ABSTRACT

Web-based Application using Flask Framework for Classifying SMS using Logistic Regression Algorithm

A type of SMS spam is a type of unwanted or unsolicited text message that is sent to a user's cell phone, often for commercial purposes. To find out how SMS spam can be classified, the author makes a research using machine learning that is integrated into a web application. The web application was chosen because only by entering an SMS message on the form, the classification results can be immediately known without the need to download additional applications such as on the mobile platform, which for sure will take up more memory space. The machine learning algorithm used is Logistic Regression. Logistic Regression was chosen because it has a fairly good performance in classifying datasets with labels 1 and 0 according to previous researches. Before training the model, the SMS dataset that has been obtained is preprocessed using a tokenizer, stop words and TFIDF to convert text data into numbers, so machine learning can process it. The dataset that has gone through the preprocessing process is tested on the Logistic Regression parameters, test size and solver. The test results show that the test size with the highest accuracy value is test size 0.4 with an accuracy value of 1 for train data, and 0.97 for test data, while the solver that produces the highest accuracy values is lbfgs, and liblinear. Each solver produces an accuracy value of 1 for train data, and 0.97 for test data. The accuracy value of the machine learning model that has been made using logistic regression with the test size parameter of 0.4, and the lbfgs solver produces 0.97 on the Accuracy scoring type. Machine learning model can detect which messages are spam and which are not spam by analyzing the pattern of the word SMS, whether the word pattern is similar to the pattern of words from SMS in the dataset that were previously labeled 0 (spam) and 1 (not spam). The machine learning model is implemented into a web application using a framework from Python, Flask.

Keywords: SMS, Spam, Machine Learning, Logistic Regression, Test Size, Solver, Flask, Web.