



Date: March 20, 2019
To: Mustafa Bilgic, Director of ML@IIT
From: Ekrem Guzelyel, Undergraduate Research Assistant at ML@IIT
Subject: Transparent Text Classification

Abstract

Machine Learning algorithms are considered as black box. The way they make decisions is unknown to the user until a further analysis is done on the model. Text documents are no different. In order to make a good justification, we propose using EDUs that are derived from the most influential keywords in the documents, which will answer the question of “why” the model picked a class for a document. However, the need for this data is huge. We will first find the best way to generate a synthetic data by automating the labeling for the rest of the EDUs; then, we will use them to return reasoning for the classification.