

1. Introduction

This study investigates the problem of predicting a company's Free Cash Flow (FCF) using the Light Gradient Boosting Machine (LGBM) algorithm. Free cash flow is defined as the cash generated by a company from operating activities after deducting capital expenditure [1]. In Indonesia, FCF data is publicly available through each company's financial statements [2]. This prediction task is particularly important, as FCF serves as a critical indicator for potential investors to evaluate whether a company's stock is a worthwhile investment. Accurate FCF forecasts provide a solid foundation for assessing company valuation and long-term profitability.

Several previous studies have focused on forecasting FCF or related financial indicators using various techniques. In 2021, Ray Ball et al, conducted a study utilizing Ordinary Least Squares (OLS) and linear regression to predict free cash flow. The objective of their research was to examine the extent to which operating cash flow and corporate earnings influence future cash flows. Their findings suggested that earnings played a more significant role than operating cash flow in predicting future cash flows [3].

In 2022, Lin Zhu et al, developed a free cash flow prediction model using a neural network optimized with a genetic algorithm. The goal was to enhance the accuracy of corporate FCF forecasting, providing valuable insights for potential investors. The results demonstrated that a model with a population size of 30 and 15 hidden layer neurons yielded the most effective prediction performance [4].

Further expanding on hybrid approaches, in 2021, Bahman Talebi et al, proposed an optimized cash flow prediction model using new data mining methods that combined artificial intelligence algorithms. Their findings highlighted that artificial neural networks enhanced with Particle Swarm Optimization (ANN+PSO) and Genetic Algorithms (ANN+GA), using 15 predictor variables, were the most optimal models [5].

In another notable work, Şirin Özlem and Omer Faruk Tan (2022) investigated the prediction of corporate cash holdings using supervised machine learning algorithms. Their study employed a diverse set of methods including XGBoost, K-Nearest Neighbors (KNN), Multiple Linear Regression (MLR), Support Vector Regression (SVR), Random Forest (RF), Decision Trees (DT), and Multilayer Neural Networks (MLNN). The results revealed that newer algorithms significantly improved prediction accuracy, with XGBoost delivering the best performance among the tested models [6].

In this study, free cash flow will be predicted using the LGBM method, which is based on gradient boosting decision trees. This method offers efficient computation, robust performance on structured data, and effective handling of feature interactions. The dataset will be analyzed and processed to detect underlying trends and patterns before the model is trained and validated to forecast future free cash flows. Unlike previous studies, this research introduces a novel approach by applying LGBM to predict FCF specifically for companies listed on the Indonesian Stock Exchange (IDX), while also incorporating multiple evaluation metrics and exploring dimensionality reduction techniques to improve model performance. The output of this prediction is expected to assist potential investors in making more informed decisions.