Identification of Mislabelled Products Using Machine Learning for an Online Retail Analytics Company

ABSTRACT

The aim of the project was to investigate the suitability of a machine learning algorithm in identifying mislabelled products within the company’s databases. The analysis conducted involved identifying the contributing factors of mislabelled products, merging data sets, and training multiple supervised machine learning algorithms, using the statistical package R. These algorithms were evaluated using a variety of performance metrics and a suitable algorithm was selected. The results indicate that the recommended machine learning algorithm will find the mislabelled products to a high degree of accuracy.