

Case 1 – Assignment

Student Name: Fang Zhang
CWID: 10433041

1. How I approach the problem?

To find the hidden patterns of SDB(Sleep-disordered breathing) in articles, I used the pdf files in pdf chest journal folder as basis of my analysis.

Firstly, I need to make sure that all the articles in folder are relevant to the SDB topic so I could find the correlations between each other.

Secondly, I need to find out what contents related to SDB people are talking about in these articles. To group these contents into potentially topics, I built LDA model (topic modeling) to find similar topics in these articles.

Thirdly, from the model result, I tried to interpret what articles may bring about by visualizing the keywords in each topic. Combined with current knowledge in the field, I got the result.

2. What tools or methods were used?

I used Python for whole analysis. To crawling the PDF files, I used pdfminer package. To build LDA model, I used scikit-learn which is a famous machine learning tool and generated the word cloud for interpretation.

My Assignment code reference: https://github.com/fzhang22/BIA686-Practicum-in-Analytics/blob/master/CASE1/Case1_text_mining.ipynb

3. Summary of my finding



From the word cloud map, some key words are popped up. Combined with the knowledge field, I suggest we should establish hypothesis with several aspects: 1.

Children snoring may heavily related to SDB. 2. Osa(obstructive sleep apnea) that

airway may correlate to SDB. 3. Obesity and high BMI index maybe connect to ahi which refers to *apnea hypopnea index*.