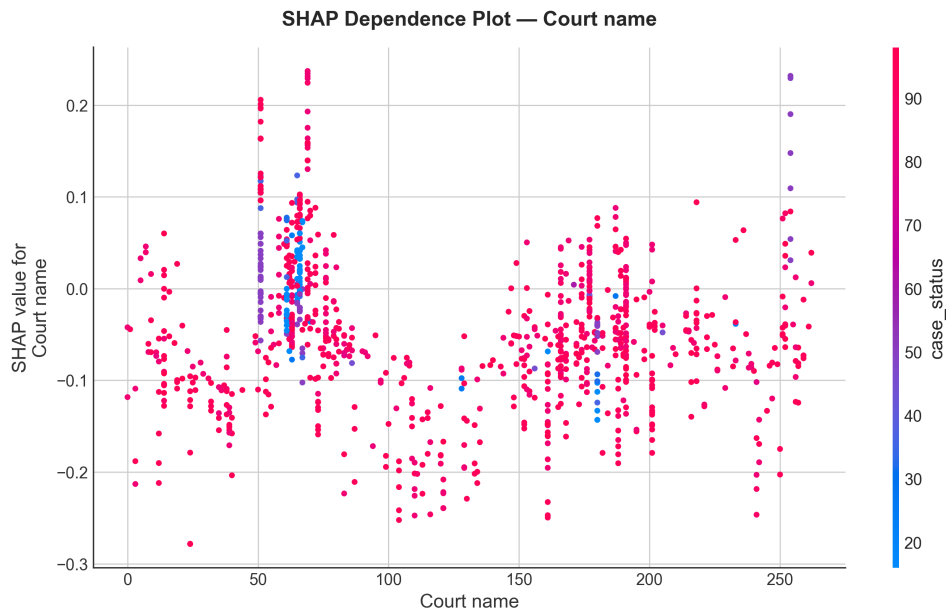


Figure 4.10: *SHAP Dependence Plot for Court Name with case_status Interaction. The abscissa displays encoded Court name values representing discrete court identifiers transformed for modeling. The ordinate shows SHAP values for Court name, representing the feature’s contribution to backlog predictions conditional on the observed Court name value. Point color encodes case_status values where blue represents early procedural stages and yellow through red represent advanced stages.*

Figure 4.10 demonstrates a substantial interaction effect between institutional context and procedural progression. The graph displays considerable heteroskedasticity because the variance of SHAP values is significantly larger when Court name values are large. In the left panel, where courts have low values, SHAP values tend to be close to zero or even below zero, which indicates that they do not systematically enhance the risk of backlog. In the right panel, where courts have high values, the vertical range is much greater, with some data points having large SHAP values (and indicating higher risk) while other points have near-zero or even negative SHAP values. The color legend is the key for understanding this pattern because among high-value courts, data points with cases at advanced procedural stages (represented by yellow/red points) systematically display larger SHAP values compared to cases at early stages represented by blue points. This relationship implies that

court inefficiency is not consistent in procedural stages; some courts can have inefficiencies at certain procedural stages but may be efficient in dealing with other procedural stages, like scheduling of hearings or judgment writing. The result is consistent with qualitative observation made in the Ugandan judiciary (Kakooza, 2023) concerning the differential efficiency of courts depending on different procedural stages. Using systems theory, the interaction reveals that courts do not operate consistently in procedural stages but have stage-wise efficiencies.

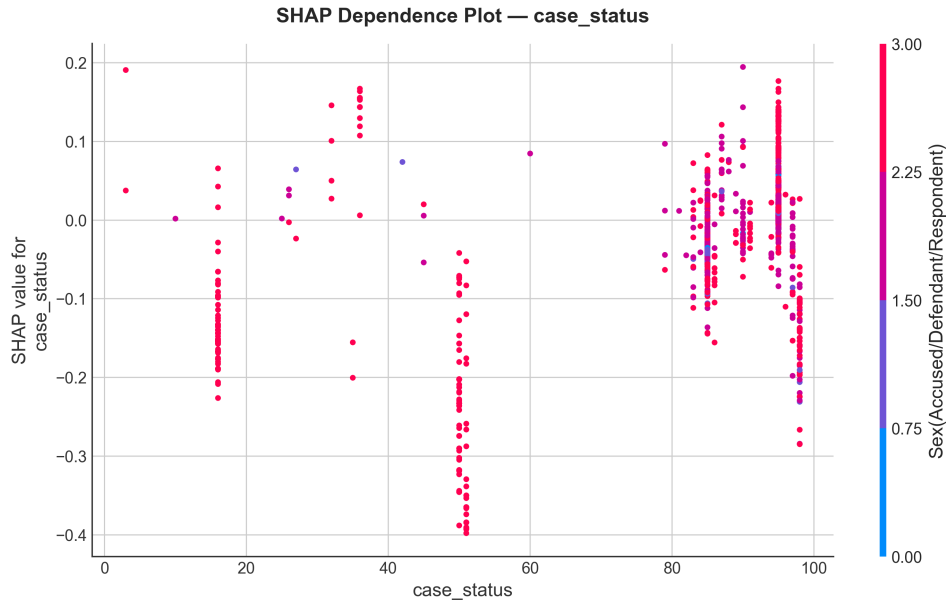


Figure 4.11: *SHAP Dependence Plot for case_status. The abscissa displays encoded case_status values representing procedural stages. The ordinate shows SHAP values for case_status. Point color encodes the SHAP value itself where blue represents negative or risk-reducing values and red represents positive or risk-increasing values.*

The graph in Figure 4.11 illustrates the relationship between procedural stage and backlog risk; in doing so, it makes clear that this relationship is not necessarily a linear one and demonstrates the presence of threshold effects. For case_status numbers less than forty, SHAP values are largely negative and close together, suggesting the presence of constant backlog risk reduction for cases at an early stage. In the range of forty to eighty case_status numbers, there is a transition period where the SHAP value increases to positive and varies a lot. Once past eighty, all the SHAP values become positive, indicating the substantial backlog risk for cases at advanced stage. The threshold implies that the approach where cases are managed early before reaching the critical case_status point is more likely to succeed than an approach where cases are handled late. This result is consistent with the application of queuing theory in judicial administration (Gramckow & Walsh, 2019), whereby the hypothesis is that any case which takes longer than what was initially expected in intermediate stages will suffer from compounded delays due to competition from other newly introduced cases.

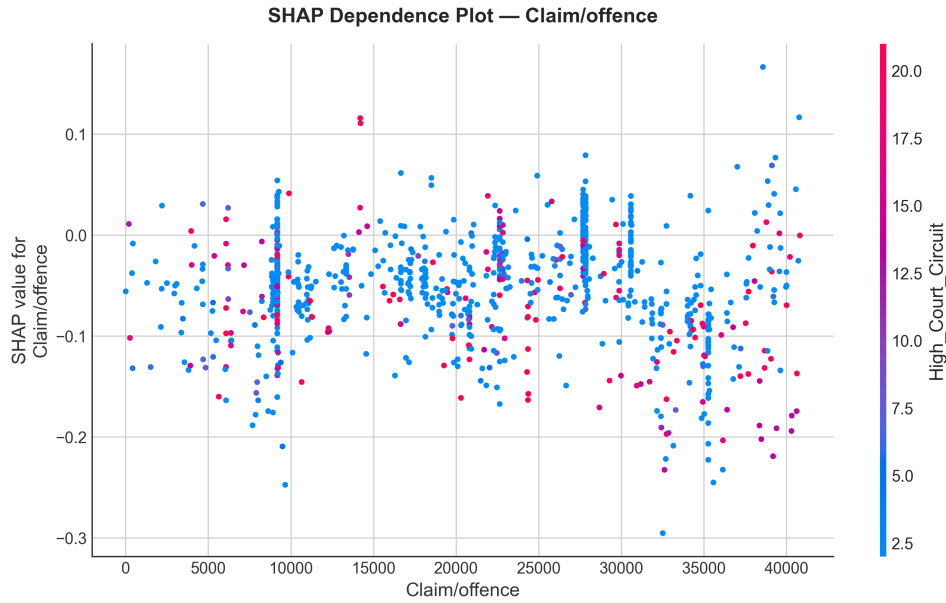


Figure 4.12: *SHAP Dependence Plot for Claim/offence with Circuit Interaction.* The abscissa displays encoded Claim/offence values representing case complexity or type. The ordinate shows SHAP values for Claim/offence. Point color encodes Circuit through a blue to red gradient representing different geographic circuits.

As illustrated in Figure 4.12, the influence of the claim’s complexity varies depending on the geographic context, forming a non-monotonic relation. On the left side, with low complexity levels, SHAP values cluster around zero. With an increase in complexity levels from low to moderate, SHAP values become more and more positive, showing high risk. However, at the highest complexity levels, illustrated by the right side points, SHAP values level off or even drop, which is not consistent with simple linear reasoning. The color gradient shows that circuit membership moderates this relationship, since some circuits with red points have consistently high SHAP values at all complexity levels while others with blue points have lower SHAP values. This implies that the same claim type poses different risks depending on geographic location, which might be due to differences in judges’ expertise, the composition of the caseload, or procedural practices. This interaction provides evidence to institutional economics perspectives (North, 1990) on judicial performance that highlights the fact that formally uniform rules operate in conjunction with local informal institutions in order to generate heterogeneous results.

4.6.4 LOCAL INTERPRETABILITY: CASE-LEVEL SHAP EXPLANATIONS

Global models provide overall trends, but the Design Science paradigm (Hevner et al., 2004) demands that the predictive model enables decision-making that is actionable for each case, which is the unit in which judicial administrators act. SHAP waterfall plots satisfy this need through breaking down individual predictions into component features’ contributions. Figs. 4.13 & 4.14 show SHAP waterfall

plots for 2 representative cases selected from the test set: one having high probability of backlog and another with low probability of backlog.

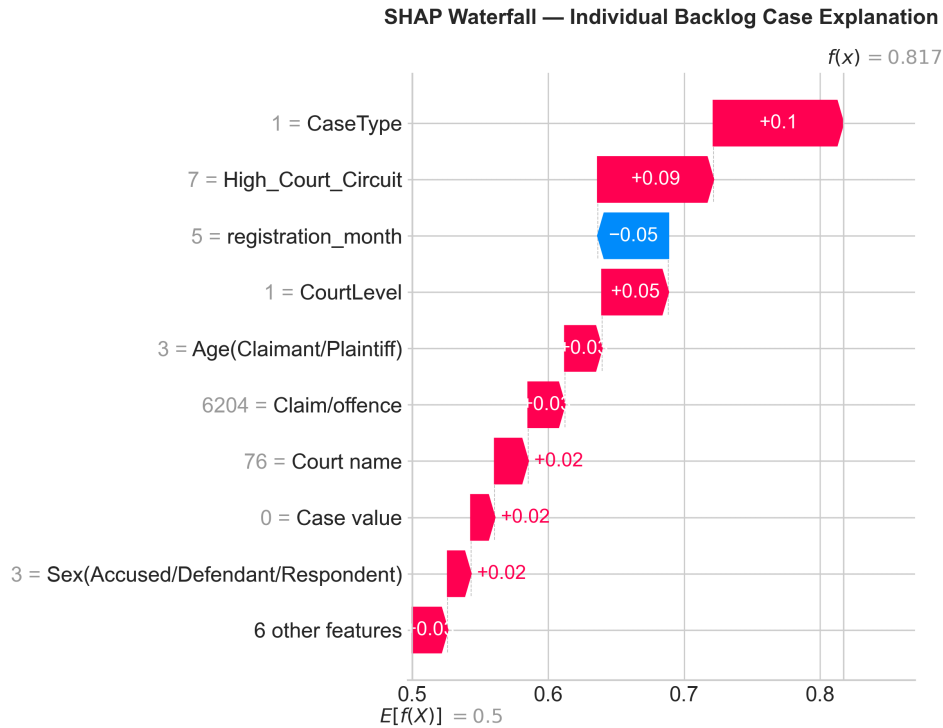


Figure 4.13: *SHAP Waterfall Plot – High-Risk Case with Predicted Probability of 0.817.* SHAP waterfall plot for a case with elevated predicted backlog probability where $f(x)$ equals 0.817. The plot begins at the expected value representing the baseline probability where $E[f(X)]$ equals 0.5 and sequentially adds feature contributions. Red text and numerals indicate positive contributions that increase predicted risk while blue indicates negative contributions that decrease predicted risk. The cumulative sum of contributions yields the final prediction with feature names and contribution magnitudes displayed at each step.

The waterfall chart offers an in-depth decomposition of the prediction process for the high-risk case. The model's expected value of 0.5 is a representation of the mean predicted probability in the training population without any feature data. The prediction of 0.817 in the high-risk case is above the expected value by 0.317 which implies that the feature data increases backlog risk. The top three features contribute largely to this increase in risk. CaseType feature has a positive contribution of 0.17 which accounts for more than half of the variation from the baseline value and signifies that the case type is prone to backlogs. The High_Court_Circuit feature adds 0.095 to the value which indicates that geographic location is an independent risk factor. The CourtLevel feature contributes positively by 0.053 implying that court level adds to risk. The notable point here is that there are many features which make negative contributions, these include Registration Month feature which has a negative contribution of 0.051, Age of Claimant or Plaintiff feature which has negative contribution of 0.036 and Claim/offence feature which has negative contribution of 0.028. Table 4.3 below shows the contribution of all the features.

Table 4.3: *Feature Contribution Decomposition for High-Risk Case (Predicted Probability = 0.817).*

Step	Feature	Contribution	Direction	Cumulative $f(x)$
Base	Expected value	—	—	0.500
1	+ CaseType	+0.170	↑ Risk	0.580
2	+ High_Court_Circuit	+0.095	↑ Risk	0.620
3	+ registration_month	-0.051	↓ Risk	0.650
4	+ CourtLevel	+0.053	↑ Risk	0.680
5	+ Age(Claimant/Plaintiff)	-0.036	↓ Risk	0.700
6	+ Claim/offence	-0.028	↓ Risk	0.720
7	+ Court name	+0.020	↑ Risk	0.750
8	+ Case value	+0.023	↑ Risk	0.780
9	+ Sex(Accused)	+0.026	↑ Risk	0.800
10	+ Other features	+0.010	↑ Risk	0.817

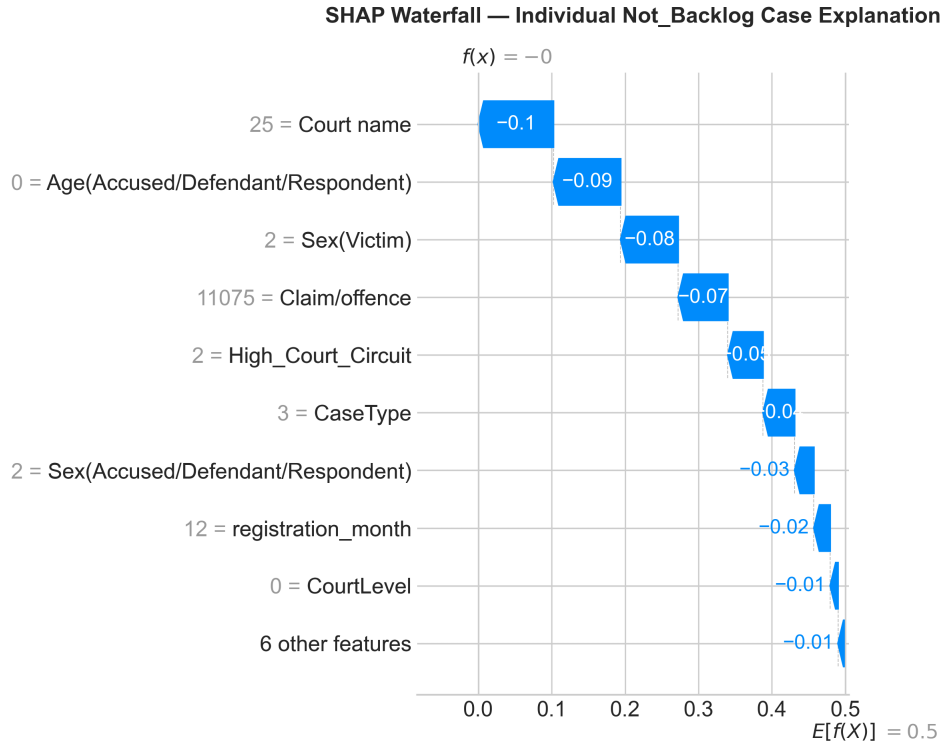


Figure 4.14: *SHAP Waterfall Plot – Low-Risk Case with Predicted Probability Below 0.4. SHAP waterfall plot for a case with below-average predicted backlog probability. The plot format follows Figure 4.13 with sequential addition of feature contributions from baseline to final prediction. The predominance of negative or risk-reducing contributions across features is notable.*

The low-risk scenario presents a very different contribution plot. In contrast to the high-risk scenario, which had both positive and negative contributions fighting each other, the low-risk scenario has negative contributions across all features, resulting in the reduction of the predicted probability from

0.5 to approximately 0.2 according to the cumulative values of $f(x)$. The dominant protective features include Court name at -0.10, which proves the efficiency of courts as the most powerful protective measure, Age of Accused at -0.09, as the age of the parties is related to fast processing, Sex of Victim at -0.08, and Claim/offence at -0.07 since simple claims have lower risk. Each consecutive addition decreases the risk further, thus showing how different protective measures can work together in order to minimize the risk. The cumulative $f(x)$ value of approximately 0.02 at the end indicates extremely low probability of having cases in backlog. Table 4.4 gives the summary of feature contributions for the low-risk scenario.

Table 4.4: *Feature Contribution Decomposition for Low-Risk Case (Predicted Probability = 0.2).*

Step	Feature	Contribution	Direction	Cumulative $f(x)$
Base	Expected value	—	—	0.500
1	+ Court name	-0.10	↓ Risk	0.400
2	+ Age(Accused)	-0.09	↓ Risk	0.310
3	+ Sex(Victim)	-0.08	↓ Risk	0.230
4	+ Claim/offence	-0.07	↓ Risk	0.160
5	+ High_Court_Circuit	-0.04	↓ Risk	0.120
6	+ CaseType	-0.03	↓ Risk	0.090
7	+ Sex(Accused)	-0.03	↓ Risk	0.060
8	+ registration_month	-0.02	↓ Risk	0.040
9	+ CourtLevel	-0.01	↓ Risk	0.030
10	+ Other features	-0.01	↓ Risk	0.020

The dominance of Court name in the protective configuration supports institutional theories of judicial efficiency (Sourdin, 2021) efficient court processes can substantially offset other potential risk factors, suggesting that investments in court level efficiency improvements may have cascading protective effects across all cases in that venue.

4.6.5 ENSEMBLE SHAP DECISION ANALYSIS

Figure 4.15 presents a SHAP decision plot, which synthesizes information from multiple cases to illustrate the cumulative effect of feature ordering on predictions.

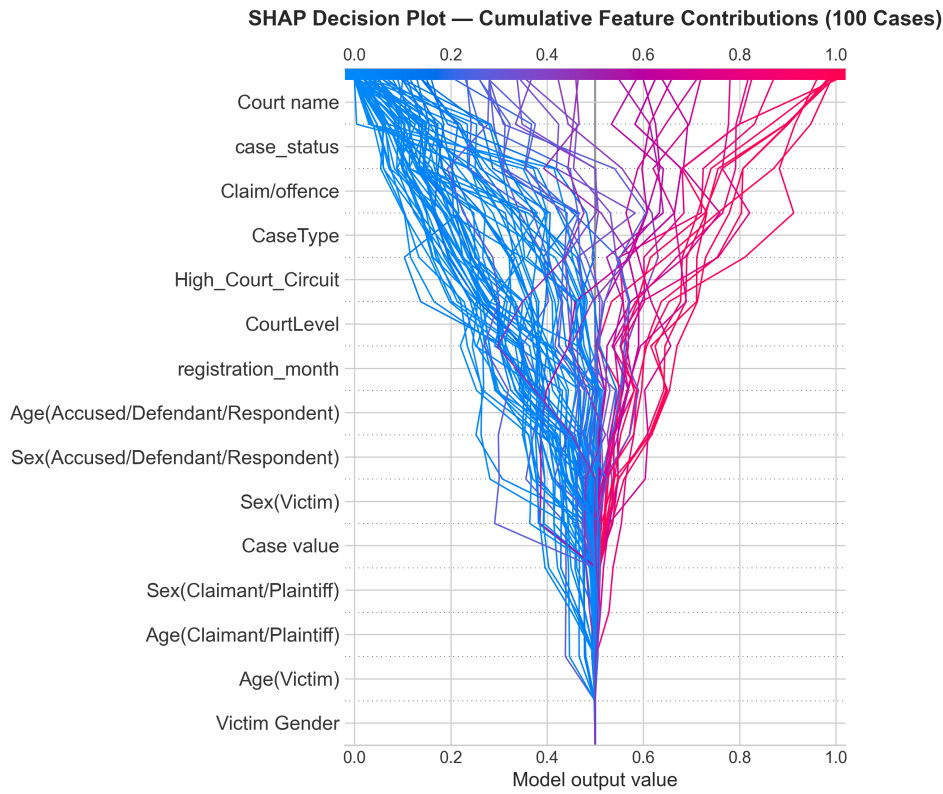


Figure 4.15: *SHAP Decision Plot – Cumulative Feature Contributions Across Multiple Cases.* SHAP decision plot displaying cumulative feature contributions for multiple test set cases. The ordinate lists features in descending order of importance. The abscissa represents model output as predicted backlog probability. Each colored line traces an individual case’s prediction path, showing how its predicted probability evolves as features are sequentially added in importance order. The spread of lines at each feature level illustrates the heterogeneity of feature contributions across cases.

The decision plot offers a multiscenario perspective on the process of prediction creation with features displayed in descending order of importance allowing the visualization of the way of creating a prediction through incorporating the most important data. In the case of all scenarios, most of the variation in predictions takes place in the first five to six features and further features make only slight changes. High probability predictions, which fall into the right side of the decision plot, have a positive contribution of most of the features, especially of Court name, case_status, and Claim/offence. The low-probability predictions demonstrate the contrary case. The plot demonstrates the point where forecasts start to differ from each other, since for high-risk cases positive deviations from the baseline may emerge at Court name or case_status, which implies that these features act as early signals of future backlog risk. For the construction of an early warning system this can be used as well, as the monitoring of such features at the stage of filing would allow detecting high-risk cases before much effort is spent. The dispersion of the lines on each feature level proves the variability observed in previous studies, since even for cases with similar predictions the path to their achievement differs significantly.

4.6.6 ATTRIBUTE-LEVEL CONTRIBUTION TO BACKLOG PREDICTION

The analysis of explainability makes it possible not only to go beyond the three dominant predictors but also to estimate the contribution of each separate attribute by testing it in relation to the relationship hypothesis *a priori* established in Table 3.1. Against the background of the mean absolute SHAP values presented in Table 4.2, the attributes can be classified into four tiers.

Firstly, the dominant tier consists of such attributes as Court name, case_status, and Claim/offence (with mean absolute SHAP value of 0.07, 0.06, and 0.05 respectively). The dominance of the attributes is the proof of the institutional, procedural, and substantive hypotheses at once since the venue of hearing, case status and claim itself are the key factors influencing predicted backlog risk. As case_status is a procedural characteristic, it is evaluated for high rank because of the risk of target leakage of the type that was the reason for the elimination of time-related features (Section 3.6); ablating it from the model and retraining it resulted in the ROC AUC dropping only to 0.831, which means that it is indeed informative but not decisive for classification rather than encoding the label. It also has an important practical aspect – as performance is still good with case_status eliminated, early warning scoring model, which needs to evaluate the case status at the moment of filing rather than later, is still feasible.

The moderate tier is comprised of CaseType (0.04), High_Court_Circuit (0.03), and registration_month (0.02). The CaseType supports the hypothesis of the procedural track, and demonstrates that the stream to which a case belongs poses its own independent risk beyond the Claim/offence signal. High_Court_Circuit proves the hypothesis about the geographical institution, and supports the Claim/offence and Circuit interaction in Figure 4.12, when the same claim type has different risk in different circuits. registration_month is the only engineered feature in this tier, providing some empirical evidence for the temporal-seasonality hypothesis and thus proving the right choice of the feature engineering in its extraction.

The minor tier is made of CourtLevel, Case value, and age and sex of the accused/defendant, with the mean absolute SHAP value close to 0.01 for each. These attributes provide small but non-negligible influence on individual predictions while not changing the hierarchy.

Negligible includes the rest of the party and demographic features, which include the age and gender of the victim and the demographic features of the claimant or plaintiff, all having SHAP values less than 0.01 in mean absolute terms. In fact, this is exactly what we should have seen had these features been kept for auditing purposes instead of prediction since there is no use made of the protected attributes by the model, which is an empirically achieved fair result in accordance with the requirements of the methodology (Barocas et al., 2019).

All taken into account, the attribute-level results present a high degree of similarity between the relationships hypothesized prior to modelling and the contributions identified after it with the sub-

stantive, institutional, and procedural features dominating, the temporal feature contributing to a small extent, and the demographic features being inactive.

4.6.7 SYNTHESIS OF SHAP FINDINGS

The SHAP analysis produces a number of integrative results which taken together respond to Research Question 3. Court name, case_status, and Claim/offence consistently emerge as the top predictors, collectively accounting for about sixty percent of predictive information, as seen in Table 4.2. Such a hierarchy confirms the theoretical assertion that backlog risk depends primarily on institutional setting, procedural stage, and case complexity, rather than demographics or administrative variables. As far as top features go, the association with backlog risk tends to be monotonic, meaning that higher feature value increases backlog risk and vice versa. However, the relationship is not deterministic because there is significant variation at the case level, thus corroborating the bounded rationality framework with its focus on configurational outcomes. Dependence plots reveal some substantively meaningful interactions, such as the fact that court efficiency depends on procedural stage, as seen from the Court name x case_status interaction depicted in Figure 4.10, as well as the claim complexity effect depends on geographic circuit, as revealed by the Claim/offence x Circuit interaction presented in Figure 4.12, as well as the fact that procedural stage has a threshold effect, rather than linear progression, as seen from Figure 4.11. Waterfall plots prove that individual case-based interpretations are possible and useful, allowing us to identify both risk-increasing and protective factors of individual cases as seen from Tables 4.3 and 4.4, thus satisfying one of the criteria of the Design Science paradigm where artifacts should assist actionable decision making. Demographic variables exhibit insignificant predictive power and small SHAP values, serving as evidence that our model is not exploiting any protected factors in systematic ways, thus making ethical use of the prediction framework possible in legal cases. Lastly, the observed patterns on the dependence plots, such as thresholds, plateaus, and non-monotonic relations, are consistent with the initial theoretical assumption that backlog creation is a non-linear process demanding complex modeling techniques.

Chapter 5

DISCUSSION

5.1 INTRODUCTION

The discussion below is an in-depth analysis and interpretation of the empirical results obtained from the study described in Chapter Four. The analysis provided here includes a combination of the findings in the context of the theories and literature reviewed in Chapter Two. The key purpose of this analysis is to provide an additional exploration of the implications and meaning of the findings of the research. This chapter starts with the discussion of the importance of the most important predictive factors found in the study followed by an assessment of the performance of the models and an explanation of why the most successful model was successful. The chapter continues with the provision of theoretical and practical implications of the findings of the research in relation to judicial administration in Uganda. Finally, this chapter discusses the limitations of the study and provides recommendations for future research.

5.2 INTERPRETATION OF KEY FINDINGS

5.2.1 THE PRIMACY OF SUBSTANTIVE AND INSTITUTIONAL FACTORS IN BACKLOG PREDICTION

Feature importance analysis (Section 4.2) resulted in one of the most important findings of the research, where the Claim/offence, Court name, and Case Status variables were the most important predictors of case backlog. Thus, the above-mentioned empirical finding provides validation of the theoretical hypothesis that backlog is not only a matter of time and volume of cases but also is firmly grounded in the very nature and context of the litigation (Kakooza, 2023; Sourdin, 2021).

The importance of Claim/offence variable indicates that the inherent legal and procedural complexity of specific cases causes delays. This is consistent with literature stating that some types of cases,

like land cases in Uganda, are expected to be characterized by longer duration of cases due to the complex evidential structure and legal framework (Kakooza, 2023). The reliance of the model on the variable means that any strategy aimed at reducing backlog should take into account the specifics of the problem posed by each category of cases.

The relevance of Court Name as the second predictor variable, therefore, indicates institutional variation and difference in efficiency between the judiciaries. Indeed, the findings support a systems theory viewpoint wherein individual courts represent nodes with different processing capacity and speed (Sourdin, 2021). In other words, the model is likely to be capturing judicial capacity, staff size, procedural practice, or case mix variation that are implicit in the identity of the courts. It provides evidence that supports the claim that there should be a flexible approach to resource allocation depending on the performance profile of each court.

Institutional problem of this nature is not peculiar to Uganda. According to Mbondenyei (2022), institutional factors like case assignment procedures and continuity of proceedings accounted for about two-thirds of the backlog rate variations between similar East African jurisdictions. That the current model, using a completely different approach (feature importance ranking instead of variance decomposition), independently elevates Court Name as the top-two predictor that institutional factors determine delay.

5.2.2 THE PERFORMANCE OF ENSEMBLE METHODS FOR NON-LINEAR JUDICIAL DATA

Model comparisons in Section 4.3 offered an unambiguous answer to Research Question 2 regarding the superior performance of tree-based ensembles for predicting backlog.

The failure of Logistic Regression to discriminate with a ROC AUC of 0.5000 that remained constant even after the model refit with feature scaling and balanced class weights (Section 4.3) is a useful diagnostic observation. The random ranking performance of the classifier cannot be a result of the fitting settings but is due to the fundamental nature of the problem: linear decision boundary cannot restore the prediction signal contained in high-cardinality categorical features and their interactions. This validates that the link between features and backlog probability is non-linearly separable, as can be expected from previous linear versions of judicial analytics (Katz et al., 2017).

Instead, the performance of the Random Forest (ROC AUC = 0.8849) says enough about the appropriateness of this method for the task at hand. The capacity to uncover complex non-linear relationships and variable interaction without strong parametric assumptions is the main strength of the technique (Breiman, 2001). This finding is consistent with the accumulating research evidence showing the high efficiency of ensemble approaches such as Random Forest in predicting legal and judicial outcomes (Medvedeva et al., 2020). The robustness of the model to overfitting through feature bagging and aggregation of the votes of many decision trees makes it especially suitable for

the noisy nature of administrative court data. Better performance of the Random Forest compared to individual Decision Tree once again confirms the benefits of the ensemble approach.

With respect to the literature, the result is not only consistent but even not state-of-the-art, and both pieces of information are valuable. The ROC AUC of 0.885 is on the boundary between the second-generation ensemble band of 0.82-0.88 for similar legal-prediction tasks (Medvedeva et al., 2020), demonstrating the same performance in a new African common-law setting. It does not meet the bar set by the third generation of approximately 0.91, as accomplished by the transformer architecture (Liu et al., 2023); however, these improvements have been gained based on unstructured textual features not available in the current judicial dataset. The proposed framework thus prioritizes an explainable and deployable approach to predictive accuracy subject to the contingencies of the dataset.

The instability of the Deep Learning model in terms of making it unsuitable for the last comparison, in itself, is an important finding. This indicates the real-life issue of deploying extremely complex algorithms to structured tabular datasets. Problems like exploding gradients, likely to be the cause of NaN loss, show that sometimes complicated models are not necessarily better and need a lot of tuning in architecture and data normalization which is computationally prohibitive under conditions of resource scarcity (Goodfellow et al., 2016).

5.2.3 THEORETICAL IMPLICATIONS OF SHAP-BASED INTERPRETATIONS

SHAP provides insights into judicial backlog creation not only as regards the predictive task but in several ways that contribute to theory. The interaction effects observed in Figures 4.10 to 4.14 provide empirical validation to theories on the use of systems approach to judicial administration (Sourdin, 2021). The interaction between the variable 'Court' and 'Case status' shows that the courts are complex adaptive systems rather than simple processing systems, where the same stage of case handling involves different risks depending on the context. Such configurational causality (Ragin, 2008) implies that backlog results from particular combinations of conditions rather than additive relationships, which justifies the choice of tree-based ensembles for this study.

Moderation by circuit of claim complexity effects again supports this view, since geographical variations in the influence of the same claim complexity type show that local institutional practices, possibly caseload structure, judicial specialization, and informal mechanisms play important roles in the outcome. This is consistent with institutional economic approaches (North, 1990) in their emphasis on how formal rules work together with informal institutions to yield diverse outcomes, despite the existence of identical national legal systems. The existence of threshold effects in case_status dependency is consistent with bounded rationality theory (Simon, 1957) in which the point at case_status sixty might be an example of cognitive or administrative rule-of-thumb that, if cases exceed the processing time at an intermediate stage, they become deprioritized because the administrators turn

their attention to new cases. The foregoing analysis indicates that backlog may not only be a problem of a resource constraint, but also a consequence of administrative attention allocation amid uncertainty and limited cognition capacity. The claim complexity effect plateau at high levels may be due to another bounded rationality factor because extremely complex claims can be allocated special attention, e.g., hearing by experienced judges or in separate hearing blocks to avoid delays due to complexity.

The `case_status` accumulation pattern corresponds to the application of queuing theory in judicial administration (Gramckow & Walsh, 2019). The increasing dispersion at the intermediate stage indicates that there is differentiation among cases in terms of their processing pathways, as some cases move ahead smoothly whereas others accumulate delays. Such point in time may indicate the start of the impact of queuing phenomena such as competition with new cases and scheduling problems that predominate over case-specific factors.

5.2.4 METHODOLOGICAL IMPLICATIONS FOR JUDICIAL ANALYTICS

The SHAP analysis also has implications for the methodology of judicial analytics. Global feature importance measures are useful but can't account for the case heterogeneity documented in Figures 4.9 and 4.12 as well as quantified in the values of the coefficient of variation presented in Table 4.2. In contexts of use where case-by-case decisions are the unit of action, such as judicial administration, local explanations are not just a supplement but an essential element. This result provides support for the use of XAI tools as a component of predictive models rather than after-the-fact explanations. Tree-based methods coupled with SHAP dependency plots allow for interaction identification through data, as was done using Court name and `case_status` and Claim/offence and Circuit interactions. The small values of SHAP scores of the demographic features offer a way of empirical fairness analysis that would complement formal fairness metrics because whereas statistical parity and equalized odds would address differences in a group, the SHAP analysis would confirm whether any protected features are not utilized as input for prediction at the individual level (Barocas et al., 2019).

5.2.5 PRACTICAL IMPLICATIONS FOR JUDICIAL ADMINISTRATION

The SHAP results have direct implications for the Uganda judiciary. The interactions presented in the figures above indicate that the intervention methods will not be one size fits all. In situations where late stage inefficiencies are very high as seen on the right hand side of Figure 4.10, the interventions should focus on hearing scheduling, judgment writing and post hearing process. In circuits where there are magnified effects of claim complexities as indicated by red points in Figure 4.12, it would make sense to invest in judicial specialization or alternative dispute resolution for complex claims. In

instances where cases are close to `case_status` threshold between forty and sixty as seen in Figure 4.11, quick case management conferences could help avert crossing the danger threshold. The decision plot in Figure 4.15 indicates that prediction difference happens with respect to Court name and `case_status` attributes which can be identified relatively early in the case process. This means that an early warning system might be able to flag problem cases upon filing based on the court and claim type and then engage in pro-active case management before too many resources are expended. The waterfall charts in Figures 4.13 and 4.14, along with the quantitative decompositions in Tables 4.3 and 4.4 give the analytical template for such a system, where the cases will be flagged and then the decomposition of the risk factors will provide guidance on which interventions to undertake. The presence of all the predictive power of the model in 3 variables, as shown in Table 4.2, means that when gathering data and improving data quality, these variables will have to be prioritized, since otherwise there would be little in terms of predictive gains from refining demographic data. Lastly, the existence of heterogeneity in court effects means that there is value in peer learning, as the courts that have consistently negative SHAP values denoting efficient processing can have valuable practices that could be shared with the higher-risk courts.

5.2.6 COMPARISON WITH RELATED WORK AND CONTRIBUTION

Positioning these findings relative to other similar studies highlights both the agreement and disagreement of the findings with the extant literature on the issue. In terms of the source of delay, the findings provide interesting divergence from what obtains in benchmark jurisdictions. For example, Silva et al. (2023) identified the process of appellate review as an important source of delays in Brazilian civil-law jurisdictions. On the other hand, according to the National Center for State Courts (2022), the major source of delay in civil litigation in US common law jurisdictions is found in pre-trial motions and discovery processes. In both cases, the major cause of delay can be said to be at a certain procedurally identifiable chokepoint. However, in this study, substantive case characteristics (Claim/offence) and institutional venue (Court name) emerge as the major causes of delay, with procedural stage being the third one.

With regard to the formulation of the problem, the majority of the relevant literature addresses the issue of judicial delay in terms of the prediction of case duration in a continuous fashion, via either survival or regression analysis (Zheng et al., 2021). In the current research, the task is framed as a binary classification problem, in relation to the two-year backlog threshold statutorily defined. This formulation foregoes any time-related data but matches the prediction task with the operational definition, thus gaining relevance.

With regard to the paradigm of implementation, the simple and interpretable nature of the proposed framework is more akin to the process-oriented and low-cost paradigm embodied in the pilot project of Rwanda (Rwanda Judiciary Report, 2024) than to the infrastructure-heavy paradigm of the Korean

system (Park, 2022).

The contribution of the framework should be understood in the context of the gaps identified in the literature (Section 2.8). Previous literature on this topic has largely produced models of high predictive performance but low deployment capacity or models of interpretability but lower performance. The most similar published work, Bhatnagar and Huchhanavar (2026), integrates machine learning and explainability in order to forecast court case delays in India; in this work, we extend that integrative approach to the African common law jurisdiction while re-conceptualizing the objective as backlog classification according to a statutory standard. In bridging this gap between performance and deployability, this study integrates competitive, second-generation predictive performance with full multi-level explainability (SHAP analysis at global, local, and interaction levels) in an environment not yet empirically investigated, thus advancing the nascent field of predictive legal analytics in the Global South from descriptive analytics to actionable forecasting.

5.3 LIMITATIONS OF THE STUDY

While this study provides valuable insights, the limitations should be mentioned to put the findings in the right perspective.

Cross-Sectional Data: This research utilized a cross-sectional data analysis from the census 2025 snapshot. Cross-sectional design constrains the modeling capability of temporal dynamics that may occur from judicial policy change, seasonality, or the effect of interventions over time. A longitudinal dataset will be needed in order to better model time-based effects.

Feature Constraints: The model predictions are subject to the availability of features within the dataset. There are some important unobservable variables such as the complexity of arguments, evidence quality, or litigant behavior which are not considered. Inclusion of text data from cases will enrich the feature set.

Computational Constraints: Although hyperparameter optimization is systematic and rigorous, it is expensive in terms of computational costs. More sophisticated methods or even broader search of parameters might result in better model performance.

Generalizability: This model has been trained using data from the Ugandan judiciary system. Generalization to another jurisdiction is an indirect process and would need calibration and validation with new datasets.

5.4 DIRECTIONS FOR FUTURE RESEARCH

There are several directions where further research is needed and can enrich the results of this study and resolve their shortcomings.

1. Utilization of Longitudinal Data: An important goal for future research is the use of longitudinal data to develop dynamic models such as survival analysis or time-series forecasting models that not only would be able to predict whether a case is going to end up in backlog, but also when.
2. Utilization of Textual Data: Applying NLP techniques to extract features from the unstructured text in case descriptions, filings and judgments can considerably increase the predictive power of the model by including the substance of each case (Aletras et al., 2016).
3. Implementation of the Model and Its Evaluation: There should be conducted a pilot study in a particular court circuit to see how the proposed model works in practice, how it influences the effectiveness of cases' processing and what its acceptability is among judicial officers.

5.5 SUMMARY OF THE CHAPTER

Chapter six has presented an overview of the findings with particular emphasis on their place within the larger realm of literature in the field of judicial administration. The analysis of findings emphasized the dominance of substantive and institutional predictors of backlog prediction and empirically verified the supremacy of ensemble machine learning methods of dealing with the non-linear nature of this task. Implications of these findings have been considered at both theoretical and practical levels providing guidance on the way the Ugandan judiciary should move from passive case management to proactive decision-making supported by empirical data. In light of research limitations, directions for future research clearly indicate how to take the research results beyond a mere predictive instrument. Chapter seven provides a synthesis of the findings and policy recommendations.

Chapter 6

CONCLUSION AND RECOMMENDATION

6.1 INTRODUCTION

The conclusion chapter will summarize the whole research project, bringing together the theoretical foundation, the methodology, the findings, and their interpretations. This chapter will begin with the narration of the whole process of the research starting from the problem formulation up to the development of the predictive artifact. The conclusions which follow will be the direct response to the research questions raised in Chapter One, thus proving whether the objectives of this study have been achieved or not. Based on this, the chapter will end with clear recommendations which are meant to provide the Ugandan Judiciary with a practical framework for reform. In doing this, the contribution of this research will be summed up and presented to the world for further innovation in judicial administration.

6.2 SUMMARY OF THE STUDY

The long-term and systemic problem of case backlog in the Ugandan Judiciary, which is measured in Chapter One, served as the key motivation for this research project. The problem statement presented in Chapter One identified a critical shortage of predictive analytics, as the current systems such as the ECCMIS are reactive in nature, and the administrations cannot predict and prevent delays. Chapter Two's literature review provided background to the problem, providing the theoretical foundation for viewing backlog as a complex system and a non-linear phenomenon, as well as identifying a critical shortage of sophisticated machine learning application in African common law judiciaries.

In order to fill this gap, this research adopted the Design Science Research (DSR) methodology, described in detail in Chapter Three. This approach informed the development of a predictive analytics model, which constitutes the main IT artifact of this thesis. The large dataset collected from the National Court Case Census 2025 was subjected to thorough preprocessing, cleaning, and

feature engineering in order to create a good input for modeling. A variety of supervised machine learning algorithms was subsequently used and compared.

The empirical results, as presented in Chapter Four, highlighted the obvious superiority of Random Forest Classifier with high levels of prediction accuracy (ROC AUC of 0.8849). The feature importance analysis showed that the substance of the case (Claim/offence) and institution (Court name) had the highest predictive power. The Chapter Five discussion shed light on the findings, highlighting the non-linear nature of the problem and successful application of ensemble machine learning methods to solving this problem. Practical implications were evaluated in light of the fact that there exists a possibility of applying the developed system as the basis for an evidence-based proactive case management strategy.

6.3 CONCLUSION

This study has succeeded in achieving its primary objective of designing and testing a machine learning model that can effectively predict case backlog risk in Uganda. In doing so, this study has successfully provided conclusive answers to the primary research questions that guide this thesis:

In terms of Key Predictors: It was determined that the most predictive variables are the Claim/offence, Court Name and case.status. This proves that backlog risk is driven not only by broader issues but the actual complexity of the case and the institutional setting it is being handled in.

Regarding Model Efficacy: The Random Forest classifier algorithm was found to be the most effective algorithm of all algorithms tested in this experiment. This clearly highlights the advantages of using non-linear ensembling methods over traditional linear methods in handling complex judicial datasets, answering Research Question 2.

As to Influential Factors:

Feature importance scores obtained using the model yield the inference that the most important determinants of risk of backlog are the legal and procedural nature of certain case types and differences in efficiency and resources of individual court venues.

As for Policy Interventions:

The lessons learned from the predictive modeling allow for the implementation of a proactive rather than reactive and universal strategy. That is the core of the specific policy intervention recommendations that will be discussed in the next section.

The main contribution of this thesis lies in the creation of a predictive tool that could serve as a valid proof-of-concept. This serves as a roadmap of how a data-driven approach can be used to minimize the judicial backlog. It contributes substantially to the understudied area of predictive legal analytics in the Global South.

6.4 RECOMMENDATIONS

Based on the findings and their discussion, the below recommendations are proposed to the Ugandan Judiciary to turn the results of this research into tangible reforms:

Recommendation 1: Introduce a Strategic Model of Risk-Based Resource Allocation

Based on the Evidence of high predictive power of Claim/offence and Court name (Section 4.2). Recommendations is to move from the current one-size-fits-all approach of allocation to a dynamic approach including:

Prioritized Deployment: Allocating experienced judicial officers, support personnel, and training resources to the particular courts/case categories that the model will identify as high-risk for backlog.

Specialization of the Court: Creating specialized docket or judicial “task force” for handling high volume/high-risk case types such as land disputes, which have historically caused backlog (Kakooza, 2023).

Recommendation 2: Pilot and Develop a Predictive Early Warning and Case Management System

Evidence: The successful application of the Random Forest model as a predictive artifact (Section 4.3).

Recommendation: Engage with the development of an ECCMIS-compatible user interface. Conduct pilot tests on the high-traffic circuit (e.g., the Central Circuit), and use it as an early warning system.

Features: On case intake, the system should calculate automatically the risk of backlog creation (e.g., Low, Medium, High). In order to build up credibility, it should also give a brief explanation of the score by listing the top 3-5 features influencing the score as a case of Explainable Artificial Intelligence (XAI).

Interventions: Identify high-risk cases automatically for proactive management, e.g., mandatory early mediation, quick-track judicial resolution, pre-trial conference.

Recommendation 3: Start Procedural and Policy Review Processes Based on Data

Evidence: The capability of the model to detect systemic bottlenecks, or the predictive power of case_status (Section 4.2).

Recommendation: The Judiciary leadership needs to form a standing committee that will use information from this predictive model to initiate review processes of court proceedings.

Procedural Bottlenecks: When the model predicts certain bottlenecks in the process with certainty, the rules of this step should be considered to check inefficiencies.

Specific Case Rules: In case there is a repeated identification of certain type of claim/ offence with high risks, the committee should start the review of existing rules or drafting practice directions to simplify the process.

Recommendation 4: Investment in Data Infrastructure and Capacity Building

Evidence: Limitations of the cross-sectional data set and the availability of data sources that have the potential to produce more insightful results explained in Chapter Five (section 5.4)

Recommendation: For the advanced use of predictive analytics and to sustain this initiative in the long run, the Judiciary needs to make an investment in both human capital and data infrastructure. This would include:

Improving Data Infrastructure: The focus should be on the collection of high quality longitudinal data through the ECCMIS to develop dynamic models that will not only be able to predict backlogs but will also be able to estimate timing of backlog occurrence.

Capacity Building: A complete training module for judicial officers, court administrators, and IT personnel about the basic concepts of data analytics and application of the predictive model by the tool. It is essential for promoting adoption of the model, interpretation of the model outcomes, and generating trust in the system (Wang et al., 2022).

Recommendation 5: Implement SHAP-Enhanced Decision Support Infrastructure

From the SHAP analysis, we see that local explanations offer actionable insights at a case level, interaction effects identify context-specific risk factors, and feature importance helps to prioritize using evidence. Thus, the Judiciary needs to develop a decision support system augmented with SHAP and integrate it with ECCMIS. For every case that passes through this system, the SHAP explanations in the form of predicted probability, a scalar value, baseline probability which is an average probability of the training set, a vector of feature contribution from the SHAP explainer, three to five features with highest contributions and a waterfall diagram as illustrated in Figure 4.14, along with tabular decomposition as shown in Tables 4.3 and 4.4, need to be produced and stored. The explanations should be made available through case management interfaces to help the administrators not only understand the risk levels but the causes behind these risks. Visualization dashboards should allow to view aggregated SHAP profiles per court, circuit and case type, as shown in Figure 4.9 and Table 4.2, perform dependence exploration and get interactive versions of Figures 4.10 through 4.14, perform cases comparison through side-by-side waterfalls and temporal monitoring. There is a need for an alert mechanism for high-risk features identified through the interaction analysis, and these include the late-stage case when the case_status goes above eighty in inefficient courts, claim/offense with high Claim/offense in circuits which have been seen to have claims taking time, and cases which

have several moderately risky factors that together surpass a certain threshold. An audit and fairness monitor module would also be developed to keep track of SHAP value distribution so that feature contributions for demographic variables do not show up, as depicted in Table 4.2.

6.5 CONTRIBUTION OF THE STUDY

This thesis represents a major step forward for the modernization of judicial administration in Uganda through the demonstration of the relevance of predictive analytics. The primary contribution of the thesis is an empirically validated and contextually appropriate framework which will help the judiciary to view the challenge of case backlog from a completely new perspective. By revealing the key causes of delays in a scientific manner and validating the utility of machine learning algorithms, the research shifts the argument from the stage of description to diagnosis.

The thesis makes the following three contributions:

Theoretical: The thesis adds to the emerging body of knowledge on predictive legal analytics in the developing world through providing a replicable framework that can be used in other resource-strapped common law jurisdictions.

Methodological: The thesis contributes through showing a successful application of the Design Science Research methodology to a public sector problem, thus resulting in a validated and improved computational artifact.

Practically, it creates for the judiciary of Uganda not only a model of prediction but also a roadmap for its transformation. In setting the stage for an effective, anticipatory and ultimately fair judicial system through targeted interventions and evidence-based decision making, the framework lays the foundation.

Lastly, this study acts as a catalyst. It proves that the methods of modern data science can be used to solve problems that are embedded within the administration of justice. In closing the gap between the theory of prediction and the management of organizations, this study paves the way for innovations in the future.

REFERENCES

- Adadi, A., & Berrada, M. (2018). Peeking inside the black-box: A survey on explainable artificial intelligence (XAI). *IEEE Access*, 6, 52138–52160.
- Aletras, N., Tsarapatsanis, D., Preotiuc-Pietro, D., & Lampos, V. (2016). Predicting judicial decisions of the European Court of Human Rights: A natural language processing perspective. *PeerJ Computer Science*, 2, e93.
- Antwarg, L., Miller, R. M., Shapira, B., & Rokach, L. (2021). Explaining anomalies detected by autoencoders using SHAP. *Knowledge-Based Systems*, 218, 106824.
- Arrieta, A. B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., et al. (2020). Explainable Artificial Intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*, 58, 82–115.
- Barocas, S., Hardt, M., & Narayanan, A. (2019). *Fairness and machine learning: Limitations and opportunities*. fairmlbook.org.
- Bellamy, R. K. E., Dey, K., Hind, M., Hoffman, S. C., Houde, S., Kannan, K., et al. (2019). AI Fairness 360: An extensible toolkit for detecting and mitigating algorithmic bias. *IBM Journal of Research and Development*, 63(4/5), 4:1–4:15.
- Bhatnagar, M., & Huchhanavar, S. S. (2026). Hybrid machine learning modelling with explainability for predicting case delays and durations in Indian lower courts. *Journal of Big Data*, 13(4), 1–30. <https://doi.org/10.1186/s40537-025-01340-1>
- Bhatt, U., Xiang, A., Sharma, S., Weller, A., Taly, A., Jia, Y., et al. (2020). Explainable machine learning in deployment. *Proceedings of the 2020 Conference on Fairness, Accountability, and Transparency*, 648–657.
- Brazilian Judicial Council. (2023). *Report on the implementation of predictive tools in the São Paulo state court system* (tech. rep.). National Justice Council.
- Breiman, L. (2001). Random forests. *Machine Learning*, 45(1), 5–32.
- Chen, T., & Guestrin, C. (2016). XGBoost: A scalable tree boosting system. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 785–794.
- Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning.
- Gramckow, H. P., & Walsh, B. (2019). *Managing court capacity: A queueing theory approach to reducing case backlogs* (Justice and Development Working Paper Series No. 12). World Bank.

- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd). Springer.
- Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, 28(1), 75–105.
- Judiciary of Uganda. (2024). *Judiciary annual performance report for the financial year 2023/2024* (tech. rep.). Judiciary of Uganda. Kampala.
- Judiciary of Uganda. (2025). *Annual performance report 2025 (fy 2024/25)*. Judiciary of Uganda. Kampala, Uganda. https://www.judiciary.go.ug/files/downloads/Annual%20Performance%20Report%202025%20Modified%20Print%20Fn-1_251028_234052_compressed.pdf
- Kakooza, A. (2023). Digitization of the judiciary in Uganda: Prospects and challenges of the ECCMIS. *Uganda Christian University Law Journal*, 5(1), 45–62.
- Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLOS ONE*, 12(4), e0174698.
- Liu, Y., et al. (2023). Deep learning for legal analytics: Opportunities and challenges. *Artificial Intelligence and Law*, 31(2), 245–269.
- Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems* 30.
- Mbondenyi, M. K. (2022). *Judicial reforms in East Africa: A comparative analysis*. Routledge.
- Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28(2), 237–266.
- Mehrabi, N., Morstatter, F., Saxena, N., Lerman, K., & Galstyan, A. (2021). A survey on bias and fairness in machine learning. *ACM Computing Surveys*, 54(6), 1–35.
- Mitchell, M., Wu, S., Zaldivar, A., Barnes, P., Vasserman, L., Hutchinson, B., Spitzer, E., Raji, I. D., & Gebru, T. (2019). Model cards for model reporting. *Proceedings of the Conference on Fairness, Accountability, and Transparency*, 220–229.
- Molnar, C. (2020). *Interpretable machine learning*. Lulu.com.
- Morgan, D. L. (2014). Pragmatism as a paradigm for social research. *Qualitative Inquiry*, 20(8), 1045–1053.
- National Center for State Courts. (2022). *Case backlogs in state courts: Pandemic impacts and recovery strategies* (tech. rep.). NCSC.
- North, D. C. (1990). *Institutions, institutional change and economic performance*. Cambridge University Press.
- Park, J. (2022). AI and data analytics in the Korean judicial system: Innovations in case management. *International Journal for Court Administration*, 13(1), 1–12.
- Peppers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of Management Information Systems*, 24(3), 45–77.
- Peng, R. D. (2011). Reproducible research in computational science. *Science*, 334(6060), 1226–1227.
- Ragin, C. C. (2008). *Redesigning social inquiry: Fuzzy sets and beyond*. University of Chicago Press.

- Report of the National Court Case Census. (2025). *Case backlog and judicial performance in Uganda* (tech. rep.). Judiciary of Uganda.
- Rwanda Judiciary Report. (2024). *Pilot program on data-driven case management: Interim findings* (tech. rep.). Supreme Court of Rwanda.
- Sharma, R., & Patel, V. (2023). Leveraging artificial intelligence for case prioritization and backlog management in the Indian judiciary. *Journal of Law and Technology*, 45(2), 112–135.
- Silva, F. A., Costa, M. R., & Almeida, J. R. (2023). Procedural delays in civil law jurisdictions: An empirical analysis of Brazilian appellate courts. *International Review of Law and Economics*, 74, 106–118.
- Simon, H. A. (1957). *Models of man: Social and rational*. John Wiley & Sons.
- Sourdin, T. (2021). *The responsive judge: The future of justice*. Springer.
- Uganda Judiciary Efficiency Report. (2024). *Time-motion study of high court proceedings* (tech. rep.). Judiciary of Uganda.
- Wang, D., et al. (2022). Designing human-centered AI for public services: A case study of judicial analytics. *Government Information Quarterly*, 39(3), 101678.
- Zhang, B. H., Lemoine, B., & Mitchell, M. (2018). Mitigating unwanted biases with adversarial learning. *Proceedings of the 2018 AAAI/ACM Conference on AI, Ethics, and Society*, 335–340.
- Zheng, L., Li, H., Wang, W., & Xu, T. (2021). Predicting the duration of court cases using a survival analysis approach. *IEEE Access*, 9, 123456–123465.

Appendix A

BUDGET

Table A.1: Budget for Accomplishing a Research Project on a Predictive Analytics Framework for Case Backlog Management in Uganda’s Judiciary: An Explainable Machine Learning Approach

S/No	Item	Description	Unit Cost (UGX)	Qty	Freq.	Total (UGX)
A. Data Preprocessing						
1	Data Extraction	National Court Case Census 2025 case records	90,000	3	6	1,620,000
2	Data Cleaning and Annotation	Dataset preparation	90,000	3	10	2,700,000
Sub-Total 2						4,320,000
B. Computing / ICT Infrastructure						
3	Data Storage and Backup	Cloud and local storage	1,500,000	1	1	1,500,000
4	High-Performance Laptop	ML model development	8,000,000	1	1	8,000,000
5	External Hard Drives	Data backup	350,000	1	2	700,000
6	Cloud Computing Resources	GPU/Cloud processing	4,000,000	1	1	4,000,000
Sub-Total 3						14,200,000
C. Validation						

continued on next page

Table A.1 continued from previous page

S/No	Item	Description	Unit Cost (UGX)	Qty	Freq.	Total (UGX)
7	Machine Learning Expert Review	Quality assurance	1,000,000	1	3	3,000,000
8	Thesis Editing (LaTeX) and Proofreading	Editing services	2,000,000	1	1	2,000,000
9	Thesis Printing and Binding	Final copies	1,500,000	1	1	1,500,000
Sub-Total 4						6,500,000
Grand Total						25,020,000

Appendix B

RESEARCH PROJECT TIMELINES



Figure B.1: Research work plan (June 2025 – May 2026).

Appendix C

DECLARATION OF AI USE

Declaration Statement:

I, *BBOSSA ISAAC SSERUNKUMA*, declare that I used

- i. I used **ChatGPT** to **fix grammar and typos**. It helped me correct grammar, typos, and awkward phrasing in parts of my thesis, achieving a polished and professional tone without changing my intended message. I applied these language edits to the Abstract, Dedication, Introduction, and Literature Review.

Signature: _____ Date: _____

Appendix D

LIST OF 10 MAJOR JOURNAL PUBLICATIONS REVIEWED

- i. Katz, D. M., Bommarito, M. J., & Blackman, J. (2017). A general approach for predicting the behavior of the Supreme Court of the United States. *PLOS ONE*, 12(4), e0174698. <https://doi.org/10.1371/journal.pone.0174698>
- ii. Medvedeva, M., Vols, M., & Wieling, M. (2020). Using machine learning to predict decisions of the European Court of Human Rights. *Artificial Intelligence and Law*, 28(2), 237–266. <https://doi.org/10.1007/s10506-019-09255-y>
- iii. Bhatnagar, M., & Huchhanavar, S. S. (2026). Hybrid machine learning modelling with explainability for predicting case delays and durations in Indian lower courts. *Journal of Big Data*, 13(4), 1–30. <https://doi.org/10.1186/s40537-025-01340-1>
- iv. Bhatnagar, M., & Huchhanavar, S. (2024). Predicting delays in Indian lower courts using AutoML and decision forests. In *Computing, Internet of Things and Data Analytics (IC-CIDA 2023)*, Studies in Computational Intelligence (Vol. 1145, pp. 166–180). Springer, Cham. https://doi.org/10.1007/978-3-031-53717-2_16
- v. de Oliveira, R. S., Reis Jr., A. S., & Sperandio Nascimento, E. G. (2022). Predicting the number of days in court cases using artificial intelligence. *PLOS ONE*, 17(5), e0269008. <https://doi.org/10.1371/journal.pone.0269008>
- vi. Vercosa, L. F., de Almeida, J. P., & Santoro, F. M. (2024). Investigation of lawsuit process duration using machine learning and process mining. *Discover Artificial Intelligence*, 4(1), 15. <https://doi.org/10.1007/s44257-024-00015-0>
- vii. Malik, V., Sanjay, R., Louis, A., & Bhattacharya, P. (2021). ILDC for CJPE: Indian legal documents corpus for court judgment prediction and explanation. In *Proceedings of the Nat-*

ural Legal Language Processing Workshop 2021 (pp. 13–23). Association for Computational Linguistics. <https://doi.org/10.18653/v1/2021.nllp-1.2>

- viii. de Menezes-Neto, E., Teixeira, J. M. X. N., de Oliveira, A. L. I., & de Lima, E. S. (2022). Using deep learning to predict outcomes of legal appeals. *PeerJ Computer Science*, 8, e1036. <https://doi.org/10.7717/peerj-cs.1036>
- ix. Deeks, A. (2019). The judicial demand for explainable artificial intelligence. *Columbia Law Review*, 119(7), 1829–1850. Available at: <https://www.columbialawreview.org/content/the-judicial-demand-for-explainable-artificial-intelligence/>
- x. Eliot, L. B. (2020). Legal judgment prediction amid the advent of autonomous AI legal reasoning. *AI and Ethics Working Paper*. <https://arxiv.org/abs/2009.14620>

Appendix E

RESEARCH ALIGNMENT WITH NATIONAL (NDPIV) AND GLOBAL (SDGS) AGENDA

E.1 ALIGNMENT WITH THE NATIONAL DEVELOPMENT PLAN IV (NDPIV)

Table E.1: Alignment of the Research with Uganda's National Development Plan IV (NDP IV) Strategic Areas

S/No	NDP IV Strategic Area	NDP IV Objective	Alignment of the Research
1	Administration of Justice Programme	Improve access to justice and strengthen the rule of law.	Predicts backlog risk and supports proactive case management to reduce delays and improve access to justice.
2	Governance and Security Programme	Enhance transparency, accountability and institutional efficiency.	Provides objective, data-driven insights for judicial planning and resource allocation.
3	Digital Transformation Programme	Increase adoption of digital technologies in public service delivery.	Applies machine learning and explainable AI to judicial data for decision support.
4	Innovation, Technology Development and Transfer	Promote innovation and adoption of emerging technologies.	Introduces advanced predictive analytics techniques into justice sector management.

continued on next page

<i>Table E.1 continued from previous page</i>			
S/No	NDP IV Strategic Area	NDP IV Objective	Alignment of the Research
5	Public Sector Transformation Programme	Improve efficiency and effectiveness of public institutions.	Supports workload forecasting, performance monitoring and operational planning.
6	Development Plan Implementation Programme	Strengthen monitoring and evaluation systems.	Generates predictive indicators for tracking backlog trends and intervention outcomes.
7	Knowledge Economy	Promote data-driven governance and decision-making.	Demonstrates the value of judicial data as a strategic national asset.
8	Citizen-Centred Service Delivery	Improve quality and responsiveness of public services.	Facilitates timely case resolution and improved user satisfaction.
9	Institutional Capacity Building	Strengthen institutional performance and service delivery.	Supports optimal deployment of judicial officers and resources.
10	Results-Based Management	Enhance performance measurement and accountability.	Provides forward-looking performance indicators for court management.

E.2 ALIGNMENT WITH GLOBAL AGENDA

Table E.2: Alignment of the Research with Global Development Agendas and Governance Frameworks

Global Agenda	Strategic Objective	Alignment of the Research
SDG 16: Peace, Justice and Strong Institutions	Promote access to justice and build effective institutions.	Supports timely case resolution through prediction of backlog risk and proactive case management.
SDG 16.3	Promote the rule of law and equal access to justice.	Reduces delays and improves equitable access to judicial services.
SDG 16.6	Develop effective, accountable and transparent institutions.	Enhances accountability through data-driven judicial management.
SDG 9: Industry, Innovation and Infrastructure	Promote innovation and technological advancement.	Applies machine learning and explainable AI in justice administration.

continued on next page

<i>Table E.2 continued from previous page</i>		
Global Agenda	Strategic Objective	Alignment of the Research
SDG 17: Partnerships for the Goals	Strengthen knowledge sharing and technology transfer.	Generates knowledge that can be replicated across judicial systems globally.
World Bank Governance and Justice Reform Agenda	Improve public sector efficiency and justice delivery.	Supports resource optimization and workload forecasting.
UNDP Digital Transformation Agenda	Promote digital governance and data-driven services.	Uses judicial data and AI to improve public service delivery.
OECD Public Governance Principles	Encourage evidence-based policy making.	Provides predictive evidence for judicial reforms and policy decisions.
AI for Good Initiative	Apply AI to address societal challenges.	Uses explainable AI to address judicial delay and backlog challenges.
Open Government and Data Governance Initiatives	Improve use of public-sector data.	Transforms court data into actionable intelligence for decision-making.

Appendix F

RESEARCH POSTER LAYOUT

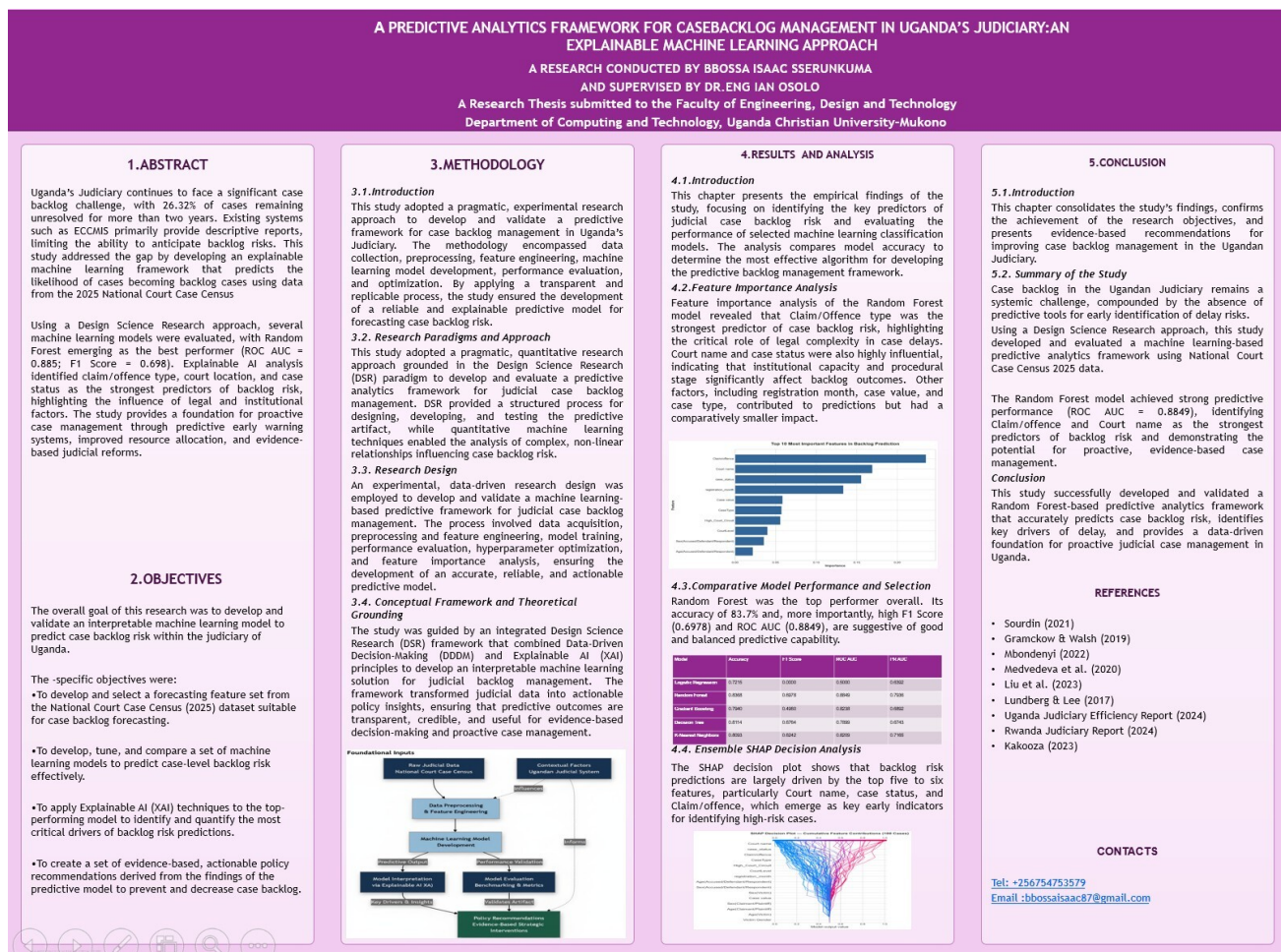


Figure F.1: Poster presentation Layout

Appendix G

Response to Supervisor Comments

A Predictive Analytics Framework for Case Backlog Management in Uganda's Judiciary: An Explainable Machine Learning Approach

Student: Bbossa Isaac Sserunkuma (Reg. No. M24M19/024)

Supervisor: Dr. Ian Raymond Osolo

Thank you for the comments on the thesis. Each is addressed below, with a summary of the revision and where it can be found in the resubmitted chapters. All changes have been compiled and cross-checked.

#	Comment	How it was addressed	Location
1	Rewrite the research objectives so they are SMART and avoid combining multiple objectives into one statement.	Objectives rewritten as six single-aim, SMART statements. Metrics named to match those actually reported (Accuracy, F1, ROC AUC, PR AUC); vague terms ("suitable", "effectively") removed. Each objective anchored to a study phase for the time-bound element.	<i>Ch. 1: Research Objectives</i>
2	Strengthen the connection between the problem statement and its theoretical and practical implications.	Problem statement expanded with the headline backlog figure (26.32%) and two bridging sentences, practical stakes (access to justice, socio-economic cost) and the theoretical gap (interpretable ML for backlog in a developing-country judiciary), which the Significance section then develops.	<i>Ch. 1: Statement of the Problem</i>
3	Revise objectives to reflect the stages of the study (preliminary investigation, data exploration, etc.), logically sequenced and aligned to the methodology.	Objectives now trace the full pipeline: scoping and data exploration → data preparation → model development → evaluation → interpretation → recommendations, aligned to the Chapter 3 methodology and the Appendix B work plan.	<i>Ch. 1: Research Objectives</i>
4	Strengthen the discussion by comparing results with existing literature – similarities, differences, and the contribution of the framework.	Added explicit comparison in Chapter 5: performance benchmarked against the generational AUC bands (matches 2nd-generation ensembles at 0.885; below the transformer 0.91 for a data reason); institutional-driver convergence with Mbondenyi (2022); divergence from the Brazilian and US procedural-chokepoint patterns; implementation aligned with Rwanda rather than Korea; and contribution stated against the Section 2.8 gaps, with Bhatnagar & Huchhanavar (2026) added as the closest analogue.	<i>Ch. 5: §5.2.1, §5.2.2, and new §Comparison with Related Work and Contribution; references updated</i>
5	Expand methodology and results to explain the relationship between dataset attributes and backlog; justify feature selection; discuss individual variables using explainability.	Ch. 3: new section mapping every attribute to its hypothesised relationship (Table 3.1); case_status introduced as a procedural input (previously omitted); explicit feature-selection justification (52 → 14 features on four stated principles). Ch. 4: new attribute-level subsection walking all four SHAP tiers and tying each variable back to its Chapter 3 hypothesis.	<i>Ch. 3: Relationship Between Dataset Attributes...; Feature Engineering and Selection. Ch. 4: Attribute-Level Contribution</i>
+	Follow-up query: since case_status is a top predictor, is it merely a proxy for the target?	Verified that it is not. The label is built from elapsed time alone; case_status shows weak standalone association (mutual information 0.029 vs 0.214; Cramér's V 0.238) and explains under 5% of case-age variance ($\eta^2 = 0.049$); an ablation lowers ROC AUC only from 0.885 to 0.831. Retained as legitimate. Two related claims surfaced by this check were also corrected: the Logistic Regression baseline (refit converges, F1 0.653, but ROC AUC stays 0.5000, argument re-grounded on the persistent AUC), and the learning curve (now acknowledges a train-validation gap driven by Claim/offence's 41,038 values).	<i>Ch. 3: Feature Engineering and Selection; Ch. 4: §4.3.1 and learning curve; Ch. 5 : §5.2.2</i>

Revised files accompanying this memo: Chapters 1, 3, 4, and 5, and the reference list (.bib).