IP Enforcer: A Data Science Approach to Identify Potential Patent Infringers

Harsh Vardhan Singh Kalra, Harish Reddy Gavva, Sandeep Mukhopadhyay, Vinay Krishna Devulapalli, Matthew A. Lanham

Purdue University, Krannert School of Management

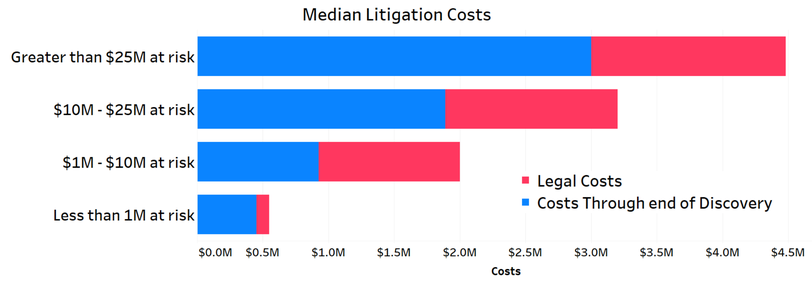
hkalra@purdue.edu; hgavva@purdue.edu; mukhopa6@purdue.edu; vdevulap@purdue.edu; lanhamm@purdue.edu

ABSTRACT

Patents are an essential feature of innovation and commerce. This research was conducted to develop a tool which facilitates an automated detection of potential patent infringers. The motivation behind this tool is to decrease the overall time and cost spent by a business on identifying potential Intellectual Property Rights infringers. The method incorporated in this tool concatenates patent claims to each other. Thereafter, BERT is trained on this data to identify the level of similarity between two patent claims. In addition, we make use of a similarity score, assigned to each compared patent claim, to inform the user of the level of similarity between two patents.

INTRODUCTION

According to the American Intellectual Property Law Association, the median litigation cost for patent infringement suits constitutes a jaw-dropping 65% of the claims which are less than a million dollars. Due to these increasing costs, many companies have started looking into viable alternatives like analytics and machine learning. Our project deals with the idea of using NLP techniques to gather focus on the key areas of claims and descriptions and thereby compute a similarity score between the patents of the plaintiff and the defendant to ascertain the possibility of infringement.



Research Questions:

- What are the most important components of a patent application to be considered while examining an infringement?
- What is the appropriate methodology to score similarities between two patents?

LITERATURE REVIEW

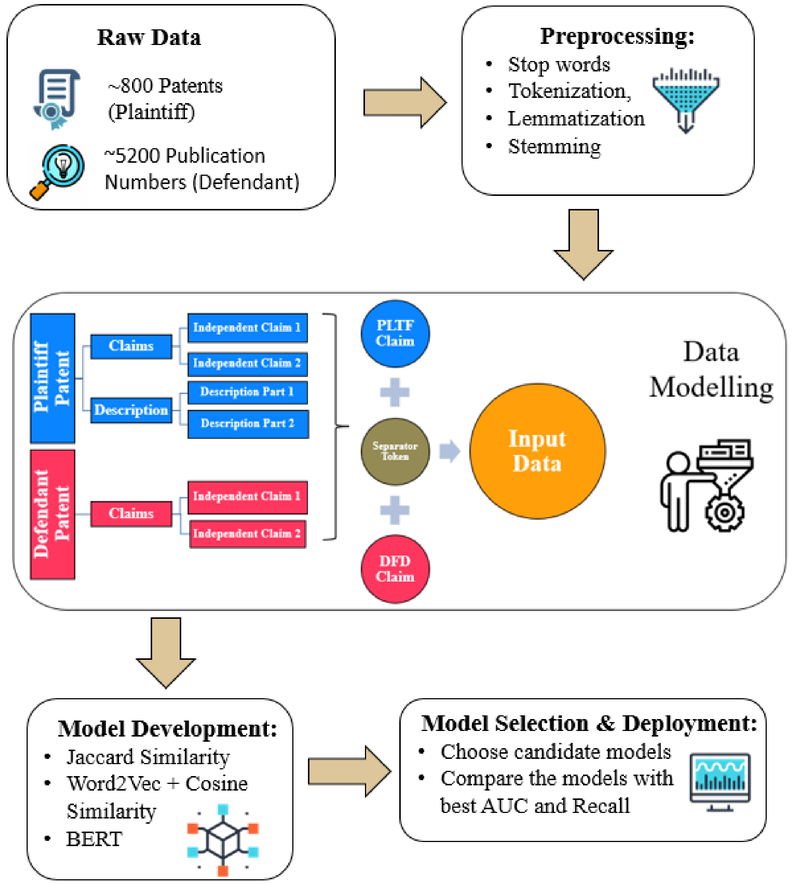
***Freunek M. & Bodmer A.* (2021)**

- Proposed a method to concatenate the claims and descriptions within the patent applications to train BERT
- Scoring according to count of relevant labels or sigmoid of relevant labels.

Ambedkar et al. (2007)

- Developed a method to ascertain similarity between two claims based on:
  - Simple lexical matching
  - Knowledge-based semantic matching

METHODOLOGY



STATISTICAL RESULTS

We compared different models based on the Recall and the AUC score and found that BERT was generating significantly better results in our experiments.

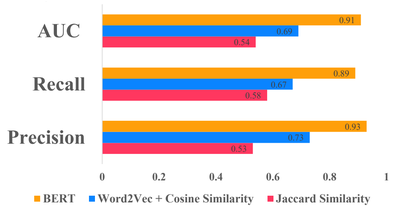


Fig 1. Model Results

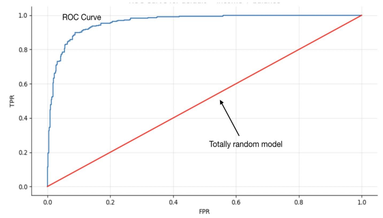


Fig 2. ROC curve for BERT

EXPECTED BUSINESS IMPACT

Mid to small companies focused on innovation do not have the resources to monitor all the patents that could suffer infringement. The tool presented herein has the potential to enhance search efficacy while reducing the time and effort involved. This project has a financial impact by directing the employee time and effort towards evaluating possible infringing instead of looking at all patents. Moreover, if a business decides to enter litigation, this tool can add value by helping identify relevant materials to share with domain experts. For instance, an expert can quickly identify relevant excerpts from large amounts of textual data based on the similarity score. Overall:

- A major improvement over the traditional machine learning models (AUC 0.6 – 0.7) is observed with BERT (AUC 0.91).
- With higher accuracy and the ability to identify relevant material, the tool is expected to reduce the labor effort by 70% and eliminate patent search costs.

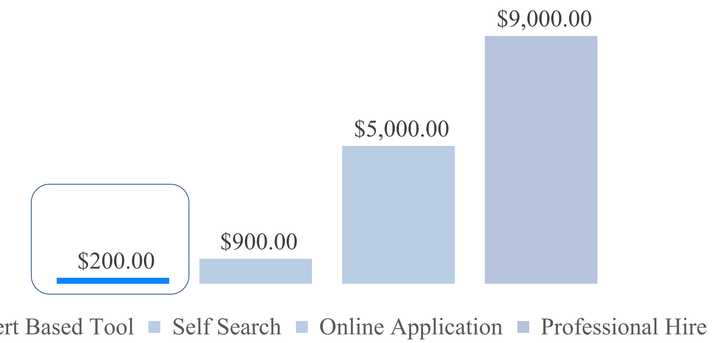


Fig 3. Approximated Patent search cost through different mediums.

CONCLUSION

This project presents a focused study on patent similarity and proposes a deep learning-based application (BERT) for this problem. Each possible combination of defendant and plaintiff claims was generated as input for the BERT. Based on the output, a similarity score was created to pre-empt the possibility of infringement. Finally, extensive experimental results on real-world data proved that BERT could help businesses identify patent infringement more precisely. We hope this work leads to further studies.

ACKNOWLEDGEMENTS

We would like to thank Professor Matthew Lanham and our corporate partner for their guidance and support on this project