International Stroke Trial: Predicting the Patient

TOM HAIGHT

SCOTT HORVATH

ASMAA BOUALEM

ERIC WANG

Primary Focus

Our Question: What are the potential outcomes of someone experiencing an acute stroke at the end of a 6-month period?

- What effects this mortality rate?
- How can we use factors to predict potential outcome?

Data Source

Primary Source: The International Stroke Trial
- Conducted between 1991-1996
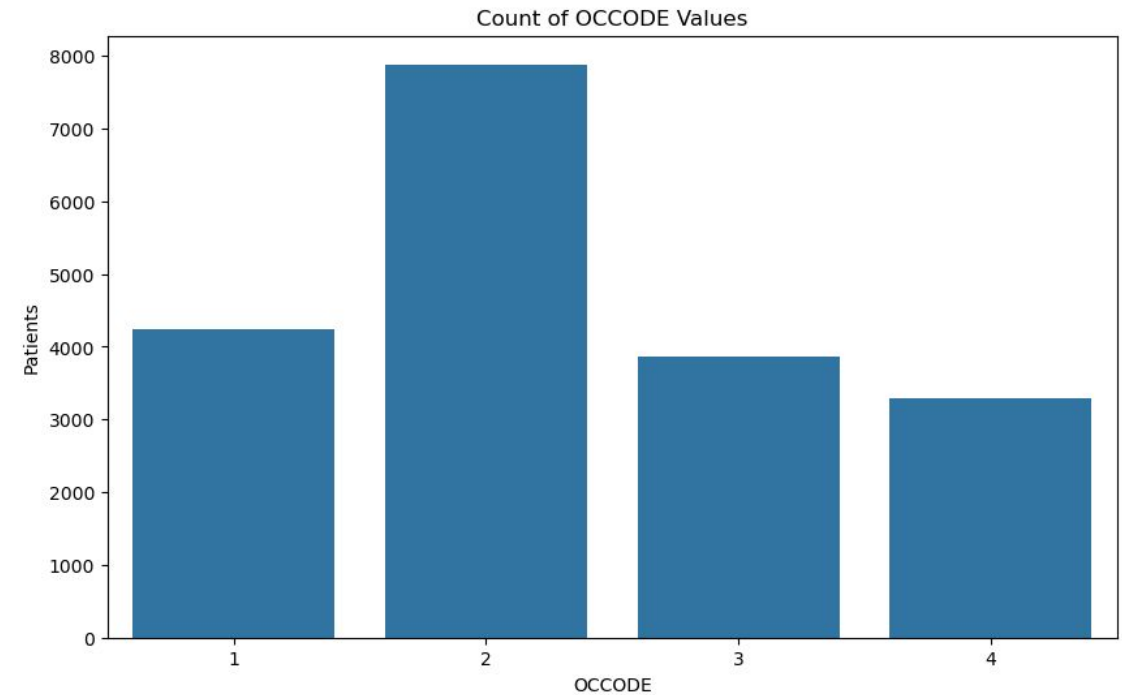- Establish effectiveness of medicine on patients

BMC (BioMed Central)
- Purpose
- Credibility
- Trial information

BMC

Initial Steps

DISCOVERING THE QUESTION

❖ What is it we want to walk away with?

❖ What is "OCCODE": (1 – Deceased, 2 – Dependent, 3 – Not Recovered, 4 – Recovered)

Count of OCCODE Values

Initial Steps

CLEANING THE DATA

- ❖ Converting data into integers

- ❖ Eliminating gaps (treatments, null)

- ❖ Other obstacles (inconsistent labeling)

Model Creation

Pipeline Testing

Necessary Corrections

Preferred Results

- Gradient boosting classifier

- Approx 70% accuracy needing correction
- Removing redundancies

- Correction of "OCCODE" to remove bias

Primary Results

Model Chosen: Gradient Boosting Classifier

Accuracy Score: 87.3%

Conclusion: We are able to predict future outcomes of patients in a 6-month time period utilizing factors recorded within first two weeks of incident

Evaluate the Performance of the Model

Confusion Matrix

```
array([[5248,    265],
       [ 673,   867]])
```

Classification Report

	precision	recall	f1-score	support
1	0.89	0.95	0.92	5513
0	0.77	0.56	0.65	1540
accuracy			0.87	7053
macro avg	0.83	0.76	0.78	7053
weighted avg	0.86	0.87	0.86	7053

Obstacles and Fixes

CORRECTING THE DATA

❖ From day 1, consistent correction

❖ Finding more null values

❖ Removing irrelevant rows

HYPERPARAMETER DILEMMA

❖ Implementation of the hyperparameter scoring

❖ Changes and cleanup required to fix score

❖ Ultimate result: 87.1% , 0.2% improvement

The Next Iteration 🚀

| Sampling Different Models | Honing in on Specific Date Ranges in the Data | Resampling |