14July, 15:30
Review on the first draft and discussed submission details.
Feedback included in the overleaf file

21July, 15:30
Discussing the results from the CNN
Talking about the first draft
Ways to interpret cnn results
Plots of auc, precision and recall
Compare results to other models

28July, 15:30
- Ensuring that Introduction and conclusion is coherent
- Reflect on citations
    - Add more
    - Automate on latex
- Make sure use of abbreviations is consistent
- Compare train and test results
- Methodology
    - Include the use of validation set
- Conclussion
    - Other possible approchaes
    - Prior research on binary classification?
- Reflect on f1-scores
- Conslusion
    - Key insights
    - Limitations
    - Future researchs

INCLUDE
- CODE
- REPOSITORY
- REFERENCES

2July, 17:00
Binary classification
Discussion on the best way to address the motivation and research objective
Outline for next steps

25June, 17:00
Doc2vec and tfidf result interpretations
Ways to improve them
Plot required curves
Next steps

18 June, 16:00

- SVM Model
    - Cross validation
    - Class_weight https://scikit-learn.org/stable/modules/generated/sklearn.linear_model.SGDClassifier.html
    - Measure: consider precision instead of accuracy?
- Literature Review

Notes

Literature Review

- Patent Data
    - Patents as an indicator for innovation / Climate change mitigation technologies
    - Patent classification
- ML with patent data
    - Natural language processing
    - Technology classification
- Research gap

Models

- Check Y02 instead of Y labels
- Report results for cross validation
- Include confusion matrix

Next steps

- Compare performance of abstract, abstract and title, abstract/title/claims
- Run different algorithms
- Grid search / hyperparameter tuning
- Vectorization

11 June, 13:00

- Revised Table of contents
- Methodology section
- CPC for design rights
- TFIDF Vectors
- SVM model for binary classification

Notes:

- https://onlinelibrary.wiley.com/doi/full/10.1002/smj.2699

7 June, 16:00

- Description for each of the methods used in preprocessing?
    - Lowercase
    - Stop words
    - Punctuation
    - Apostrophe
    - Single Characters
    - Stemming/lemmatization
- Separate code for preprocessing for each of the steps?
- Is it okay to use TFIDF after stemming? Or should I try Bag of words?
- Package for TFIDF? Gensim, Spacey or Sklearn?
- Check CPC code
- **Revise** Table of Contents
- Dropping variables?
    - Null values in Claims and Abstract
- Variable names

26 May, 15:40

- Draft for Table of Contents
- Target variable
- Methodology section
- Literature Review
- Update Minutes section
- CCMT Papers:
- https://www.merit.unu.edu/publications/working-papers/abstract/?id=8718
- https://www.researchsquare.com/article/rs-266803/v1
- https://www.hbs.edu/faculty/Pages/item.aspx?num=59272&dm_i=53XH,D4XW,3Q42GG,1I4LY,1
- https://www.journals.uchicago.edu/doi/10.1093/reep/req023

Can I consider publication_type?
Check legal_status

19 May, 15.30 BST
Overleaf draft: https://www.overleaf.com/1824625374hjvtxmvhftyf

- Have a final dataset
- Methodolody
- Preprocessing
- Final columns for the dataset
- Drop jurisdiction column

Notes:

- Focus on Title, date, lens_id, abstract, claims
- Number of rows = number of patents
- Check the subclasses
- Focus on binary classification

- Draft on table of contents
    - Subsections?
- More on Literature Review

13 May, 9am GMT

- Added papers to Mendeley

Patent features:

- As discussed in the previous meeting, main focus on **title** and **abstract**
    - D'hondt and Verberne (2010) showed that the title and abstract sections are more informative than the full-text representation of the patent document for patent retrieval.
- Multi-label classification
    - Y tags
    - Looking into other possible labels
- Can I find similarity between patents? Using a particular measure?
- Sub-classes?
- Do we have patents that are not a part of Y tags?

https://worldwide.espacenet.com/classification?locale=en_EP#!/CPC=Y02

Notes:

- Y tags as labels
- http://ir.nsfc.gov.cn/paperDownload/ZD4277723.pdf

Literature Review

- https://www.scopus.com/search/form.uri?display=basic
- https://scholar.google.com/

Follow up

- 3-5 papers
    - Research question
    - Data, features and labels
    - Methodology
        - Processing
        - Classification algorithms
        - Evaluation and validation
- Initial research questions
- Initial methodology including high-level steps
    - Data
    - Patent Features
    - labels

5 May, 2pm GMT

Agenda

- Proposal included in the overleaf link above.

Notes

Project Management

- Minutes
- Trello.com
- Mendeley: mhce2@cam.ac.uk

Lens API

- https://docs.api.lens.org/response-patent.html

Patent features for NLP

- Title
- Abstract
- Claims
- Descriptions

Potential features

- Text
- Classification tags
- Citations

Next meeting

- Narrow down the patent dataset
- Label for climate change mitigation technology. Y-tags?
- Patent features

8 April, 2pm GMT

Agenda

- Logistics
- Patent databases
- Literature

Notes

Logistics

- Project management, Minutes document, JIRA TBC
- Code in faculty: uceimhc@ucl.ac.uk
- Mendeley Reference Manager
- Dissertation Template: https://www.overleaf.com/1964469159tgtdvhrdbdtf

Patent databases:

- Overview: https://wipo-analytics.github.io/databases.html
- Lens API: https://www.lens.org/
- PATSTAT: https://www.epo.org/searching-for-patents/business/patstat.html
- Patentsview API, USPTO: https://patentsview.org/, https://patentsview.org/apis/api-endpoints/patents
- Google Patens

Y02 class for Climate Change Mitigation Technologies (CCMT):

- Y02 section: https://www.uspto.gov/web/patents/classification/cpc/html/cpc-Y.html#Y02
- https://www.mendeley.com/catalogue/9834871b-dd44-33e4-846a-6ceba3fd0b89/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7Bb0563c53-3fed-4547-a657-e9b8203dff53%7D
- https://www.mendeley.com/catalogue/dd509e21-3585-3377-be2b-1b5fe390a44d/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7B497525ad-7e23-49fe-bbc4-74e973618557%7D

NLP for patent classification: (Added to Mendeley)

- https://www.emerald.com/insight/content/doi/10.1108/ECAM-09-2019-0480/full/html?skipTracking=true
- https://services.google.com/fh/files/blogs/bert_for_patents_white_paper.pdf
- https://arxiv.org/pdf/1906.02124.pdf
- https://www.sciencedirect.com/science/article/pii/S0048733320302195
- https://www.mendeley.com/catalogue/7d71e944-9977-3e62-a246-39cb6cf19f5d/?utm_source=desktop&utm_medium=1.19.8&utm_campaign=open_catalog&userDocumentId=%7B7bf359d0-f50f-420d-965f-2b0320bba648%7D

Academic Literature:

- https://www.scopus.com/search/form.uri?display=basic#basic
- https://scholar.google.com/

Preparation for the next meeting:

Proposal:

- Objective
- Methodology
- Expected outcome