



UGANDA CHRISTIAN UNIVERSITY

A Centre of Excellence in the Heart of Africa

SCHOOL OF RESEARCH & POSTGRADUATE STUDIES DISSERTATION CORRECTION COMPLIANCE FORM (POST VIVA FORM)

Date: 10th Sept 2026

Name of Candidate: Mbabazi Fiona

Reg.No: J24M19/009

Title of Dissertation: PREDICTING PHARMACY CLOSURES IN UGANDA USING MACHINE
LEARNING: A RISK-BASED APPROACH TO REGULATORY PRIORITIZATION

S/N	COMMENTS BY EXTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	Empirically compare all candidate models – particularly Logistic Regression, Decision Tree, Random Forest and XGBoost – before claiming that XGBoost is the best model.	Logistic Regression, Decision Tree, Random Forest and XGBoost were empirically evaluated using a common chronological train/test split and group-aware cross-validation framework. Comparative ROC-AUC, PR-AUC, precision, recall, F1-score and Precision@200 results were added to Chapter three, four and five.	Section 3.6.2, 3.6.3, 3.6.5, 4.5, 4.6, 5.2.2, 5.2.4
2	Strengthen the definition of pharmacy closure, distinguishing regulatory closure from voluntary closure, business failure and inactive/expired licences	The operational definition of pharmacy closure was revised to distinguish regulatory closure from voluntary closure, commercial/business failure, inactive status and licence expiry. The methodology now clarifies that the model outcome was derived from closure-related regulatory statuses available in the NDA dataset and does not necessarily represent independently verified physical business cessation. The corresponding limitation and interpretation of the outcome were also strengthened in Chapters 3 under section 3.11 Operational Definition of Pharmacy Closure”	Chapters 3 under section 3.11, page 31
3	Improve model validation and reporting through baseline comparisons, uncertainty estimates, calibration assessment and, where appropriate, temporal/prospective validation	Model validation was strengthened by empirically comparing four candidate models under a common framework, including Logistic Regression as a conventional baseline and test-set closure	

4	Strengthen the discussion by explaining why XGBoost performs as it does and distinguishing prediction from causation.	Chapter Five discussion under section 5.2.4 was revised to explain XGBoost's ability to capture nonlinear relationships and feature interactions in structured regulatory data, together with its regularisation and handling of class imbalance.	Section 5.2.4 page 58.
5	Demonstrate operational value by investigating whether risk-based ranking actually improves regulatory prioritisation compared with existing practices.	I investigated operational value by asking a practical regulatory question: if NDA could only prioritise 200 pharmacies, would model-based ranking identify more subsequent closure-related outcomes than non-targeted selection? Using Precision@200, Decision Tree identified 59 closures, Random Forest 50, XGBoost 41 and Logistic Regression 33 among the top 200 ranked pharmacies. This shows that risk-based ranking can concentrate higher-risk pharmacies within a limited inspection capacity. However, I did not claim that the models have already outperformed existing NDA practice because the dataset did not contain a complete record of historical prioritisation decisions. Therefore, I present this as evidence of potential operational value that should be confirmed through prospective implementation.	Chapter Four - Operational Evaluation / Precision@200; Chapter Five - Practical Interpretation and Recommendations
6	Strengthen the literature review through more critical synthesis and engagement with recent work on temporal validation, imbalanced learning, calibration and explainable AI.	The literature review was strengthened with critical synthesis and recent methodological literature on temporal validation, imbalanced learning, calibration and explainable AI. The revised discussion links these issues to the study's validation framework, performance metrics, class imbalance, interpretation of predicted probabilities and SHAP analysis.	Under Literature review, added sections 2.10, 2.11. 2.12
7	Correct presentation and formatting problems, including the department template, repeated content, terminology inconsistencies, broken bookmarks and reference formatting.	Presentation and formatting were reviewed and corrected, including use of the required departmental template, removal of repeated content, harmonisation of terminology, correction of broken bookmarks and standardisation of reference formatting.	Done No broken bookmarks in my Report

S / N	COMMENTS BY INTERNAL EXAMINER	ACTION TAKEN	INDICATOR
1	You identify a gap in the literature concerning ML-based pharmacy	The literature review was revised to define the research gap more precisely. The study	Added under section 2.14

	closure prediction in LMICs.	now identifies limited empirical evidence at the intersection of longitudinal pharmacy regulatory data, machine-learning closure prediction and risk-based regulatory prioritisation in Uganda rather than claiming that machine-learning closure prediction itself is entirely new.	page 15
2	How did you establish that this was genuinely a research gap rather than simply a gap in your literature search?	The literature search strategy and research-gap justification were strengthened in Chapter Two. The revised section documents the databases/search sources, search terms, date range, inclusion criteria, and backward and forward citation chaining used to identify relevant literature. The search was broadened beyond pharmacy closure to include related studies on pharmacy regulation, business and healthcare-facility closure prediction, regulatory compliance prediction, and machine learning in health and regulatory settings.	Strengthened in Chapter Two
3	What are some of the factors that lead to non-compliance and why did you choose pharmacy closure as an outcome and not regulatory non-compliance?	Regulatory non-compliance would certainly be an important outcome to predict, but I selected pharmacy closure because I needed a clearly defined subsequent outcome that could be constructed longitudinally from the historical NDA data. The dataset contained repeated inspections and regulatory-status information that allowed me to define whether a pharmacy entered a closure-related regulatory state within 180 days following an inspection. This made it possible to frame the study as a prospective prediction problem: using information available at an inspection to predict a later outcome. Non-compliance, on the other hand, is multidimensional. A pharmacy may be non-compliant because of licensing, staffing, storage, documentation, premises, product handling or other requirements, and these violations differ considerably in severity. A simple compliant/non-compliant label could therefore combine very different regulatory situations unless a validated	

		composite definition and sufficiently detailed inspection data were available. I therefore selected closure as the primary outcome because it provided a more consistently identifiable endpoint within the available historical data. However, I now define it carefully as a closure-related regulatory outcome rather than assuming that every inactive or expired licence	
4	What is the contribution of your study to pharmacy regulation in Uganda?	The study contribution was strengthened to show how routinely collected NDA inspection and licensing data can be converted into prospective risk rankings to support risk-based pharmacy regulation. The contribution emphasises targeted inspection planning, more efficient allocation of limited regulatory resources, empirical model comparison and a foundation for future data-driven regulatory surveillance.	
5	How would you classify your research and what informed your choice of models? Are you doing classification or regression?	My research is a retrospective predictive analytical study using supervised machine learning. Specifically, it is a binary classification problem, not a regression problem. The outcome has two classes: whether a pharmacy experiences the defined closure-related regulatory outcome within 180 days after inspection or does not." I selected the four models to represent different levels of complexity and modelling approaches. Logistic Regression provided a conventional interpretable linear baseline; Decision Tree provided a simple nonlinear classifier; Random Forest represented a bagging ensemble; and XGBoost represented gradient boosting. This allowed me to empirically determine whether increasing model complexity improved prediction rather than assuming that XGBoost would automatically be superior.	Chapter Three: Research Design, Model Development and Candidate Model Selection
6	Did you feed the model with your data as displayed before the panel?	Yes, as displayed raw regulatory records were not entered into the models unchanged; dates were parsed, the 180-day outcome was constructed, predictive features were engineered, missing values were handled, numeric variables were	Section 3.6.1 specifies it was wasn't a random selection

		scaled where applicable, categorical variables were encoded, and the data were divided using the chronological train-test framework.	
7	Wrong choice of the dataset.	The suitability of the dataset was reviewed against the study objectives. The study specifically seeks to predict pharmacy closure using routinely collected regulatory information; therefore, historical NDA licensing and inspection records were retained because they contain longitudinal inspection, licensing, regulatory status and geographic information relevant to the prediction task. Important variables such as financial performance, staffing, ownership changes and market conditions were unavailable. I therefore interpret the model as predicting closure risk from available regulatory data, rather than providing a complete causal explanation of pharmacy closure.	
S/N	COMMENTS BY VIVA VOCE PANNEL	ACTION TAKEN	INDICATOR
1			
2			
3			

Candidate's Name Mbabazi Fiona

Signature



Supervisor's Name/ Signature

Prof. Francis Otho  19/9/2026

NB: Post Viva compliance form is designed to capture all the corrections recommended by internal examiner (supervisor), external examiner and viva panel.