UNIVERSITI TEKNOLOGI MARA

POVERTY RISK PREDICTION BASED ON SOCIOECONOMIC FACTORS USING MACHINE LEARNING APPROACH

NUR FARHANA ADIBAH BINTI MOHD ZAWARI

BSc

July 2025

UNIVERSITI TEKNOLOGI MARA

POVERTY RISK PREDICTION BASED ON SOCIOECONOMIC FACTORS USING MACHINE LEARNING APPROACH

NUR FARHANA ADIBAH BINTI MOHD ZAWARI

Proposal submitted in partial fulfilment
of the requirements for the degree of
Bachelor of Science (Hons.) Management Mathematics

Faculty of Computer and Mathematical Sciences

July 2025

APPROVED BY:

……..

BALKIAH MOKTAR

Supervisor

Faculty of Computer and Mathematical Sciences

Universiti Teknologi MARA

ABSTRACT

Poverty remains a persistent socioeconomic issue in Malaysia, affecting the quality of life, access to education, employment opportunities, and long-term wellbeing. The process of classifying individuals or households that may be at risk of poverty can be time consuming and less accurate in relation to traditional methods like Poverty Line Income (PLI). As the concept of data analytics grows, machine learning provides a potent solution that can be used to reduce poverty via predictive modelling. This study seeks to develop a predictive model of measuring poverty risk using socioeconomic factors based on a machine learning framework. A secondary dataset that considered 635 households of Terengganu was used, and the following aspects were identified as important indicators of poverty: age, income, education, occupation, and health. Information gain was used in the feature selection and four classification algorithms namely, Logistic Regression, Random Forest, Decision Tree, and Gradient Boosted, were implemented and tested with the incorporation of 10-fold cross-validation and splitting 70:30 in WEKA. The findings indicated that the Logistic Regression outperformed the other algorithm with 99.06% using cross-validation and 98.42% using the splitting method, and with the best value of precision, recall, and F1-score. The feature that was found to be the most influential predictor of poverty risk was age. These findings imply that Logistic Regression is the suitable and interpretable model that can be used with structured data in the classification of poverty. Although the research is limited with respect to its sample size and geographical scope, it has provided important findings that can be used when implementing data-driven methods in social policy formulation and poverty mitigation strategies.

TABLE OF CONTENTS