Abstract

During the last decade, particularly after 2010, there has been a sharp increase in the number of people who chose a meatless lifestyle. This trend is continuously growing rapidly around the world resulting in a higher demand of children's books for vegetarian families. Many authors already created numerous books in this context; however, we wonder if traditional folktales would suit this need. In this thesis, we took advantage of computational advancement to classify folktales written in or translated into English around the 1900s from different cultures and locations into five different dietary classes. For classification, rule-based and hybrid machine learning systems were implemented. Due to the imbalanced nature of the small corpus we created, the method of oversampling with virtual examples was used to augment and balance out our dataset for the latter algorithm and both models achieved fairly similar results. The final application is deployed online aiming to assist anyone who loves traditional fairy tales to find out if the story they are intending to read suits their dietary preferences. In the future, one could extend the system to be able to process more specific contents.

Keywords: folktale, fairy tale, classification, support vector machine, vegetarian, vegan, fruitarian, oversampling, rule-based, virtual examples