

Can Data Science Technologies Predict Insurance Claim Lifespan?

Sheikh Hyder Ali
Supervised by Paul Barry

Introduction

- Unum, a leading insurance company is looking out for ways to automate medical claim lifespan prediction.
- An average claim settlement period ranges between 15 - 30 days.
- There is a yearly 10% increase in the number of claims received.
- An effective prediction system will help the company in better resource allocation & utilization.

Research Objective

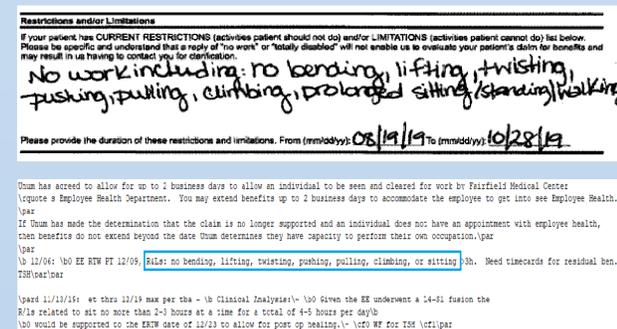
- This study will try building a machine learning model capable of discerning key important details from medical prescription notes.
- The model will try predicting claim lifespan with $\pm 10\%$ accuracy score.

Literature Review

- Natural Language Processing (NLP) is a branch of Data Science that expertise in imparting language understanding to machines.
- Clinical NLP has experienced a surge in its research base post advancements in the Data Science domain.
- Most of the previous works have focused on improving the medical knowledge of the models by using Named Entity Recognition (NER) techniques.
- Sentiment analysis of clinical texts has further helped in understanding the diagnosis procedures.
- This work will try to implement NLP in a way to predict the claim lifespan by deriving key important features from clinical texts.

The Data

- Anonymised prescription notes converted into text files have been made available by Unum for this study.
- These are past claim records along with their respective details.



Methodology

- The text data is in RTF and it has a variety of inconsistencies in it.
- We will begin by cleaning the data to reduce the noises present.
- Cleaned data will then be Tokenized, Lemmatized to proceed with NER.

- Recurrent Neural Networks (RNN) will be used to train the model based on the data.
- Model performance will be analysed to make further changes in it.

Technologies



Future Work

- Further works can be carried out to read images digitally, eliminating the manual process of text conversion.

References:

1. Nadkarni PM, Ohno-Machado L, Chapman WW. Natural language processing: an introduction. J Am Med Inform Assoc. 2011;18(5):544-551. doi:10.1136/amiajnl-2011-000464
2. Wang, Y., Wang, L., Rastegar-Mojarad, M., Moon, S., Shen, F., Afzal, N., Liu, S., Zeng, Y., Mehrabi, S., Sohn, S., & Liu, H. (2018). Clinical information extraction applications: A literature review. *Journal of biomedical informatics*, 77, 34-49.
3. Yang, H., Spasic, I., Keane, J. A., & Nenadic, G. (2009). A text mining approach to the prediction of disease status from clinical discharge summaries. *JAMIA*, 16(4), 596-600.