MAKE THES CHANGES:

In the data section, change the first word to “The” instead of “TThe”

In the data section, add paragraph breaks per original document

Replace ABSTRACT section with the following

Add section D about association rules to the PROPOSED SOLUTION SECTION

Add the section immediately following that below to the end of the results section

Add the CONCLUSION section included below

ABSTRACT

Parkinson’s Disease is a neurodegenerative disorder of dopamine systems. It can be treated but the effectiveness of pharmacological options faces diminishing returns and anything that might reduce the occurrence of overmedication would have strong positive impact. [PREVIOUS PAPER] collected data comparing features of speech in 52 individuals diagnosed with Parkinson’s and analyzed it using deterministic linear regression. In this work, we used various Data Mining tools to prepare and analyze the data speculatively in order to discover interesting information for use in Machine Learning or other subsequent analysis. K-means and K-means++ demonstrated 93.62 and 95.57 respectively while OPTICS showed 96.62. DBSCAN was not viable due to density and resulted in 100 indicating overfitting. Validation of clustering was performed using [METHOD FOR VALIDATION]. Classification after Data Mining was performed using SVM [SPELL THAT OUT]. We also approached the problem as an Association Rule problem and discovered nothing significant though we believe further analysis is warranted.

Key Words: Data Mining, k-means, k++, OPTICS, DBSCAN, Parkinson’s Disease, SVM

D. Frequent Itemsets and Association Rules

For the analysis of the data as Association Rules, first the data needed to be re-imagined. After normalization, the data was put through a second encoding process. Three methods for encoding were used. The first was Equal Size indicating equal splits of 0.2 in the normalized date. The second was Bell Curve Like meaning the ratios more closely represented standard deviations. The third was based on how extreme the data was. Anything above or below a small threshold was considered *outlier* and anything outside of one standard deviation *different*. This was calculated based on magnitude. The data was encoded separately from the source of truth. Different combinations of the two, as well as various threshold values were performed.

RESULTS of the ASSOCIATION RULES

The Association Rules analysis did not provide any useful results though it is plausible that more analysis could yield potentially valuable results. We recommend further analysis.

CONCLUSION

We conclude that there is value in performing analysis using Data Mining techniques. We believe that a larger data set would provide valuable insight into individualized treatment. Additional clustering analysis should focus on OPTICS and related algorithms.

We also believe that the novel approach of re-imaging the data as transactions and using an Association Rule approach should take priority over clustering techniques. Using that framework, additional analysis could be performed by considering it from a longitudinal perspective where time trial x0 relates to time trial x1.