



PHYSIOTHERAPY VS CHIROPRACTIC

A Subreddit Study

AGENDA

01

PROBLEM
STATEMENT

02

BACKGROUND

03

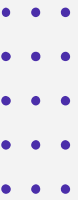
METHODOLOGY

04

RESULTS

05

CONCLUSION

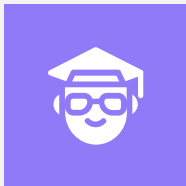


01

PROBLEM STATEMENT

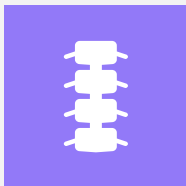


PROBLEM STATEMENT



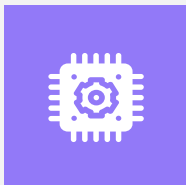
WHO

Prospective undergraduates choosing a college major



WHAT

- Understand, compare and contrast: physio vs chiro
- Which is a better fit for their interests?



HOW

Natural Language Processing (NLP) on Subreddit threads



BACKGROUND




Are they the same?



Source: <https://www.ansellchiropractic.com.au/wp-content/uploads/2017/04/Physiotherapy-1.jpg>



COMPARISON

	Treatment	Focus	Patient Involvement	Minimum Education	Scope*
PHYSIO	Non-invasive treatment of Back pain, Sports injuries, etc	Movement of joints and muscles	Active	BSc, MSc in Physiotherapy	Hospitals, private clinics
CHIRO		Spinal Manipulation	Passive	Doctor of Chiropractic (Degree)	Private Clinics

*Singapore context. Reference: <https://physioclinic.sg/physiotherapy-or-chiropractor-in-singapore/>



03

METHODOLOGY



METHODOLOGY

WEB-SCRAPING

DATA CLEANING

PREPROCESSING

MODELING

VISUALIZATION OF
RESULTS



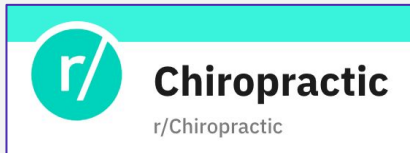
WEBCRAPING & DATA CLEANING



- Pushshift API
- 1500 submissions each
- No comments



DATA
CLEANING



1417 posts

51.8%

Dropped:

- Duplicates
- [removed], [deleted]

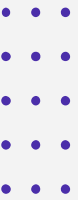
Merged:

- Selftext & Title

48.2%

1325 posts

PREPROCESSING



NO HTML
CODE



NO STOP
WORDS



LOWER
CASE



NO URL
LINKS

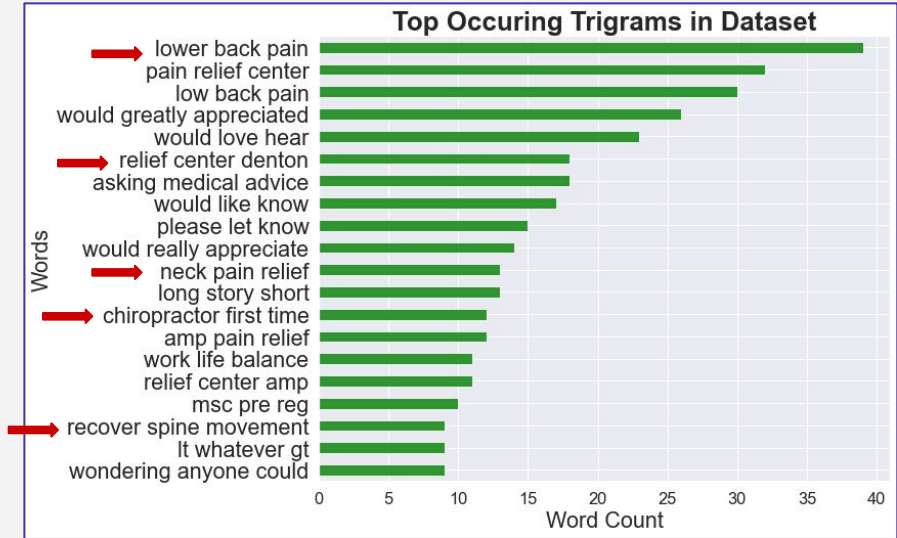
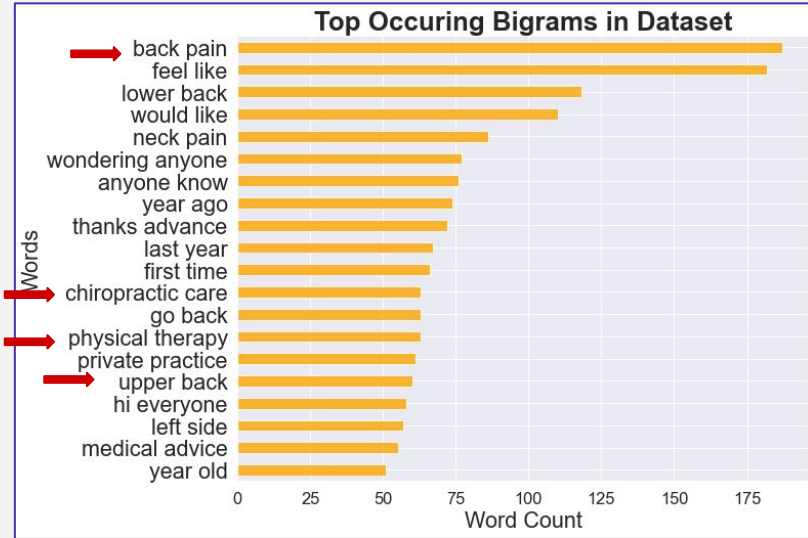


LETTERS
ONLY



LEMMATI
-ZED

Most frequent words in combined dataset



It appears that the general theme across the two subreddits is "pain relief" for "lower back pain"



MODELLING

**COUNT
VECTORIZER**

**LOGISTIC
REGRESSION**

**COUNT
VECTORIZER**

**NAIVE
BAYES**

**COUNT
VECTORIZER**

**RANDOM
FOREST**

**TFIDF
VECTORIZER**

**LOGISTIC
REGRESSION**

**TFIDF
VECTORIZER**

**NAIVE
BAYES**



- Test size: 0.25
- GridSearch CV, 5-folds cross validation





04

RESULTS



RESULTS



BEST MODEL:

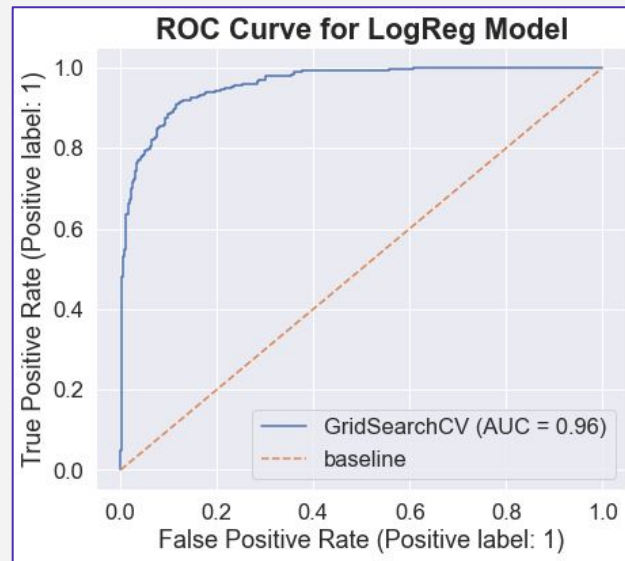
- Transformer: Count Vectorizer
- Estimator: Logistic Regression

BEST PARAMETERS:

`CountVectorizer(max_df=0.85,
max_features=2000, ngram_range=(1, 3))`

BEST METRICS:

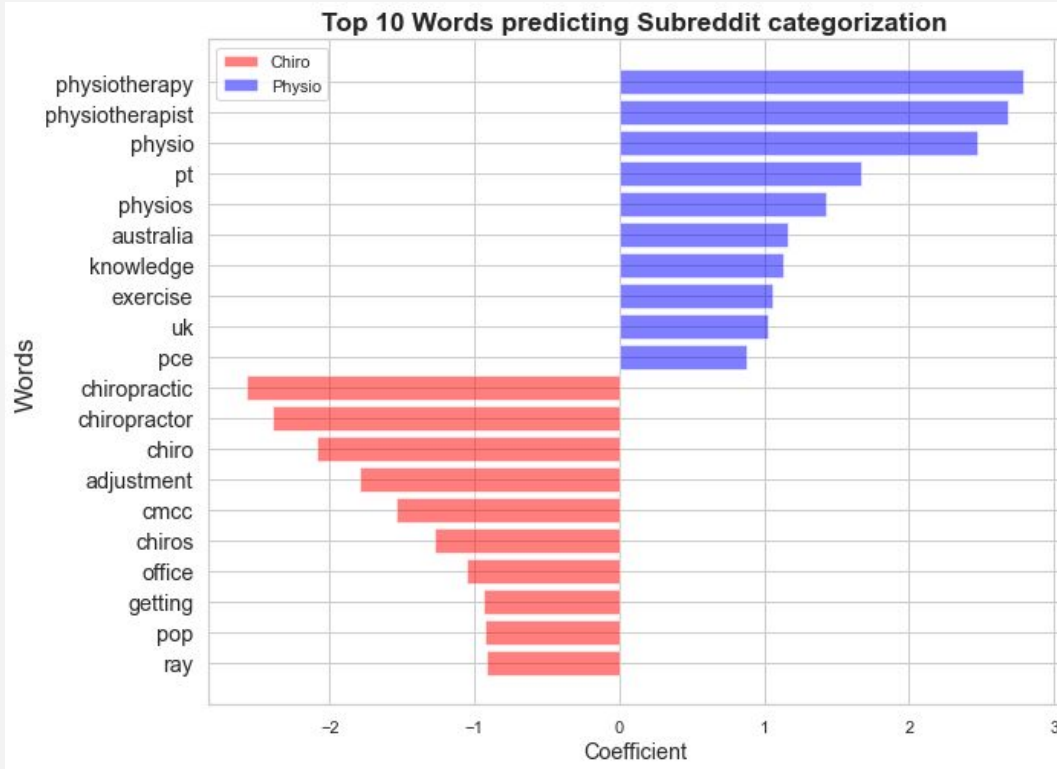
Accuracy = 89.2%
Specificity = 88.5%
Sensitivity = 89.9%
Precision = 89.3%



ROC AUC Score: 0.9584



Top 10 Words Predicting Classification



- Top predictors: variations of the word "physio" or "chiro".
- 'Exercise' is significant in predicting Physio but not Chiro.
- Conversely, 'adjustment' is significant in predicting Chiro.
- PCE: Physiotherapy Competency Examination.
- CMCC: Canadian Memorial Chiropractic College





05

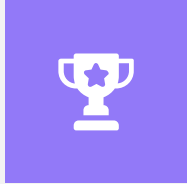
CONCLUSION



CONCLUSION

1

Logistic Regression
coupled with Count
Vectorizer



BEST MODEL

2

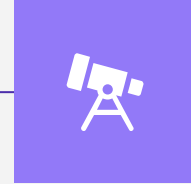
- Differences in approach
 - Education requirements/options



INSIGHTS

3

Further training on more
data to improve model
accuracy



FUTURE



THANKS!

CREDITS: This presentation template was created by **Slidesgo**, including icons by **Flaticon**, infographics & images by **Freepik**

