Supplementary Material for: Predicting decisions of the European Patent Office's Boards of Appeal using Machine Learning (Submission to JURIX 2023 Doctoral Consortium)

David Bareham[0009−0000−7542−8966]

Department of Computer Science, University of Liverpool, UK
d.bareham@liverpool.ac.uk

1 Data and Pre-Processing

1.1 Extraction

Using simple keyword matching in SpaCy on the Summary of Facts section of the textual content, I created and tested a series of patterns to identify the type of appeal. The types of appeal are Opposition Division appeal, Examining Division appeal, Admissibility and Other. Admissibility corresponds to cases which solely concern admissibility rather than patent validity, and Other corresponds to cases which cannot be classified from the existing patterns.

The patterns were created by manually analysing a set of 50 randomly sampled appeals and splitting them into types based on the keywords used. The patterns were as follows[1]:

- Opposition Division = [“LOWER”: ‘opposition’}; {“LOWER”: ‘division’}]
- Admissibility = [{“LOWER”: ‘restricted’};{“OP”: ‘*’};{“OP”:‘*’};{“LOWER”: {“FUZZY”: “admissibility”}}]
- Examining Division = [{“LEMMA”:‘refuse’};{“OP”:‘*’};{“LOWER”: ‘european’,“OP”:‘*’};{“LOWER”:‘patent’,‘OP’:‘*’};{“LOWER”: ‘application’}]

The patterns were initially tweaked until 100% accuracy was achieved on the original 50 appeals. To test their generalisability, another random sample of 50 appeals were selected with an initial success rate of 47/50 classified correctly. The 3 misclassified appeals were Examining Division appeals using previously uncaptured keyword patterns. After alteration, the patterns achieved 50/50 and I was satisfied with the accuracy I conducted a final generalisability test on a group of 50 cases cases achieving 98% accuracy.

Furthermore, I checked a small sample of cases in the ’Other’ category for glaring omissions in the final dataset but none were observed. I then separated the the Opposition Division cases from the Examining Division cases.

[1] LOWER = lowercase; OP = optional; * = wildcard token; FUZZY = alternate spellings are acceptable; LEMMA = any acceptable lemmatisation of the word

Examining Division Appeals The final extraction task for the examining division cases was to extract the target label using the Order section of the decision which provides the board's outcome. The phrasing of the outcomes are relatively homogeneous and four different types of outcome were observed: the appeal was dismissed, the appeal was rejected for being inadmissible, the decision under appeal was set aside and Other outcomes. These first two outcomes were treated the same as they both result in the original decision by the examining division being maintained, thus they were labelled as 'Affirmed', whereas the previous decision being set aside reverses the prior outcome so was labelled as 'Reversed'. The Other outcomes refer to unique or infrequent outcome decisions such as referrals to the Enlarged Board of Appeal, which the patterns could not detect and thus were excluded from the analysis. The patterns were as follows:

– Dismissed = [{"LOWER":{"FUZZY": "appeal"}},{'OP':'*'},{"LOWER": "dismissed"}]
– Rejected = [{"LOWER":{"FUZZY": "appeal"}},{'OP':'*'},{"LOWER": "rejected"}]
– Set Aside = [{"LOWER":{"FUZZY": "appeal"}},{'OP':'*'},{"LOWER": "set"},{"LOWER": "aside"}]

I created the patterns on the same manually analysed random sample of 50 appeals, used to identify case type, before sampling 50 more appeals to test generalisability. For both samples the pattern achieved 100% accuracy providing confidence in its labelling abilities on this dataset.

Opposition Division appeals An additional complexity in regard to appeals against decisions from the Opposition Division are that the appellant can be the patentee, the opponent or both parties. Furthermore, there may be multiple opponents involved. In order to identify which parties brought the appeal further pattern matching was performed for these appeals. The categories were Patentee, Opposition, Both and Other. The patterns were as follows:

– Both Parties 1 = [{"LOWER": {'FUZZY': "patent"}}, {"LOWER": 'proprietor'},{"LOWER": "and"},{"OP":'{,5}'}, {"LOWER": {"FUZZY": "opponent"}}]
– Both Parties 2 = [{"LOWER": {"FUZZY": 'proprietor'}},{'OP': '{,3}'},{"LOWER": '('},{"OP": '{,5}'},{"LOWER": {"FUZZY": "appellant"}},{"OP": '{,5}'},{"LOWER": ')'},{"OP": '{,10}'},{"LOWER": {"FUZZY": "opponent"}},{'OP': '{,3}'},{"LOWER": '('},{"OP": '{,5}'},{"LOWER": {"FUZZY": "appellant"}},{"OP": '{,5}'},{"LOWER": ')'}]
– Both Parties 3 = [{"LOWER": {'FUZZY': "opponent"}},{'OP': '{,3}'},{'LOWER': {'FUZZY': "proprietor"}}
,{'LOWER': {'FUZZY': "appeal"}}]
– Opposition 1 = [{"LOWER": {'FUZZY': "appellant"}},{"OP": '{,3}'},{"LOWER": "("},{"OP": '{,5}'},{"LOWER": {"FUZZY": "opponent"}},{"OP": '{,5}'},{"LOWER": ')'}]

- Opposition 2 = [{“LOWER”: {‘FUZZY’: ‘opponent’}},{‘OP’: ’{,3}’},{“LOWER”: ‘(’},{“OP”: ’{,5}’},
  {“LOWER”: {“FUZZY”: “appellant”}},{“OP”: ’{,3}’},{“LOWER”: ‘)’}]
- Opposition 3 = [{“LOWER”: {‘FUZZY’: ‘opponent’}},{‘OP’: ’{,3}’},{“LOWER”: {“FUZZY”: “appeal”}}]
- Opposition 4 = [{“LOWER”: {‘FUZZY’:‘appeal’}},{’LOWER’: ’of’},{“LOWER”: {“FUZZY”: “opponent”}}]
- Patentee 1 = [{“LOWER”: {‘FUZZY’: “appellant”}},{“OP”: ’{,3}’},{“LOWER”: ”(”},{“OP”: ’{,5}’},{“LOWER”: {“FUZZY”: ”patent”}},{“OP”: ’{,3}’},{“LOWER”: ‘)’}]
- Patentee 2 [{“LOWER”: {‘FUZZY’: ‘proprietor’}},{‘OP’: ’{,3}’},{“LOWER”: ‘(’},{“OP”: ’{,5}’},{“LOWER”: {“FUZZY”: “appellant”}},{“OP”: ’{,3}’},{“LOWER”: ‘)’}]
- Patentee 3 [{“LOWER”: {‘FUZZY’: “appellant”}},{“OP”: ’{,3}’},{“LOWER”: ”(”},{“OP”: ’{,5}’},{“LOWER”: {“FUZZY”: “proprietor”}},{“OP”: ’{,3}’},{“LOWER”: ‘)’}]
- Patentee 4 [{“LOWER”: {‘FUZZY’: ‘patent’}}, {‘OP’: ’{,3}’},{“LOWER”: ‘(’},{“OP”: ’{,5}’},{“LOWER”: {“FUZZY”: “appellant”}},{“OP”: ’{,3}’},{“LOWER”: ‘)’}]
- Patentee 5 [{“LOWER”: {‘FUZZY’: ‘proprietor’}},{‘OP’: ’{,3}’},{“LOWER”: {“FUZZY”: “appeal”}}]
- Patentee 6 [{“LOWER”: {‘FUZZY’: ‘appeal’}},{“LOWER”: ‘by’},{“LOWER”: ‘the’},{“LOWER”: {“FUZZY”: “proprietor”}}]

First a sample of 70 cases was taken and manually annotated to build the patterns. After a sufficient degree of accuracy was achieved on this set the patterns were evaluated on a final group of 50 opposition appeals achieving 90% accuracy. This score was lower than for the other pattern matching tasks and upon analysis it became evident that the ‘Both’ and ‘Other’ categories were causing the decrease in accuracy. All the cases the patterns identified as ‘Other’ were not in fact ‘Other’ and two thirds of the cases identified as ‘Both’ were not actually ‘Both’. Due to the unreliability of these categories the decision was made to exclude them from the analysis. However, all appeals identified as either ‘patentee’ or ‘opposition’ were correct providing confidence in using these categories for the model building.

A final outcome extraction task was performed on the opposition and patentee appeals, using the same patterns as for the Examining Division outcomes. This achieved 100% accuracy on 50 sampled cases.

1.2 Pre-Processing

Before the Summary of Facts section is ready to become the input of our ML models, a number of pre-processing steps must be performed to reduce noise and ensure consistency across all appeals. These steps are as follows:

1. Remove: whitespace, punctuation, XML tags, HTTP links, non-alpha characters, individual letters other than ‘i’ and ‘a’ as not valid words, the first 35 characters from each case as they are boilerplate and not case-specific

2. Lowercase all text
3. Outcome is labelled 1 for Affirmed and 0 for Reversed
4. Vary the inclusion of numerical characters, stopwords and lemmatisation as pre-processing hyperparameters

1.3 European Patent Full-Text Data for Text Analytics

This dataset consists of XML-tagged titles, abstracts, descriptions, claims and search reports of European patent publications from 1978 onwards. The data is split into 40 different files, each averaging around 5-6GB in size and covering patent publications associated with 100,000 publication numbers. Due to the size of the total dataset, a subset of 5 files was chosen from this dataset to train the embedding models to ensure the models would train quickly whilst still providing sufficient text from 500,000 publication numbers.

As an individual file from this dataset is quite large, a batch streaming approach was used to load the data in increments of 10,000 publications at a time for training the embedding models. From this a pandas dataframe was created to filter only English entries and exclude HTTP links, to the original documents as PDFs, before a number of pre-processing steps were undertaken:

1. All data including titles, abstracts, claims, descriptions and amendments were used
2. Step 1 from the EPO Decisions dataset pre-processing was repeated
3. Stopwords were not excluded to provide more accurate context for embeddings i.e. 'not' is a stopword but is contextually important. Similarly, no lemmatisation was performed

2 Parameters and Full Results

This section includes the full nested cross-validation results and the best models alongside their associated hyperparameters and feature representations. All nested cross-validation results for experiment 2 are calculated using a weighted-average over the outer cross-validation scores due to the nature of the time series split in the cross-validation procedure. To preserve the time series, the initial splits are smaller than the later splits as each split permits a greater number of training cases, across a larger period of time. The weighted average is calculated for n weights and x splits:

$$Weights : w_i = \frac{i}{n} \; for \; i = 1, ..., n.$$

$$WeightedAverage : \sum_{i=1}^{n} w_i * x_i$$

Table 1. Hyperparameters

Name	Value	Description
N-Gram Parameters		
ngram_range	(1,1),(1,2),(1,3),(1,4),(2,2),(2,3),(2,4),(3,3),(3,4),(4,4)	Length of the n-grams
norm	None, 'L2'	Normalisation term for vectors
min_df	2, 5, 10	Minimum document frequency the terms must appear in to be included
use_idf	True, False	Use Inverse Document Frequency weighting (TF-IDF)
Model Parameters		
C	0.1, 1, 10, 100	SVM and LR: Regularisation strength
solver	'lbfgs', 'sag'	LR: Algorithm to use in the optimisation problem
penalty	None, 'L2'	LR: The norm of the penalty parameter
max_iter	100, 250, 500	LR: Maximum iterations for the solver to converge
n_estimators	100, 200, 300	RF and XGB: Number of trees in the forest
max_features	'sqrt', 'log2'	RF: Number of features to consider when looking for the best split
max_depth	10, 50, 100, None	RF: Maximum depth of the tree
learning_rate	0.01, 0.02, 0.05	XGB: Step size shrinkage used in update to prevent overfitting
gamma	0.0, 0.1, 0.2	XGB: Minimum loss reduction to make a further partition on a leaf node of the tree
Pre-Processing Parameters		
stopwords	True, False	Include or exclude stopwords
lemmatisation	True, False	Perform lemmatisation or not
numbers	True, False	Include or exclude numerical tokens

Table 2. Patent Refusal: Experiment 1 (2 d.p.) - Mean and std dev of 10-fold cross-validation

	N-Grams			**TF-IDF**			**Word2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	85.34 ±1.14	85.21 ±0.98	70.72 ±2.27	85.63 ±1.26	85.45 ±1.18	71.31 ±2.52	71.63 ±1.44	**71.53** ±1.77	43.29±2.88
LR	85.29 ±1.57	85.11 ±1.56	70.61 ±3.14	85.34 ±1.28	85.15 ±1.19	71.72 ±2.57	**71.66** ±1.26	71.49 ±1.57	**43.35** ±2.51
RF	84.79 ±0.83	84.74 ±0.95	69.60 ±1.66	85.00 ±1.31	84.85 ±1.31	70.03 ±2.61	67.22 ±0.99	65.10 ±1.20	34.71 ±2.00
XGB	**86.68** ±1.09	**86.58** ±1.16	**73.39** ±2.18	**86.88** ±0.85	**86.81** ±0.79	**73.78** ±1.72	70.06 ±1.31	69.60 ±1.61	40.15 ±2.60

	Patent2Vec			**Law2Vec**			**PatentDoc2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	**73.16** ±1.46	**73.17** ±1.59	**46.35** ±2.92	**70.25** ±1.72	**70.15** ±1.89	40.54 ±3.43	72.24 ±1.60	71.76 ±1.79	44.55 ±3.20
LR	73.07 ±1.82	73.03 ±1.99	46.18 ±3.63	69.75 ±1.34	69.54 ±1.45	39.52 ±2.68	73.14 ±1.71	72.90 ±1.70	46.30 ±3.42
RF	69.30 ±1.05	67.51 ±1.13	38.83 ±2.11	67.43±1.33	65.54 ±1.27	35.09 ±2.72	70.14 ±1.77	69.13 ±1.93	40.39 ±3.55
XGB	72.23 ±1.46	71.78 ±1.51	44.49 ±2.92	69.88 ±1.13	69.28 ±1.30	**49.83** ±2.30	**73.27** ±1.51	**72.91** ±1.61	**46.56** ±3.02

Table 3. Patent Refusal: Experiment 2 (2 d.p.) - Weighted Average 10-fold TimeSeriesSplit

	N-Grams			**TF-IDF**			**Word2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	83.66 ±3.55	84.76 ±5.30	66.61 ±6.97	82.88 ±4.04	84.08 ±5.52	65.02 ±7.77	68.46 ±2.13	**70.56** ±4.51	36.07 ±4.04
LR	83.78 ±3.61	84.90 ±5.12	66.88 ±7.12	83.02 ±3.88	84.23 ±5.51	**65.38** ±7.53	**68.52** ±2.27	70.45 ±4.64	**36.20** ±4.37
RF	82.21 ±3.51	82.81 ±7.99	63.86 ±6.43	81.99 ±3.80	82.57 ±8.05	63.35 ±6.84	64.49 ±2.67	61.07 ±8.42	28.33 ±4.23
XGB	**85.34** ±3.83	**86.11** ±6.41	**70.02** ±7.65	**84.15** ±3.20	**85.07** ±5.69	67.61 ±6.25	67.63 ±2.18	68.23 ±6.81	34.68 ±4.16

	Patent2Vec			**Law2Vec**			**PatentDoc2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	70.94 ±2.42	**73.45** ±5.31	40.55 ±4.41	64.73 ±3.14	67.29 ±5.57	28.19 ±5.34	**69.79** ±1.88	**71.66** ±4.68	**38.88** ±3.89
LR	**70.96** ±1.72	73.34 ±4.75	**40.57** ±2.90	67.08 ±1.60	**69.47** ±4.91	32.80 ±2.84	69.46 ±1.89	71.45 ±4.19	37.88 ±3.74
RF	65.01 ±2.07	63.49 ±7.97	30.74 ±3.57	63.63 ±2.51	60.72 ±9.11	28.97 ±4.11	66.92 ±2.27	65.84 ±9.62	34.02 ±3.53
XGB	68.83 ±2.37	69.82 ±6.23	36.88 ±4.67	**67.73** ±2.66	67.86 ±7.93	**35.10** ±4.45	69.72 ±2.38	70.29 ±7.40	38.72±4.32

Table 4. Opposition Division: Experiment 1 (2 d.p.) - Mean and std dev of 10-fold cross-validation

	N-Grams			**TF-IDF**			**Word2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	70.74 ±1.08	71.27 ±1.11	41.51 ±2.17	70.29 ±1.11	70.87 ±1.11	40.63 ±2.20	62.60 ±1.34	62.15 ±1.45	25.22 ±2.70
LR	70.86 ±1.68	71.30 ±1.67	41.74 ±3.36	70.55 ±0.91	71.05 ±1.01	41.13 ±1.82	62.90 ±1.55	62.43 ±1.94	25.82 ±3.10
RF	76.56 ±1.51	77.00 ±1.56	53.18 ±3.03	76.83 ±0.96	77.28 ±0.79	53.72 ±1.89	65.49 ±1.51	66.61 ±1.84	31.08 ±3.06
XGB	**79.90** ±0.94	**80.66** ±0.83	**60.01** ±1.84	**79.14** ±1.15	**79.77** ±1.14	**58.42** ±2.29	**66.50** ±1.41	**67.78** ±1.45	**33.13** ±2.85

	Patent2Vec			**Law2Vec**			**PatentDoc2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	62.76 ±2.25	62.28 ±2.68	25.54 ±4.48	61.90 ±1.25	61.14 ±1.39	23.81 ±2.49	62.77 ±1.78	63.07 ±1.53	25.57 ±3.55
LR	62.81 ±2.26	62.45 ±2.43	25.64 ±4.52	61.94 ±1.73	61.23 ±2.10	23.89 ±3.46	62.76 ±1.70	62.91 ±1.47	25.54 ±3.40
RF	66.21 ±1.57	66.55 ±1.57	32.44 ±3.14	**65.11** ±1.91	**65.90** ±2.03	**30.28** ±3.83	**69.25** ±1.32	71.54 ±0.98	39.02 ±2.54
XGB	**67.38** ±2.02	**68.07** ±2.16	**34.82** ±4.05	65.26 ±1.63	66.44 ±1.37	30.61 ±3.22	69.01 ±1.09	**70.89** ±0.74	**38.34** ±2.07

Table 5. Opposition Division: Experiment 2 (2 d.p.) - Weighted Average 10-fold TimeSeriesSplit

	N-Grams			**TF-IDF**			**Word2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	68.88 ±1.68	69.39 ±2.11	37.94 ±3.30	68.17 ±1.89	68.05 ±2.79	36.50 ±3.72	62.16 ±3.11	61.63 ±3.18	24.39 ±6.24
LR	68.79 ±1.98	69.10 ±2.25	37.71 ±3.93	67.96 ±1.96	68.01 ±2.26	36.01 ±3.91	62.46 ±2.60	61.77 ±3.00	25.01 ±5.20
RF	73.17 ±2.60	72.12 ±2.93	46.55 ±5.30	73.53 ±3.00	72.66 ±3.37	47.19 ±6.02	64.02 ±3.34	61.45 ±3.95	28.40 ±6.91
XGB	**76.74** ±3.29	**77.48** ±3.09	**53.65** ±6.53	**75.91** ±2.79	**76.72** ±2.99	**51.94** ±5.56	**65.10** ±3.45	**64.48** ±4.22	**30.30** ±6.95

	Patent2Vec			**Law2Vec**			**PatentDoc2Vec**		
	Acc	F1	MCC	Acc	F1	MCC	Acc	F1	MCC
SVM	62.01 ±3.14	61.90 ±3.55	24.06 ±6.25	59.26 ±2.92	58.04 ±2.72	18.60 ±5.92	61.27 ±3.04	61.35 ±3.35	22.55 ±6.08
LR	62.24 ±2.46	62.17 ±2.77	24.55 ±4.89	60.95 ±1.95	59.80 ±2.33	22.02 ±3.99	61.28 ±2.84	61.27 ±3.15	22.58 ±5.66
RF	64.04 ±3.92	62.11 ±5.09	28.30 ±7.81	62.57 ±3.33	60.37 ±4.35	25.38 ±6.79	67.35 ±3.29	66.14 ±4.55	34.82 ±6.57
XGB	**65.64** ±3.69	**65.65** ±4.52	**31.30** ±7.34	**63.22** ±3.06	**62.66** ±3.37	**26.52** ±6.10	**69.21** ±3.02	**69.30** ±3.49	**38.48** ±6.03

Table 6. Best models and their selected parameters

Board of Appeal	Experiment	Model	Model Hyperparameters	Input	Input Hyperparameters	Stopwords	Lemma	Numbers
Patent Refusal	1	XGBoost	n_estimators: 300 learning_rate: 0.05 gamma: 0.0	TF-IDF	use_idf: True norm: L2 ngram_range: (1,4) min_df: 2	False	False	False
Patent Refusal	2	XGBoost	n_estimators: 300 learining_rate: 0.05 gamma: 0.2	Bag of Words	use_idf: False norm: L2 ngram_range: (1,4) min_df: 2	False	False	False
Opposition Division	1	XGBoost	n_estimators: 300 learning_rate: 0.05 gamma: 0.2	Bag of Words	use_idf: False norm: None ngram_range: (1,3) min_df: 10	True	False	True
Opposition Division	2	XGBoost	n_estimators: 100 learning_rate: 0.05 gamma: 0.1	Bag of Words	use_idf: False norm: None ngram_range: (1,2) min_df: 2	True	False	True