

The Spectacular Sailors - Checkpoint 4 Machine Learning

In this Checkpoint

- Can we predict whether an officer is in a crew, a community, or is unaffiliated?
- Can we whether any particular group of officers is a crew or not based on aggregated data about their member officers?

▼ Getting Started

Libraries and tools for this analysis:

- Data Structures: Numpy and Pandas
- Machine Learning: SKLearn
- Data Extraction: DataGrip (PostgreSQL)
- Data Source: CPDB
- Analyzed Data: The Spectacular Sailors/Checkpoint 4/src

```
import numpy as np
import sklearn
import pandas as pd
from sklearn.model_selection import train_test_split
from sklearn import linear_model, metrics
from sklearn_pandas import DataFrameMapper
import matplotlib.pyplot as plt

# read the officer and cohort data into pandas dataframe
url = 'https://raw.githubusercontent.com/Northwestern-Data-Sci-Seminar/Invisible-In
df = pd.read_csv(url, header = 0)
df = df._get_numeric_data()
original_headers = list(df.columns.values)
df
```

	officer_id	avg_coaccusals	avg_years_on_force_at_incident	avg_age_at_i
0	1	2.166667	5.500000	{
1	2	4.250000	4.750000	{
2	4	1.714286	27.428571	{

