Slide 4 – My name is An Nguyen, I am currently living in Grand Prairie Texas. I have a been a biologist over the last 6 years and am currently working for a Medical Device company. Am I a Data Scientist?

Slide 21 – Thanks Jodi. To visualize the 512 features vector we transformed each vector using PCA. (10sec)

Slide 22 - The untransformed vectors were combined into one dataset for K-Means unsupervised clustering. We were able to find 4 distinct groupings, which is inaccurate.

Slide 23 – So we applied supervised clustering with Data Scientists labeled and all other jobs labeled as Non-Data Scientist. But the results were also inconclusive (50% accuracy).

Slide 24 – Similarly, we then separately label each of the 6 job descriptions but was not able to locate Data Engineer.

Slide 25 – To improve our accuracy percentage we chose to employ Neural Networks to classify each job posting. And was able to achieve a 92% accuracy in classifying a job posting as a Data Scientist or not.

Slide 26 – We then applied all six labels to the neural network, 85% accuracy.

Slide 27 - We observed that there is a large overlap between Data Analyst and Statistician with Data Scientist job postings. (10 sec)