Toronto Fire Services Incident Data: Revisited
By Geoffrey Clark, for Dr. Ceni Babaoglu
October 2nd, 2018

Methodology: Compare model trained with ROSE data vs. model trained with randomly sampled data. Both datasets created by sampling 60% of entire dataset. Classifier used was logistic regression with below formula and threshold of 0.5.

```
model_logistic_bal <- glm(CRITICAL ~ ., data=I_s_ROSE, family=binomial(link="logit"))
model_logistic_imb <- glm(CRITICAL ~., data = I_s[full_idx$train,], family=binomial(link="logit"))
```

	Original Data (%)	ROSE Data (%)
Accuracy	97.4	91.8
False-Negative Rate	56.9	19.1
Precision	89.4	31.3
Recall	43.0	80.8
F1-Score	58.1	45.2
Confusion Matrix		

	0	1
0	126544	3131
1	280	2369

	0	1
0	117092	1051
1	9730	4448