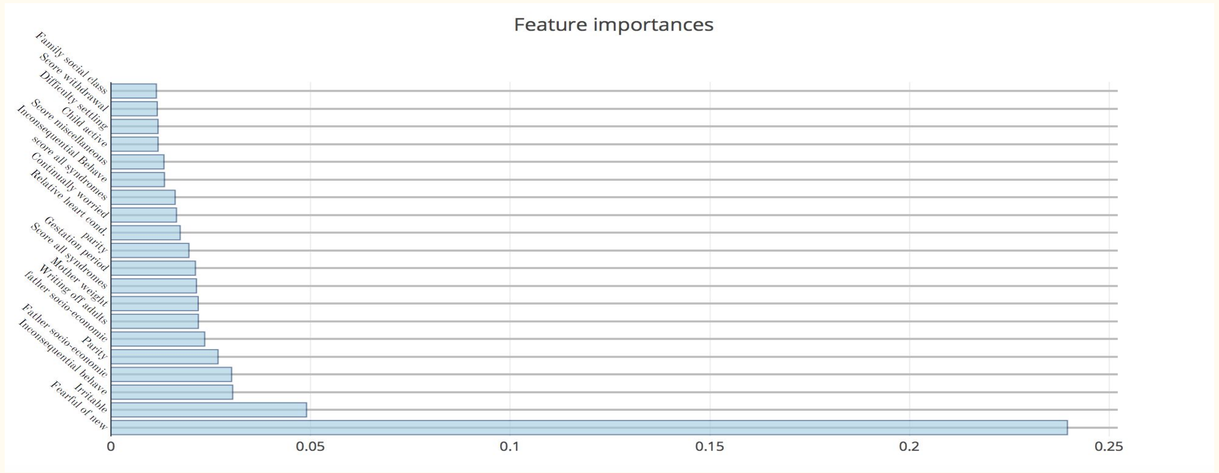
Early Childhood Seizures and Relapse in Adolescence

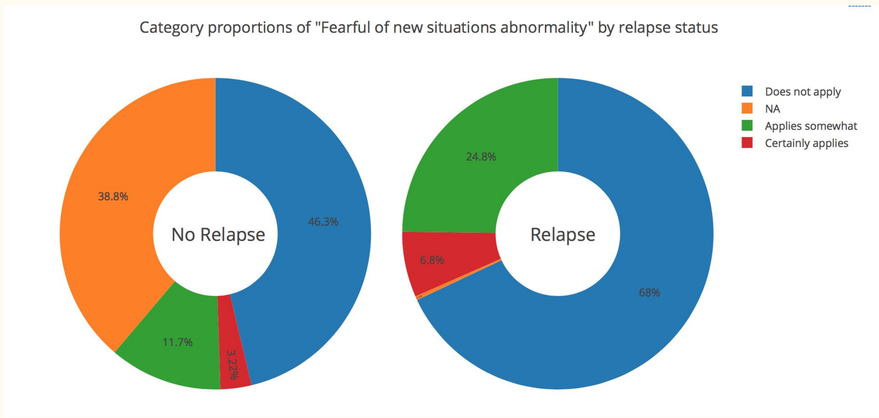
Fernando Becerra, Jessica Cao, Yi Ding, Andrew Mayo

Introduction

As one of the most far-ranging and devastating neurologic diseases, epilepsy afflicts over 50 million individuals worldwide. The most diasabling aspects of seizures are its unpredictability, potentially life-threatening impairment of consciousness, and the embarrassment of incidences, which 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 recurrence and predict 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 a 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 to get a broad view of the performance for each model. Predictions were based on majority vote.
2. The decision tree model outperformed other methods. We further tuned the model via cross validation to improve its performance.
3. We created a balanced dataset to fit the 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 the non-relapse group and a test accuracy of 0.552 for the 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, which is consistent with various existing literatures.

The 1958 NCDS dataset is a social survey-based longitudinal dataset, which imposes two difficulties. First, a large amount of missing value existed in the dataset. We assign 'missing' to missing values in categorical data, and the mean value to missing values in continuous data. We may improve the data analysis by using more complicated methods such as multiple imputation. But if the data is not missing at random, our results will remain biased. The second problem is that although 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 longitudinal study, but the problem could be addressed by designing a study targeted specifically at epilepsy analysis.

Conclusion

We have presented a seizure prediction model for the 1958 National Child Development Study. The model fit 100 decision trees in the training set to predict seizures in test set (taking the majority as the result). 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