

ABSTRACT

Windy Mega Melani, 17118365

Comparison of Linear Regression and Decision Tree Regression Methods with Uber vs Lyft Dataset

Thesis. Information Systems. Faculty of Computer Science and Information Technology. Gunadarma University. 2022

Keywords : Prediction, Comparison, Linear Regression, Decision Tree Regression

(xiv, 57, Attachment)

The development of information technology and the rapid flow of data requires proper management to produce relevant knowledge for decision-making, namely data mining. This study uses Uber and Lyft datasets as a case study testing material to predict an online transportation price. The purpose of this study is to find out how to implement the Linear Regression and Decision Tree Regression models and find out which model is better for predicting online transportation prices according to the dataset that has been prepared. The method used in this research is Cross Industry Standard Process for Data Mining (CRISP-DM). CRISP-DM has several stages, namely Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, and Deployment. The results of the accuracy test on the Linear Regression model in three ratios, namely 70:30, 75:25, and 80:20 are 52.32%, 52.36%, 52.32% while the accuracy results of the Decision Tree Regression model are 94.58%, 94.61%, 94.67%. The results of the comparison of the Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) values in the Linear Regression model are 14.5194 and 11.7207 and for the Decision Tree Regression model are 4.8600 and 2.9052. The conclusion is that the Decision Tree Regression model has greater accuracy and the RMSE and MAE values are smaller than the Linear Regression model. This shows that the Decision Tree Regression model has better performance than the Linear Regression model.

Bibliography (2012-2021)