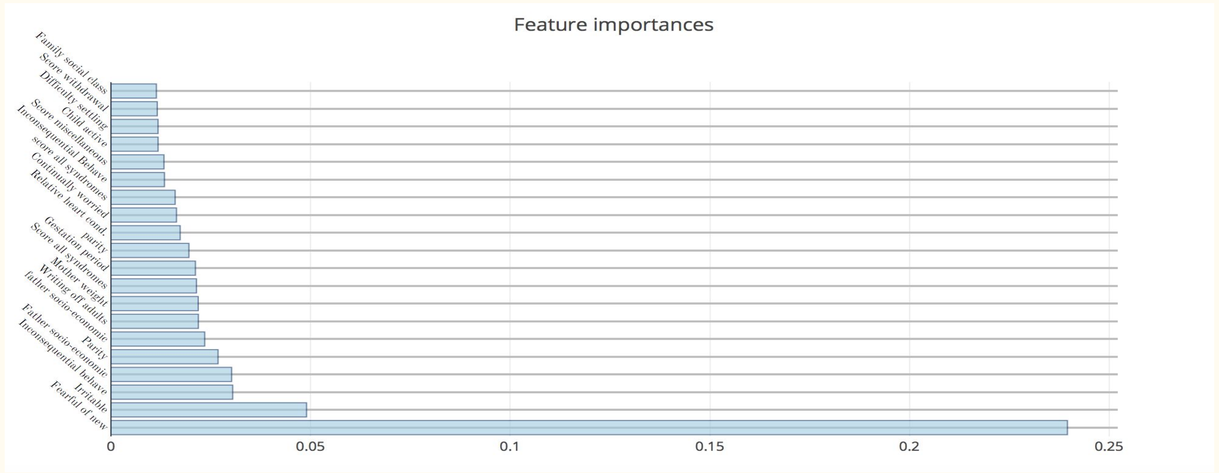
Early Childhood Seizure and Relapse in Adolescence

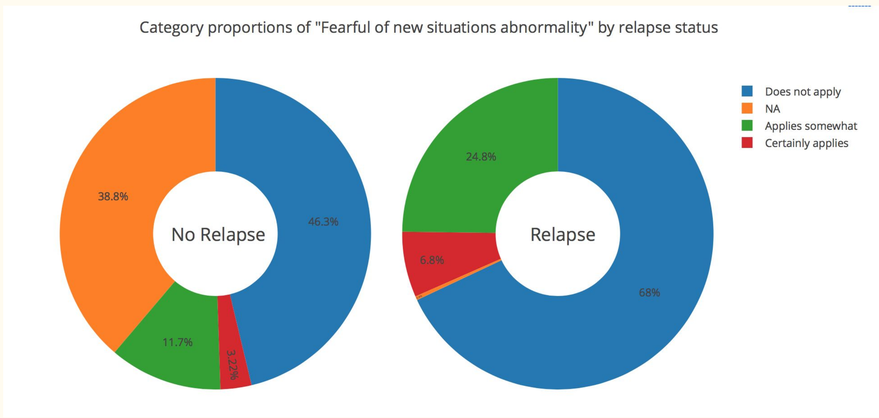
Fernando Becerra, Jessica Cao, Yi Ding, Andrew Mayo

Introduction

As one of the most far-ranging and devastating neurologic diseases, epilepsy afflicts over 50 million individuals worldwide. The most diasabling aspects of seizure is its unpredictability, the impairment of consciousness can be life-threatening, and the embarrassment of incidences can lead to social isolation and even misbehavior.

Therefore, for our CS109a project, we decided to study seizures and our ability to predict them. Our goal has been to build a model that can determine the factors that affect epilepsy recurrences and predicting future seizure incidences given the individual's identity, medical history, social behavior, etc.





Methods

Dataset: Sweep 0-3 (Age 0-16) of 1958 National Child Development Study, which tracks every child born in the UK during a single week in 1958. 1026 out of 18558 children developed syndrome like epilepsy and convulsion before age 11. 250 of them relapsed by age 16.

Preprocessing: Based on subject matter knowledge, we extract 207 social economic status and medical related variables. After dropping columns with missing rate higher than 50%, 188 variables remain for analysis.

Model building:

1. Logistic regression, LDA, decision tree, random forest and support vector machine were tried roughly to get a broad view of the performance for each model. Predictions were based on majority vote.
2. Decision tree outperformed other methods. We further tuned the model via cross validation to improve its performance.
3. We created a balanced dataset to fit model and predict seizure relapse.

Results & Discussion

Using a balanced sample and majority vote strategy, we came up with a decision tree-based model with a test accuracy of 0.734 for non-relapse group and a test accuracy of 0.552 for relapse group. The model suggested that the most important features to predict recurrence of seizure are parents' social class, neonatal birth condition, mental condition and social behavior issue, these results are adherent to various existing literatures.

The 1958 NCDS dataset is a social survey-based longitudinal dataset, which imposes us two challenging problems by nature. 1) The first is that, a large amount of missing value existed in the dataset. We assign missing value in categorical data as 'missing', and that in continuous data as mean of the non-missing ones. We may improve the data analysis by using more complicated method such as multiple imputation. But as long as the data is not missing at random, bias would still haunt our results. 2) The second is that although a number of 18558 children were followed up in the longitudinal study for almost 20 years, there is a limited amount of useful information in terms of seizure prediction. This is the nature of the complication of this project and could be addressed by designing a study targeted specifically at epilepsy analysis.

Conclusion

We have presented a seizure prediction models for the 1958 National Child Development Study. The model fit 100 decision trees in the train set to predict seizure in test set and take the majority as results. This model yielded a test accuracy of 0.734 for non-relapse group and a test accuracy of 0.552 in relapse group.

Acknowledgments & References

Thank you to the CS109a staff for providing us with an excellent class, engaging material, and wonderful support. Thank you especially to Kela Roberts, our project TF who provided us with superb feedback throughout our project.

References

Kariuki, S., Newton, C., Prince, M. J., & Das-Munshi, J. (2016). The association between childhood seizures and later childhood emotional/ behavioral problems: findings from a nationally representative birth cohort. Psychosomatic Medicine. 10.1097/PSY.0000000000000305

Hauser, W.A. and Beghi, E. (2008) First seizure definitions and worldwide incidence and mortality. Epilepsia, 49, 8-12. doi:10,1111/j.1528-1167.2008.01443.x

Carney, Paul R., Stephen Myers, and James D.Geyer. "Seizure Prediction: Methods." *Epilepsy & behavior*: E&B 22.Suppl 1 (2011): S94-S101. *PMC*. Web. 8 Dec.2016