A Comparative study of predicting the award-winning books using machine learning algorithms

Wijekoon S.M.S.T.[1] and Rupasingha R.A.H.M.[2*]

[1]*Department of Computing & Information Systems, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.*
[2]*Department of Economics and Statistics, Sabaragamuwa University of Sri Lanka, Belihuloya, Sri Lanka.*

Most people around the world are very fond of reading books offline and online. Different people have different preferences when it comes to choosing books and it has become important to find better books for readers. The awards are one of the main criteria to select the best books. There are no existing studies relevant to the same idea. Therefore, the purpose of this study is to predict the best books such as award(s) winning books based on different ten attributes such as author, average ratings, date, genre, language, pages, publisher, ratings, reviews, and title. For this, a data set has been obtained through online community platform called Kaggle. The dataset contains information about books obtained from the 'Goodread' website. Those books are related to the period between years 2000 and 2021. The dataset is pre-processed by removing duplicate data, removing unnecessary special characters and removing missing values etc. The Waikato Environment for Knowledge Analysis (WEKA) data mining tool is used to rank the preprocessed data, and hyper-parameter tuning was applied to enhance the outcomes. Six machine learning methods at the classification level including Random Forest, Support Vector Machine, Decision Trees, Multilayer Perceptron, Logistic Regression, and Naive Bayes have been used to generate prediction models. Using a Naive Bayes algorithm, we achieved 86.32% higher accuracy with higher precision, recall, and f-measure values, as well as the lowest error rates. The proposed method can predict the best books for readers using the above attributes by the Naive Bayes algorithm successfully.

Keywords: Award, Books, Classification, Machine Learning, Prediction

*Corresponding author: hmrupasingha@gmail.com