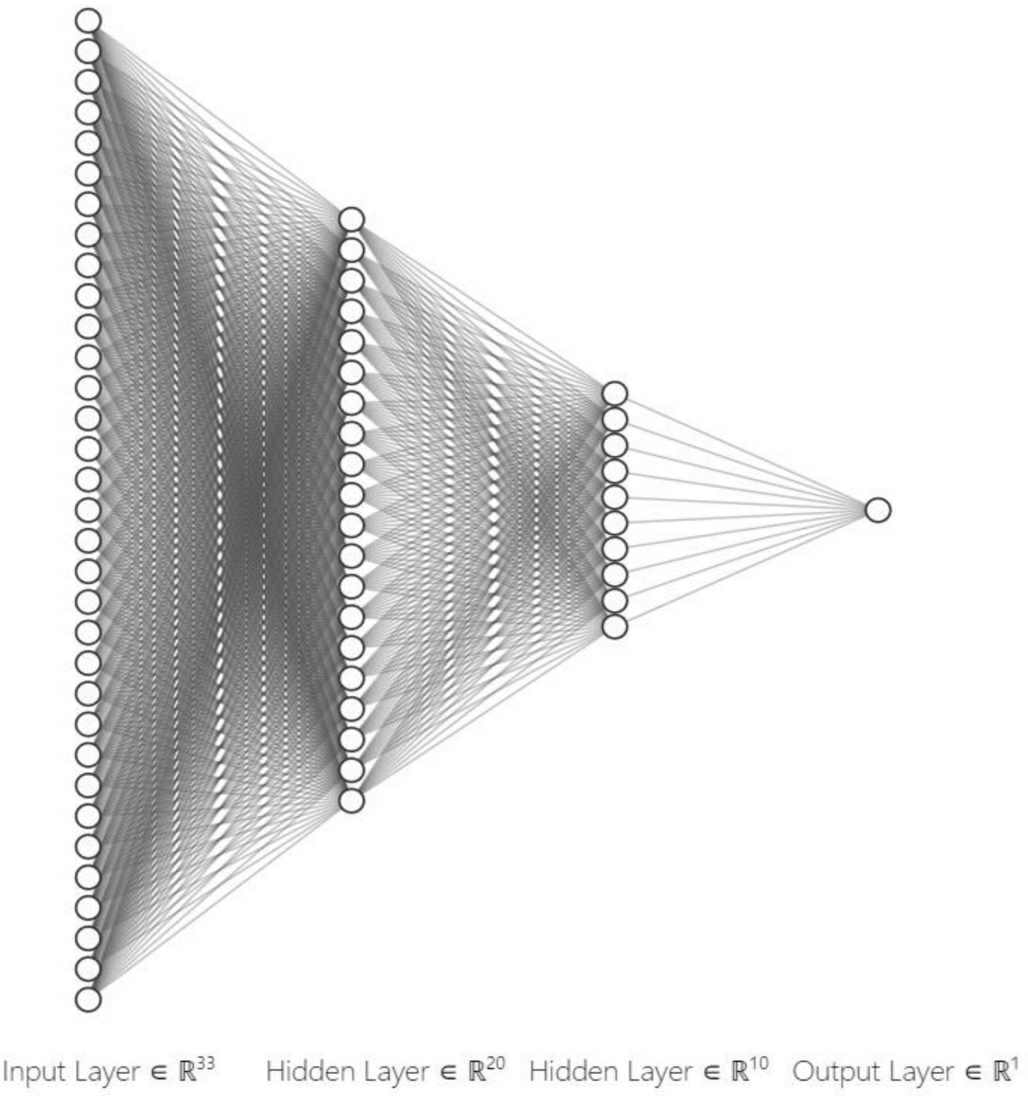
Predicting Mental Health Outcomes with Deep Learning

By Christiana Marchese

With the mentorship of Dr. Xinlian Liu



INTRODUCTION

There are many research studies linking demographical features to mental health outcomes, and many studies have also been devoted to understanding how various community characteristics influence individual's mental health. As a result of the dire and immediate consequences of poor community mental health, it is of utter importance to be able to assess a community's risk for certain mental health outcomes, such as suicide, in order that a community may be well equipped to address the underlying needs of their constituency. But in order to assess a community's risk for mental health outcomes such as suicide, a variety of community features must be taken into account and their differing affect on mental health outcomes must be weighed appropriately; to do this, I have developed deep learning models that take into account community characteristics of counties in the state of California and predict the risk level of these counties for suicide.

DATA

The data used to train the models was sourced from the U.S. Center for Disease Control and Prevention (CDC) and the U.S. Census Bureau.

- All data is from the year 2015
- From the Census's American Community Survey, the DP02 (select personal features) and the DP03 (select economic features) tables were used
- The dataset consisted of 57 out of 58 California counties
- 33 demographical features for each county were used for model prediction. Some features include:
  - Veteran population
  - Median income
  - Unemployment rate
- Counties' risk level for suicide was classified based on whether it exceeded the national average suicide rate, 13.42 per 100,000 individuals

METHODS

Collect Data

Query personal demographic and economic data from Census Bureau

Obtain CDC suicide count data

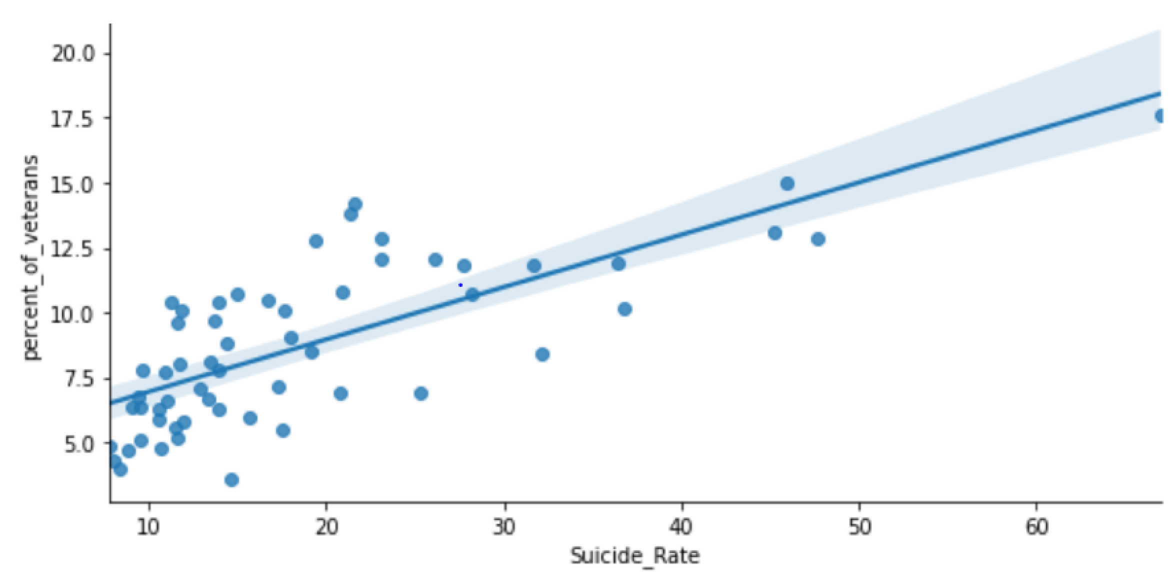
Preprocess Data

Merge and clean data

Remove incomplete data features

Classify risk level of counties based on suicide counts and population

Choose Features



Visualize features and select ones with moderate correlation

Select features of common study

Train Models

Create training set of 33 demographical features from the 10 counties with the highest suicide rate and the 10 counties with the lowest suicide rate

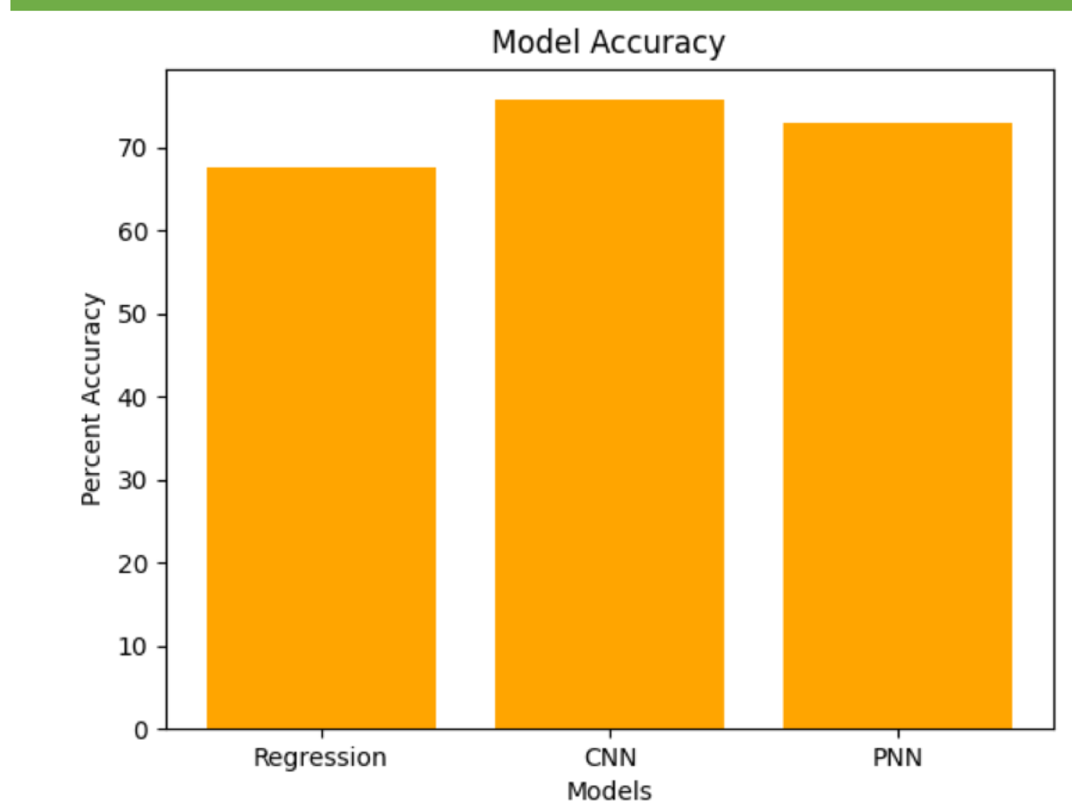
Train the linear regression model, perceptron neural network, and convolutional neural network

Optimize and Test

Test models on remaining 37 counties and score them based on accuracy of prediction

Adjust deep learning models' learning rates and number of layers to maximize accuracy

RESULTS



In predicting 37 California counties' risk level for suicide, the linear regression model had an accuracy score of 67.57%, the perceptron neural network (PNN) had an accuracy score of 72.97%, and the convolutional neural network (CNN) had an accuracy score of 75.68%.

The linear regression did the worst of all the models because it was unable to extract important features from the input data like the deep learning models did. The CNN worked the best for predicting a county's suicide risk level, due to its ability to extract important features from the data, which made its training more effective; and therefore, its predictions more accurate. The PNN was the second best at predicting county risk level for suicide as it still has the ability to extract important feature from the data just to a lesser extent than the CNN.

CONCLUSION

With the prevalence of community assessment for mental health outcomes, I set out to create deep learning models that predicted such outcomes based on 33 community characteristics. Predictions of county suicide risk level on a validation set of 37 California counties were accurate for 75.68% of counties using a CNN and 72.97% of counties using a PNN; both deep learning models outperforming standard statistical analysis. Due to the data being constrained to California counties alone (58 counties total), there was limited data for training and testing. With a greater locational span and more data, more accurate predictions could be plausible.

AREAS OF FURTHER INTEREST

A possible extension of this work could be looking into how geospatial characteristics play a role in mental health outcomes as cross-county interaction would affect the experiences of community members. This study seems especially prevalent in the case of California with many people commuting across counties for work in more prominent counties like Los Angeles, Santa Clara, and San Francisco County.

REFERENCES

1. Huang, X., Ribeiro, J. D., Musacchio, K. M., & Franklin, J. C. (2017). Demographics as predictors of suicidal thoughts and behaviors: A meta-analysis. *PloS one*, *12*(7), e0180793. https://doi.org/10.1371/journal.pone.0180793
2. 1D Convolutional Neural Networks for Time Series Modeling - Nathan Janos, Jeff Roach. (2018, November 29). [Video]. YouTube. https://www.youtube.com/watch?v=nMkqWxMjWzg&feature=youtu.be

ACKNOWLEDGMENTS

A special thank you to XSEDE and the XSEDE EMPOWER Program for making available the necessary computational resources for this project as well as allowing undergraduate students, like me, to participate in the work of XSEDE.

Thank you also to Dr. Xinlian Liu for his amazing mentorship and guidance throughout the duration of this project.

THE MODELS' REPOSITORY

https://github.com/christymarc/CA_SuicideRate_Models