Examining Copyright judicial rulings in South Africa using Machine Learning and NLP Techniques

An empirical analysis of appellate copyright jurisprudence in South Africa

INTRODUCTION

- The University of Pretoria Law Faculty has indicated that "there has been no empirical investigation of judicial sentiment on copyright issues and questions in South Africa. Yet, there are claims that copyright law involves aspects of public policy which are determined by courts".
- The objective of the project is to apply data science through the use of NLP techniques and machine learning to perform sentiment analysis and classification of the judicial rulings in the field of copyright law using the data from Supreme Court of Appeal.
- Three classification models: 1.Sentiment analysis on defence statement, 2. Classify judicial ruling to the copyright issue 3. judicial ruling into copyright work

METHODS

Figure 1 outlines the steps followed to deliver the project objective.

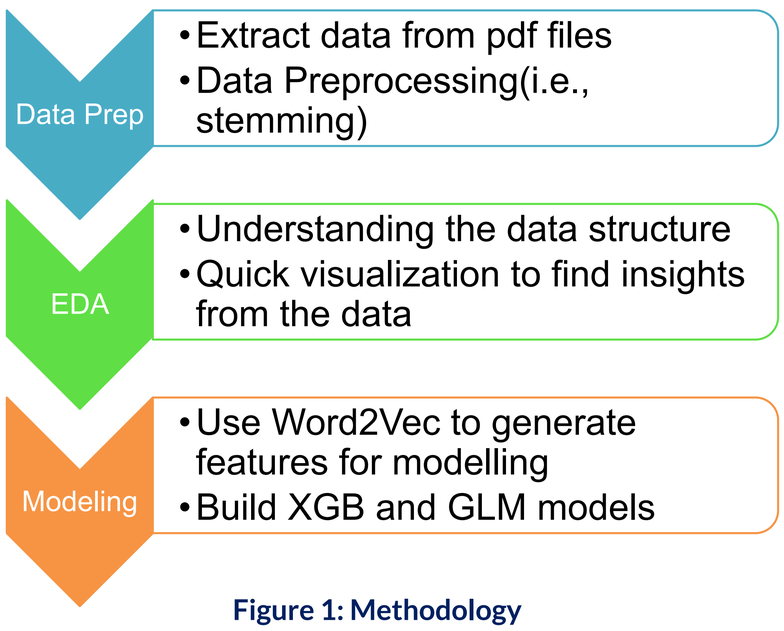

Figure 1: Methodology

RESULTS

Table 1 gives model performance for each classifier.

Objective	Model Selected	Model Performance (Accuracy %)
Sentiment analysis model	XGB	56
Copyright work Model	XGB	42
Copyright Issue Model	GBM	76

Table 1: Results

DISCUSSION

- Sentiment Analysis

The model accuracy score was 62% out of the 35 pdf's provided it identified that 57.1% of the defence statement were rejected, 33.3% were affirmed and 9.52%were unresolved.

- Copyright work classifier

The model performed poor in classifying the judicial rulings into the different types of copyright work. The model accuracy score was 45%. Most of the judicial rulings were classified as artistic work copyright work at 38.1% with the least being musical work. More data will be needed to improve the model.

- Copyright issue classifier

The results shows that the infringement copyright issue was the dominant one from the judicial ruling making up 66.7%. The model accuracy score was 77% which is higher than the two classifiers.

Exploratory Data Analysis Overview

Figure 2: Word cloud

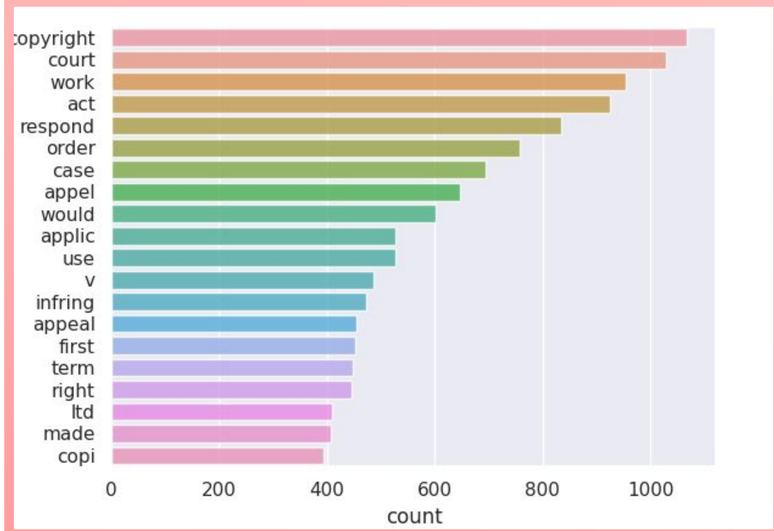

Figure 3: Top 20 words count

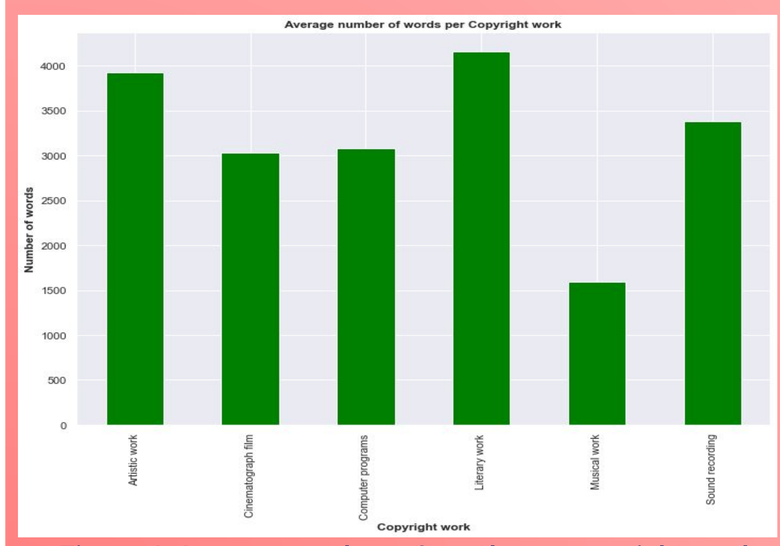

Figure 4: Average number o f words per copyright work

Xolile O'Reilly, **Kapei** Sebesho

UNIVERSITEIT VAN PRETORIA
UNIVERSITY OF PRETORIA
YUNIBESITHI YA PRETORIA

Department of Computer Science

Faculty of Engineering, Built Environment and Information Technology

Fakulteit Ingenieurswese, Bou-omgewing en Inligtingtegnologie / Lefapha la Boetšenere, Tikologo ya Kago le Theknolotši ya Tshedimošo

Capstone Project - MIT 808

Course Coordinators:
Dr. Vukosi Marivate (vukosi.marivate@cs.up.ac.za)
Ablodun Modupe (ablodun.modupe@cs.up.ac.za)

Scan me