Using Machine Learning to Predict Judicial Decisions

Conor O'Sullivan , Master of Science in Computer Science

University of Dublin, Trinity College, 2019

Supervisor: Joeran Beel

In this study, machine learning models were constructed to predict whether judgements made by the European Court of Human Rights (ECHR) would lead to a violation of an Article in the Convention on Human Rights. The problem is framed as a binary classification task where a judgement can lead to a "violation" or "non-violation" of a particular Article. Using auto-sklearn, an automated algorithm selection package, models were constructed for 12 Articles in the Convention. To train these models, textual features were obtained from the ECHR Judgment documents using N-grams, word embeddings and paragraph embeddings. Additional documents, from the ECHR, were incorporated into the models through the creation of a word embedding (echr2vec) and a doc2vec model. The features obtained using the echr2vec embedding provided the highest cross-validation accuracy for 5 of the Articles. The overall test accuracy, across the 12 Articles, was 68.83%. As far as we could tell, this is the first estimate of the accuracy of such machine learning models using a realistic test set. This provides an important benchmark for future work. As a baseline, a simple heuristic of always predicting the most common outcome in the past was used. The heuristic achieved an overall test accuracy of 86.68% which is 29.7% higher than the models. Again, this was seemingly the first study that included such a heuristic with which to compare model results. The higher accuracy achieved by the heuristic highlights the importance of including such a baseline.