

Data and Algorithm Design

Data

We have gathered data from OSHA's records that hold data about specific injuries to people so that we can hopefully predict which employees are most likely to experience injuries of specified severity. This information will allow management more clearly understand potential costs they could incur and take preventive measures to save money over time.

Algorithm Discovery

We have identified two multi-class classification tasks:

1. Predicting the degree of injury an employee is likely to experience
 - a. Using explanatory variables such as age, gender, occupation, activities in the employee's job description, we want to predict whether their injury is likely to result in hospitalization, non-hospitalization, or death.
 - b. Our preliminary attempt on our test data using a random forest classifier resulted in these metrics:

	precision	recall	f1-score	support
Fatality	0.53	0.51	0.52	2124
Hospitalized injury	0.56	0.61	0.59	2509
Non Hospitalized injury	0.17	0.12	0.14	463
accuracy			0.53	5096
macro avg	0.42	0.41	0.42	5096
weighted avg	0.51	0.53	0.52	5096

c.

2. Predicting if the injury will be fatal vs. non-fatal
 - a. Using the same explanatory variables, our random forest classifier performed slightly better:

	precision	recall	f1-score	support
Fatality	0.54	0.48	0.51	2124
Non-fatal	0.66	0.70	0.68	2972
accuracy			0.61	5096
macro avg	0.60	0.59	0.59	5096
weighted avg	0.61	0.61	0.61	5096

i.

3. Predicting if the injury will be severe vs. non-severe
 - a. Using the same explanatory variables, we were able to accurately predict if the injury would be severe:

	precision	recall	f1-score	support
Non Hospitalized injury	0.12	0.60	0.20	463
Severe	0.93	0.57	0.70	4633
accuracy			0.57	5096
macro avg	0.53	0.58	0.45	5096
weighted avg	0.86	0.57	0.66	5096

i.