# **CHAPTER THREE RESEARCH METHODOLOGY**

## **Research Design**

This study involves experimentation, wherein specific groups of variables remain consistent, while another group of variables is observed as the primary focus of the investigation.The study involves conducting experiments using various predictive modeling techniques for credit risk of commercial banks in Tanzania. The objective is to apply different Predictive modeling Techniques for Credit Risk. The performance of each trained Predictive modeling algorithm is then evaluated to determine the best predictive model.

## **Research Approach**

**3.4 Variables**

The study will consider a range of variables, including but not limited to: Loan performance metrics, Borrower demographics, Financial indicators (e.g., liquidity ratios, leverage ratios), Economic indicators, Alternative data sources.

These variables will be instrumental in constructing a comprehensive dataset for the development of predictive models.

**3.5 Predictive Modeling Techniques**

The Study will employ several machine learning algorithms to develop predictive models for credit risk management. The chosen algorithms include Logistic Regression, Decision Trees, Random Forest, Support Vector Machines, and Long Short-Term Memory. The selection of these techniques is based on their suitability for handling different aspects of credit risk and their relevance to the Tanzanian banking context.

**3.6 Model Development**

The predictive models will be developed using a two-step process. In the first step, a comprehensive dataset will be prepared by cleaning, preprocessing, and transforming the collected data. In the second step, the dataset will be used to train and validate the selected machine learning algorithms. The models will be fine-tuned to ensure optimal performance in predicting credit risk.

**3.7 Model Evaluation**

The performance of the developed models will be assessed using relevant metrics such as accuracy, precision, recall, and the area under the ROC curve. The evaluation process will provide insights into the effectiveness of the predictive models in comparison to traditional credit scoring methods.

## **Study Area**

This research is centered on the United Republic of Tanzania, a nation in East Africa known for its diverse economic landscape and a thriving financial sector. The Study will narrow its focus to three commercial banks operating in Tanzania, namely Stanbic Bank Tanzania, CRDB Bank, and NMB Bank. These banks were selected based on their significant market share, operational scale, and diverse credit portfolios. By concentrating on these institutions, the study aims to provide comprehensive insights into credit risk management practices across different banking models.

## **Tools and Materials**

This study will utilize a variety of materials, including but not limited to: -

| **S/No** | **Material** | **Tool description** |
| --- | --- | --- |
| 1 | portable computer | for computation |
| 2 | software | JetBrains PyCharm for the analysis |
| 3 | modem | for Internet access to retrieve various literature materials |
| 4 | interview | for capturing quantitative data |
| 5 | Programming Languages | Python version 3.10.4, JavaScript ECMAScript 2023 |
| 6 | Data formats | JavaScript Object Notation (JSON), Comma Separated Value (CSV), Excel (xls, xlsx) |
| 7 | Virtual Environment | Pipenv |
| 8 | Web Development Frameworks | Django 3.1.6, StreamLit |
| 9 | Version control | Git |
| 10 | Libraries | NumPy, Pandas, JupyterLab, Scikit-learn, Gensim, Keras matplotlib |

Table 3.1 tools and materials to be used in this study

## **Data Collection**

The data for this study will be gathered from the three commercial banks in Tanzania, namely Stanbic Bank Tanzania, CRDB Bank, and NMB Bank.The selection criteria are based on the market share,operational scale, and diversity of credit portfolios of the three selected banks.This ensure that the dataset encompasses a broad spectrum of banking activities, contributing to a thorough analysis of credit risk management. The data spans from 2020 to 2023.

This study will employ a combination of primary and secondary data collection methods. Historical credit data will be obtained, including information on loan performance, borrower characteristics, and relevant financial indicators.

Secondary data will be acquired from various sources such as financial reports, credit histories, industry publications, and academic journals will be consulted to gather supplementary information. And the source will be from the official records of Stanbic Bank Tanzania, CRDB Bank, and NMB Bank. These documents will serve as the primary source for evaluating credit risk management strategies and outcomes.

### **Document Review**

As part of the secondary data collection, this study will review multiple documents from diverse sources including research papers, academic articles, journals, open-data platforms, and relevant websites. The purpose of this document review is to gain insights into different techniques used for analyzing and predicting credit risk for commercial banks. It aims to gather information about the approaches adopted by other researchers, their methodologies, and the findings they obtained. Additionally, the review will aid in identifying existing Predictive modeling Techniques and selecting suitable models that can be customized to align with the specific context of this research.

## **Data Analysis**

In this research, the application of Exploratory Data Analysis (EDA) will be employed to conduct initial investigations on the collected data. The purpose is to uncover patterns, derive meaningful insights, identify anomalies, gain a comprehensive understanding of the data, and provide summary statistics and graphical representations. The initial data analysis will specifically focus on the datasets containing customer information from Commercial banks in Tanzania during the period of 2020 to 2023. To facilitate this process, a Python virtual environment will be established. Python scripts will be developed to handle the large JSON files associated with each month, enabling them to be split into smaller chunks. These chunks will be saved in the Comma Separated Value (CSV) format, ensuring ease of use and effective data manipulation.

## **Ethical Consideration**

To address the ethical considerations associated with this study, research clearance letters will be obtained from UDSM and also the research clearance letters will be provided to three commercial banks in Tanzania, which will be involved in data collection: Stanbic bank Tanzania, CRDB, and NMB Banks . The potential respondents will be informed about the benefits of the study and will be requested to participate voluntarily throughout the data collection process. They will be fully informed and assured that this research is a necessary component of academic requirements and will be utilized to develop a model for predicting Credit risk of Commercial Banks in Tanzania.