

# IDSN 542: Machine Intelligence

Final Project Part 3

Due: 12/10/2025 @ 11:59PM

## Goal

For this last part of the final project, you are to train, test, and tune on your dataset to determine the best performing algorithm from the three that you chose in part 2.

## Setup

- You get to create whatever Python scripts you need. You should use the custom transformer class we covered in class if you are adding new attributes. They should go into their own Python script file – like we did in class.

```
'''
Name
IDSN 542, Fall 2025
USC email
Final Project Part 3
'''
```

## Requirements

A document that contains the additional sections to your document from part 2. Just add to your document that you submitted for part 2.

- ML Algorithms:** You are to have trained on at least 3 different ML algorithms. Use the root-mean-squared error (RMSE) to determine your average error on the training, or use recall/precision/F1 or AUC/ROC curves. Choose the algorithm that has the smallest RMSE – or use accuracy/recall/precision/F1.
- Data Cleanup:** Discuss what you had to do to clean up your data. That may include deleting attributes/columns that have little to no correlation, filling in missing values, removing rows that have incomplete data that you cannot correct, etc. As part of this discussion, include the lines of code that you have in Python that performed this work.
- Algorithm Tuning:** Do a grid search with GridSearchCV or a randomized search with RandomizedSearchCV to tune your hyperparameters. You are to have at least 3 iterations of this hyperparameter search. Document the hyperparameters you use for each iteration and the results you obtained in your writeup. The process you follow to narrow the search for hyperparameters is important. Your final submitted code should only have the best hyperparameters. Did tuning the hyperparameters improve precision/recall/accuracy/F1 score? If you are using artificial neural networks, discuss how you trained the model.
- Results:** Discuss how much of your data you used to train and how much to test. Did you have to separate your dataset into a test set and training set or was that already available to you. How well did you do on the test set? Make it clear what Python code produces your results.
- Analysis of results:** What did you learn from this dataset – whether predictor or classifier? Depending on how much cleanup you had to do, did that bias your data or lead to a better result? If you had more time, could you have done other things to improve your result?
- Final Presentation:** Since we won't have time for everyone to present at the final exam period on December 11<sup>th</sup>, you are to record no more than a 5-minute video of your final presentation. Submit your presentation slides and your video in your ZIP file. If your submitting file size is too big, submit a separate ZIP of your final presentation slides and video.

## Deliverables

A compressed folder containing your writeup, your presentation slides, your presentation video and all Python code you used to prepare your dataset for training of your ML algorithm and your data or a link to it. Don't forget to submit your dataset, as well – or a link to it that your code can access.

## Grading

Item	Points
ML Algorithms	10
Data Cleanup/Algorithm Tuning	20
Results	20
Analysis of Results	40
Python Code	30
Final Presentation	30
Total*	150