**STATS 517 Final Presentation Review**

**Reviewer:** Dalyn McCauley

**Reviewee:** Franklin L. Sturgeon

**Project:** LDA Topic clustering on national miner health and safety data

**Coverage of 3 course areas – Supervised/Cluster/Association:**

Luke covered all three course concepts in his final presentation. Though he did not explicitly cover the clustering and association concepts introduced in this class, he utilized a similar but more complicated technique called LDA that was better suited for his dataset. From the looks of it, the LDA was able to cluster documents based on words that are associated with fatigue. Luke showed a table in his presentation that showed the probability of a word showing up in a given document. This is very similar to concept of support in itemsets, but even seems much more complicated. Luke also incorporated many supervised regression models.

**Dataset size needs stated criterion:**

The original dataset used by Luke had a size on the order of several millions. Even with his reduction, the dataset still had a size greater than 500,000, which far surpasses the requirements for this project.

**Relevancy:**

I think Luke’s project had the most real world relevancy of the projects presented in this class. His research question directly affects the health and well-being of miners, one of the most deadly industries in America. The work Luke put into this project could be utilized by the CDC to implement real policies to protect miners from fatigue related injuries.

**Difficulty:**

On a scale of 0 – 10, I give this project a difficulty of 10. The majority of the work done in this project seems to be self-taught, because we did not cover concepts of document categorization modeling or LDA in this class. The dataset was also massive, which must have taken a lot of preprocessing and a program with efficient memory allocation.

**Interestingness:**

Luke’s presentation definitely kept me engaged. The work he did seems to be novel, and could have great impacts if implemented by the CDC.

**Clarity:**

The presentation was very clear. Luke explained all variables and concepts thoroughly, and was able to answer all the question asked of him.

**Originality:**

This presentation was very original. It was great to see a different application of the concepts we learned in this class.

**Creativity:**

Luke was creative in his presentation, and chose a project topic that was beyond the scope of concepts learned in this class.