MIND YOUR METALS: EXPLORING THE IMPACT OF HEAVY METALS ON DEPRESSION WITH MACHINE LEARNING

MEKHALA KUMAR AND ROY YOON, UNIVERSITY OF MASSACHUSETTS AMHERST

MACHINE LEARNING FOR THE SOCIAL SCIENCES 695M, SPRING 2023

INTRODUCTION

The association between heavy metals and depression has been a topic of interest in scientific research. Studies have investigated the potential relationship between exposure to heavy metals and the development or exacerbation of depressive symptoms. Here are some general observations from the existing literature:

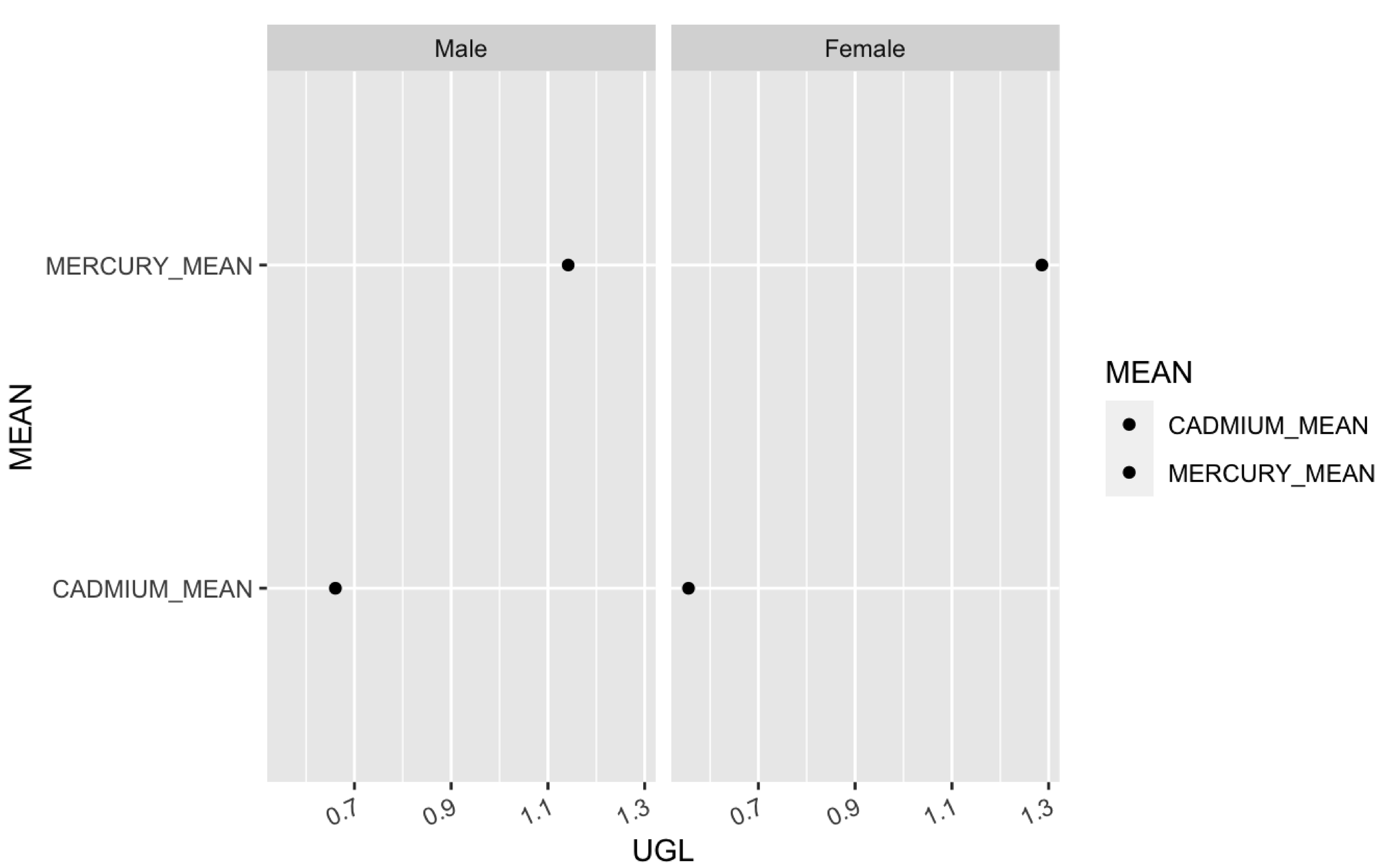
1. **Lead**: Lead is a neurotoxic metal that can affect the central nervous system and disrupt neurotransmitter function, potentially contributing to depressive symptoms.

2. **Mercury**: Mercury is a toxic metal that can accumulate in the body, including the brain. It may have adverse effects on mood regulation and neurotransmitter function, potentially contributing to depressive symptoms.

3. **Cadmium**: Cadmium is a toxic metal found in industrial pollutants and tobacco smoke. It may disrupt neuroendocrine function and contribute to the development of depressive symptoms.

DATA

The **National Health and Nutrition Examination Survey (NHANES)** is a large and comprehensive dataset that includes information about the health and nutrition status of individuals in the United States. The 2005-2006 dataset (ICPSR 25504) is a nationally representative sample of individuals aged 6 months and older who participated in the survey between 2005 and 2006.



Comparing Heavy Metal Averages Between Depressed and Non-Depressed Groups

LEAD: p-value = 0.9397
CADMIUM: p-value = 1.464e-06
MERCURY: p-value = 2.567e-06
INORGANIC MERCURY: p-value = 0.2626

These results suggest that there are statistically significant differences in the mean levels of Cadmium and Mercury between depressed and non-depressed groups. However, there are no significant differences in the mean levels of Lead and Inorganic Mercury between the two groups.

RESEARCH QUESTION

Heavy Metals and Depression aims to **predict how factors such as age, gender, and heavy metals(lead, cadmium, mercury, etc.) can have an effect on one's mental health, *specifically depression***. We will examine this topic as a classification problem.

From the National Health and Nutrition Examination Survey we chose to examine:

Independent Variable(s):

- **DS1: Demographics**
  - Age
  - Gender
- **DS102: Blood Lead and Blood Cadmium**
  - Lead
  - Cadmium
- **DS130: Blood Total Mercury and Blood Total Inorganic Mercury**
  - Mercury
  - Inorganic Mercury

Dependent Variable:

- **DS209: Depression Screener**
  - Feeling down, depressed, or hopeless

METHODOLOGY

Test-Train Split: 80-20

Predictors Used: Age, Gender, Cadmium and Mercury

Models Used: KNN, Random Forest, SVM Radial and SVM Linear

For KNN:
We tested for different values of k = 1 - 10.

For Random Forest:
1. 20 to 100 trees with an increment of 20.
2. 6 different hyper-parameter control values 1 -6

For SVM Radial:
We observed a sequence of values from 0.1 to 1 with an increment of 0.1 for the width of the RBF kernel. Additionally, we observed a sequence of values from 1 to 10 with an increment of 1 for the regularization parameter.

For SVM Linear:
We observed a sequence of values from 1 to 10 with an increment of 1 for the regularization parameter.

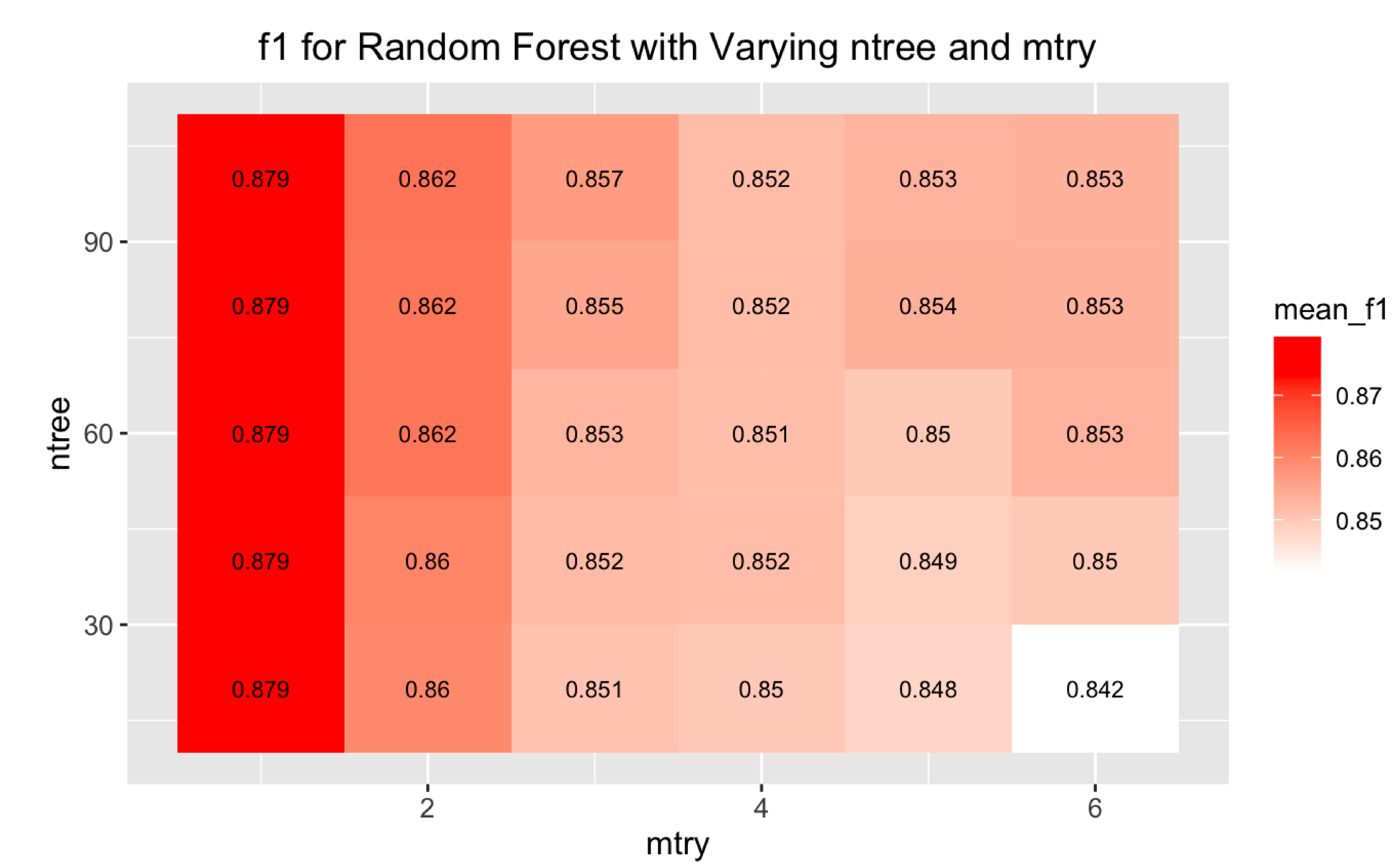
RESULTS

For the training data we observed the best F1 scores for:

- KNN:
  - K=1, F1 score= 0.99
- SVM Radial:
  - sigma = 0.8 C = 1, F1 score= 0.88
- SVM Linear:
  - : C=1, F1 score= 0.88

Model Comparison using Test Data

- **KNN**: 0.85
- **Random Forest**: 0.86
- **SVM Radial**: 0.85
- **SVM Linear**: 0.85



CONCLUSION

We found that there is a statistically significant effect of Cadmium and Mercury between depressed and non-depressed groups.
While running the models on the training set, KNN performed the best. However, when we compared the models using the test set, the difference in the F1 scores was minor, though the Random Forest Model performed slightly better.

In summary, the Random Forest model generally performs well in terms of overfitting and bias-variance tradeoff due to its ensemble nature. SVM with a linear kernel and KNN also show similar performance in terms of overfitting and bias-variance tradeoff. However, Random Forest and SVM with a radial kernel offer more flexibility in capturing complex patterns, while SVM with a linear kernel provides better interpretability with its coefficients.

REFERENCES

Oskar S, Stingone JA. Machine Learning Within Studies of Early-Life Environmental Exposures and Child Health: Review of the Current Literature and Discussion of Next Steps. Curr Environ Health Rep. 2020 Sep;7(3):170-184. doi: 10.1007/s40572-020-00282-5. PMID: 32578067; PMCID: PMC7483339.
Xia F, Li Q, Luo X, Wu J. Machine learning model for depression based on heavy metals among aging people: A study with National Health and Nutrition Examination Survey 2017-2018. Front Public Health. 2022 Aug 4;10:939758. doi: 10.3389/fpubh.2022.939758. PMID: 35991018; PMCID: PMC9386350.
Xihang Fu, Huiru Li, Lingling Song, Manqiu Cen, Jing Wu, Association of urinary heavy metals co-exposure and adult depression: Modification of physical activity, NeuroToxicology, Volume 95,2023,Pages 117-12 ISSN 0161-813X,https://doi.org/10.1016/j.neuro.2023.01.008.