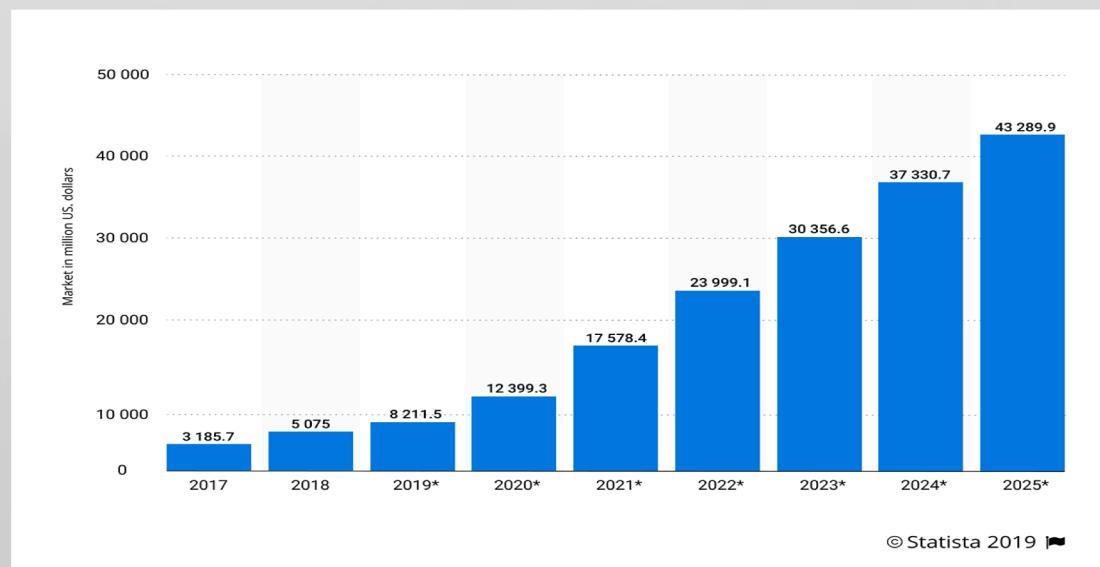
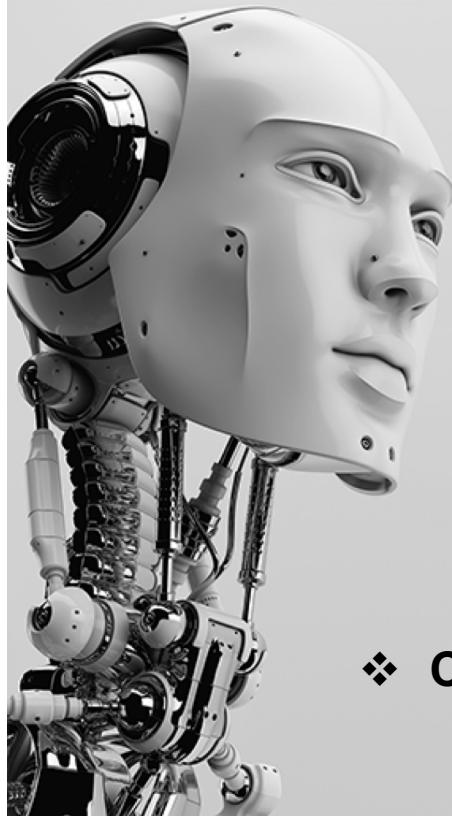


Progressive Clinical Documentation Analysis for Kaiser Permanente using NLP

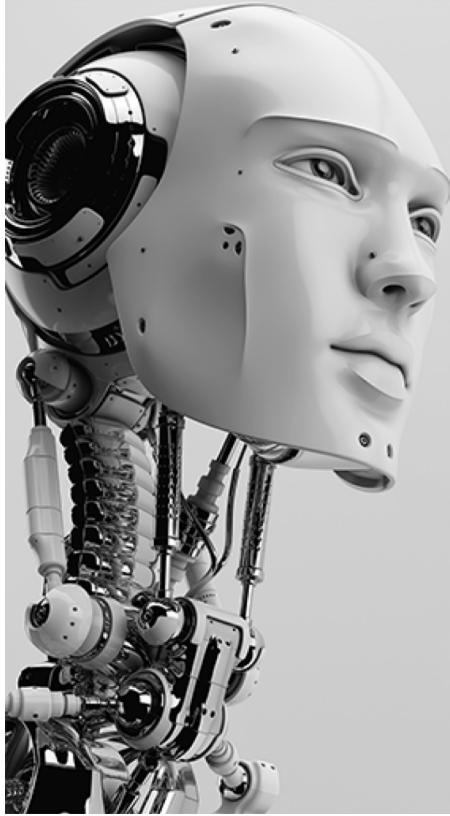
KISHORE

Clinical documents

- Natural Language Processing poses some existing opportunities across various industries.
- One of the quickest evolving AI technologies today is NLP.



- ❖ Clinical documents from Kaiser Permanente Healthcare Center



Preprocessing

- 1 • Lower case
- 2 • URL address
- 3 • Punctuations
- 4 • Stop words
- 5 • Stemming

BOW and TF-IDF

Term Frequency (TF)

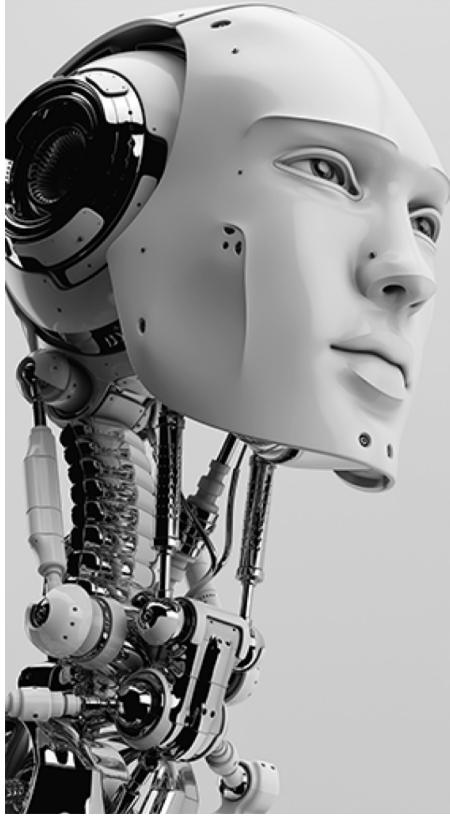
$$tf_{t,d} = \frac{n_{t,d}}{\text{Number of terms in the document}}$$

Inverse Document Frequency (IDF)

$$idf_t = \log \frac{\text{number of documents}}{\text{number of documents with term 't'}}$$

TF-IDF

$$(tf_idf)_{t,d} = tf_{t,d} * idf_t$$



Algorithms

KNN

Logistic Regression

TF-IDF Logistic Regression

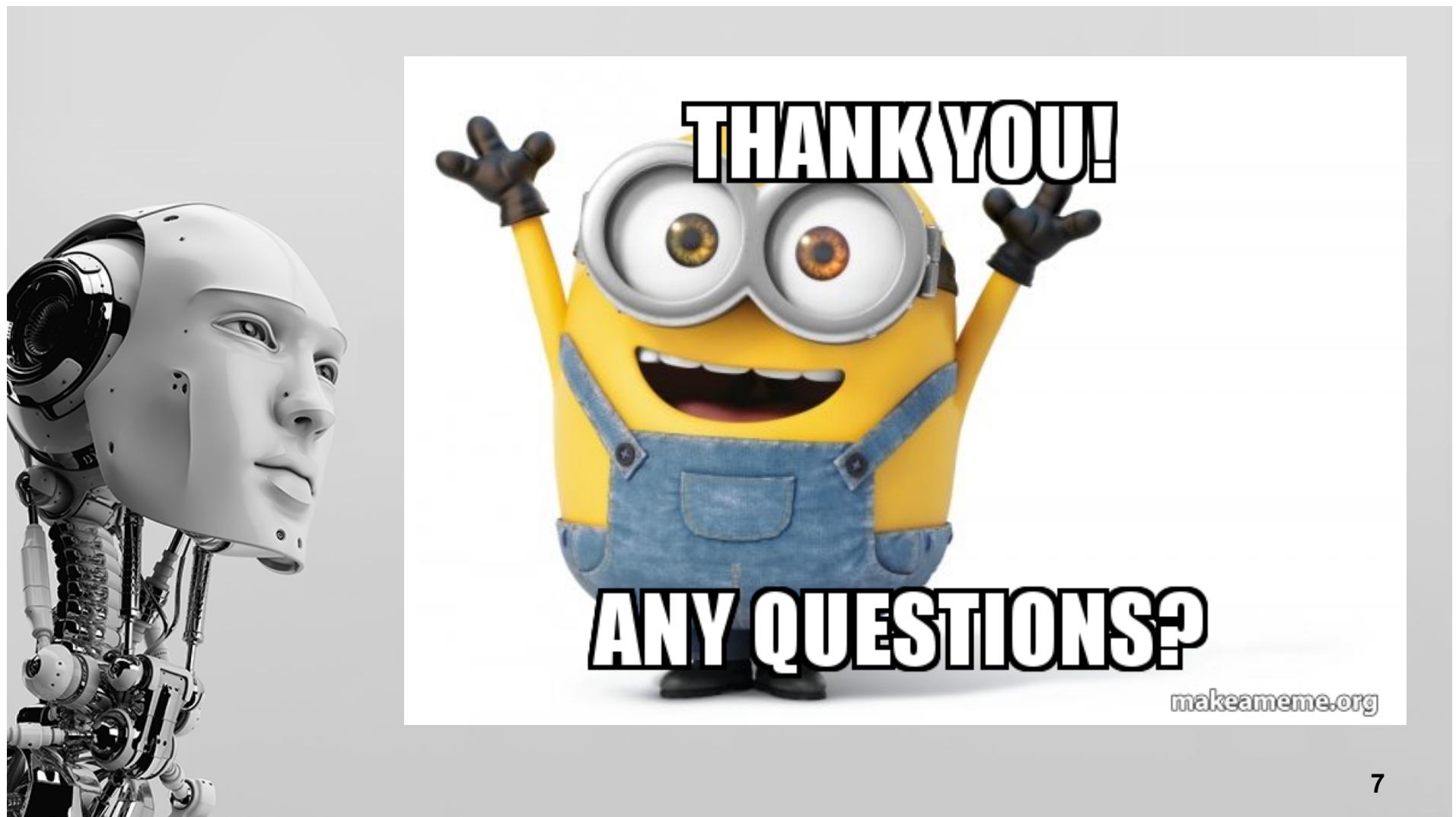


Results and Conclusions

Algorithm	Train data accuracy (%)	Test data accuracy (%)
KNN (K=20)	97.09	50.18
Logistic Regression	94.18	52.69
IF-IDF LogReg	60.54	50.18

Conclusions:

- ❖ NLP provides better understanding, aiding in clinical trial matching and clinical decision support for Kaiser Permanente healthcare centers.



Appendix

Logistic Regression cost function plot

