

IDSN 542: Machine Intelligence

Homework 4

Due: 10/15/25

Goal

This is a continuation of lab 5. You are to use the same dataset as you used in lab 5.

Setup

- Your Python file(s) must begin with comments in the following format (replace the name and email with your actual information):

```
'''
Name
IDSN 542, Fall 2025
USC email
Homework 4
'''
```

Requirements

Use the same dataset that you used for lab 5.

Create a reliable training set and test set. It doesn't have to be 80% and 20%, but that is a good place to start.

Use your lab 5 data preparation as the start of this assignment. In addition, do the following:

- Scaling your data with StandardScaler if you have numeric data that varies a lot in magnitude
- Create a Pipeline and ColumnTransformer (if you have categorical data) to apply all your transforms

Train your prepped dataset using cross validation. Choose the number of folds.

You do not have to fine tune your model – finding the best hyperparameters. Just take the defaults.

Use two of the three ML models that we've covered: linear regression, decision tree regression, or random forest regression.

Once you've trained your model, use it on your test data.

Write up a report as to what you did. This includes:

- Which two ML models you chose and why you chose them
- The results of the two ML models on your test data predictions. This could be a screenshot of the printout of each model's performance metrics – like what we did in class in the chapter 2 v6 code
- Which ML model performed better and which worse. Why do you think that is what happened?

Submit your report and all your Python code in a ZIP file and submit to Brightspace.