# Fighting FIRE with FIREpower:

## Using machine-learning models to classify wildfires by intensity

Wayne Chan
Mariam Javed
Shawn Syms

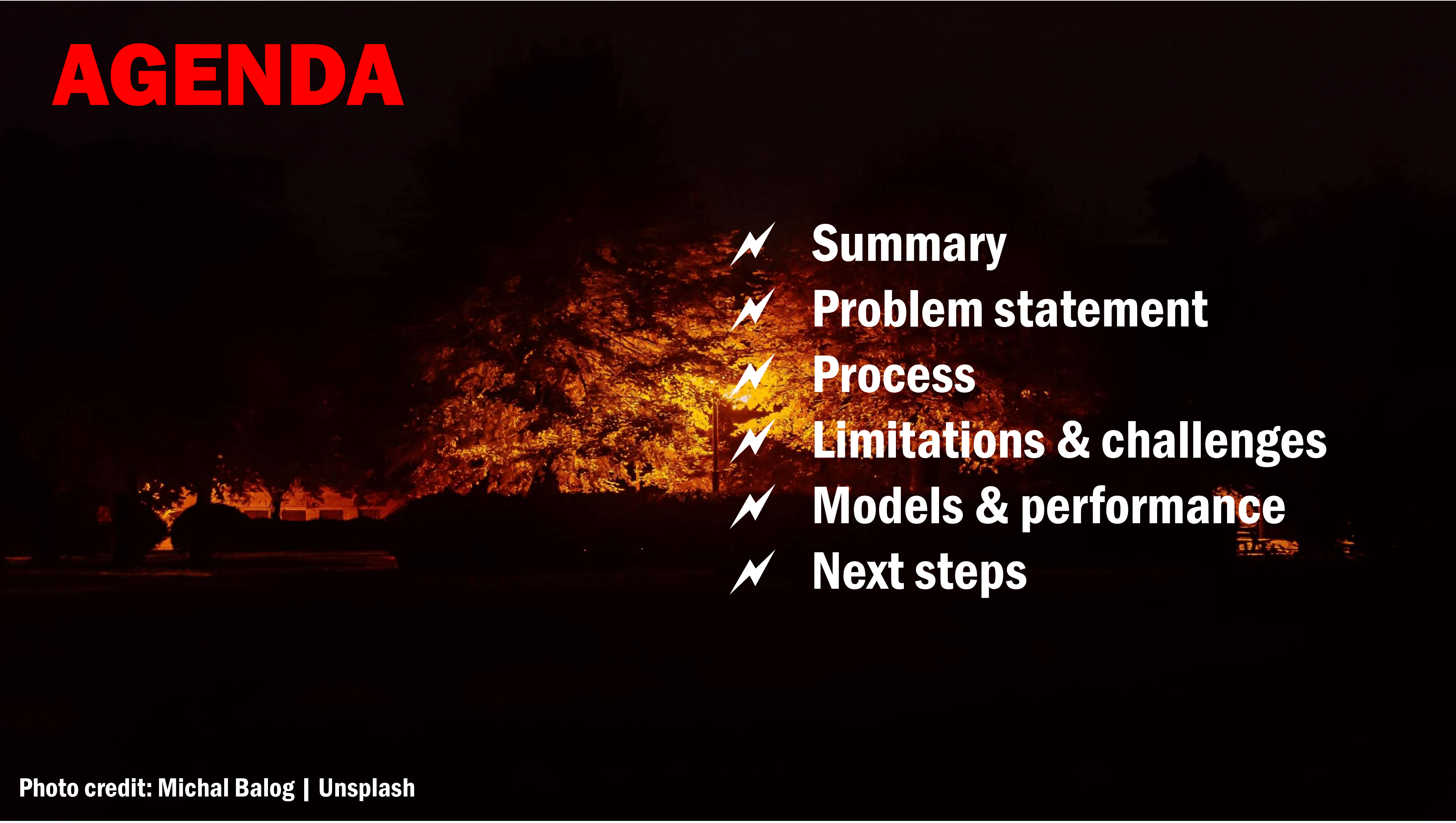



Photo credit: Michal Balog | Unsplash

# PROBLEM STATEMENT

Wildfires have a devastating impact. At risk are homes, infrastructure and wildlife, and an increase in atmospheric toxicity. The costs of fighting these fires also burn out of control: fire-fighting agencies spent an all-time high ($2.9B) fighting fires in the United States in 2018.

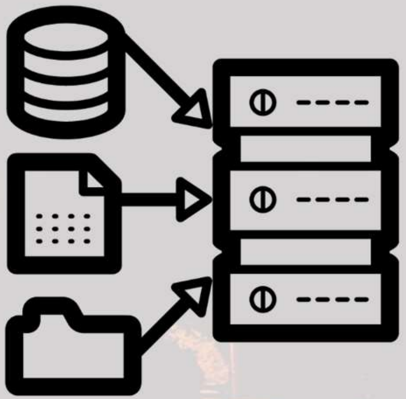
So it's crucial to develop a model to classify the size of fires, so appropriate resources can be gathered to combat them before it's too late. While 90 percent of wildfires in America are caused by human interaction with the natural environment, their ability to spread may depend largely on other variables, which we will explore and engineer to get the necessary results.

We will describe how the size of the fire changes depending on a range of such factors. Our guiding metric will be the accuracy statistical measure. Through implementing a range of machine learning models, we will determine the best possible estimator, and use it to classify fire size to a significant degree of accuracy.

Data source: National Interagency Fire Center

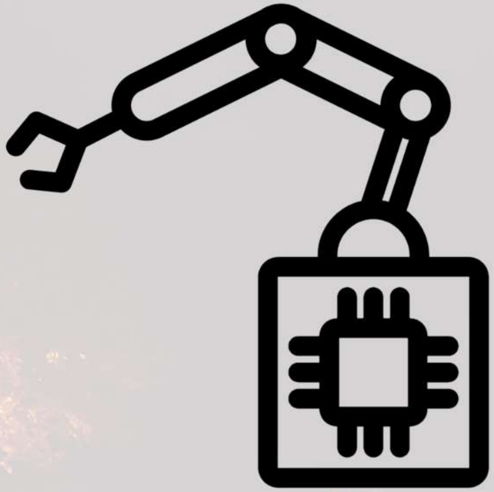
# PROCESS JOURNEY

- Sourcing a dataset was challenging: some had meteorological variables but no locations, others had full longitude/latitude but few other significant details
- We settled upon a well-known dataset of 1.88M U.S. wildfires, but augmented it with relevant weather details using the DarkSky API
- But time and cost constraints limited us to 22,000 of the 1.88M observations



# INITIAL APPROACH

- We initially approached the issue as a regression problem
- In earnest, we built and fit 12 models to see which would perform best:



- ADAboost
- Bagging regressor
- ElasticNet
- ExtraTrees regressor
- Gradient Boosting regressor
- Lasso
- LinearSVR
- Logistic regression
- Random Forest regressor
- Ridge
- Stochastic Gradient regressor
- SVR

# INITIAL **RESULTS**

- Our models performed abysmally
- We added our weather data and still witnessed little to no correlation to our initial target, fire size

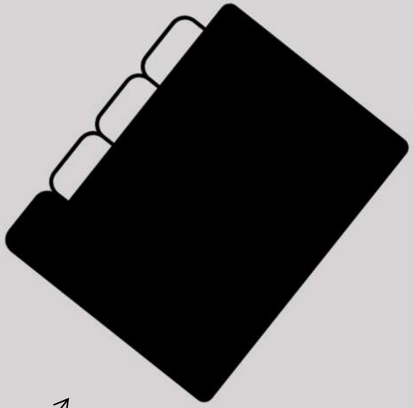
- ~~ADAboost~~
- ~~Bagging regressor~~
- ~~ElasticNet~~
- ~~ExtraTrees regressor~~
- ~~Gradient Boosting regressor~~
- ~~Lasso~~
- ~~LinearSVR~~
- ~~Logistic regression~~
- ~~Random Forest regressor~~
- ~~Ridge~~
- ~~Stochastic Gradient regressor~~
- ~~SVR~~

# INITIAL RESULTS

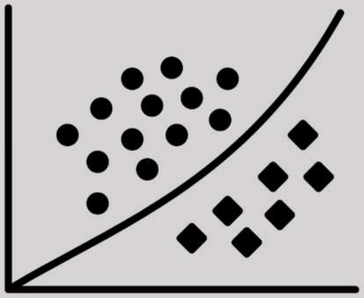
| Model | Best Score | Best Params |
|---|---|---|
| ADA Boost | -63.6354% | {'ada__n_estimators': 10, 'ada__random_state': 42} |
| Random Forest | -13.2784% | {'rf__max_depth': 3, 'rf__max_features': 5, 'rf__min_samples_leaf': 15, 'rf__min_samples_split': 2, 'rf__n_estimators': 100, 'rf__random_state': 42} |
| Linear | -3.5782% | {} |
| ElasticNet | -1.2579% | {'enet__alpha': 0.9950000000000004} |
| Bagging | -1.0297% | {'bag__max_features': 3, 'bag__max_samples': 50, 'bag__n_estimators': 100, 'bag__random_state': 42} |
| Gradient Boost | -0.8349% | {'gbr__max_depth': 3, 'gbr__max_features': 10, 'gbr__min_samples_leaf': 10, 'gbr__min_samples_split': 2, 'gbr__n_estimators': 1, |
| Lasso | -0.6311% | {'lasso__alpha': 100000.0} |
| Ridge | -0.5656% | {'ridge__alpha': 100000.0} |
| Linear Support Vector | -0.4928% | {'lsvr__C': 100} |
| Stochastic Gradient Descent | -0.3272% | {'sgd__alpha': 1, 'sgd__penalty': 'elasticnet', 'sgd__random_state': 42} |
| Support Vector Machine | -0.0872% | {'svr__C': 100, 'svr__kernel': 'rbf'} |
| ExtraTrees | 0.7970% | {'et__max_depth': 3, 'et__max_features': 5, 'et__min_samples_leaf': 3, 'et__min_samples_split': 2, 'et__n_estimators': 100, 'et__random_state': |

# FRESH APPROACH

- We recast the issue as classification and returned to the full dataset
- Fire-size class, a binning of the fires into groups by number of affected acres, would be our new target:

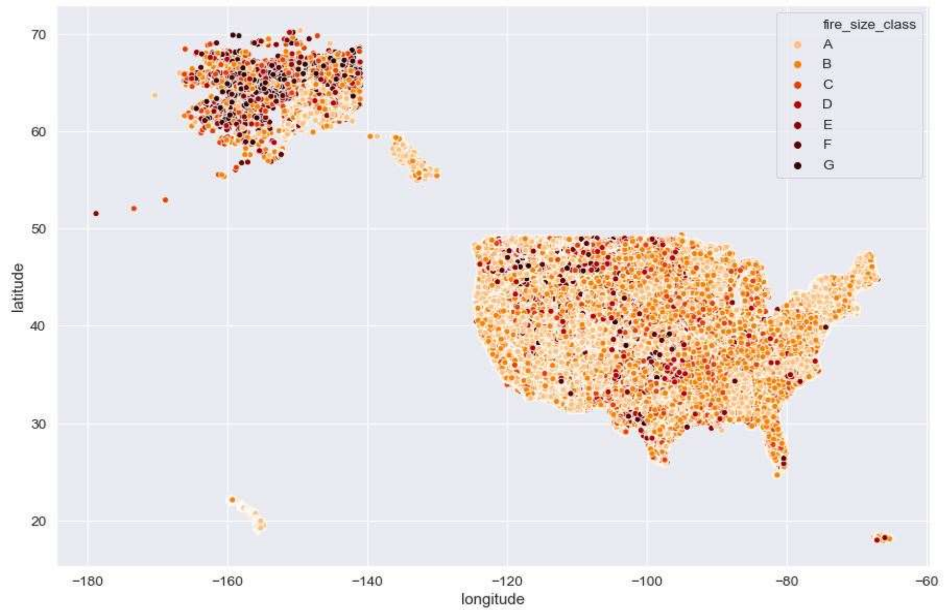


A: greater than 0 but <= 0.25 acres
B: 0.26-9.9 acres
C: 10.0-99.9 acres
D: 100-299 acres
E: 300 to 999 acres
F: 1000 to 4999 acres
G: 5000+ acres

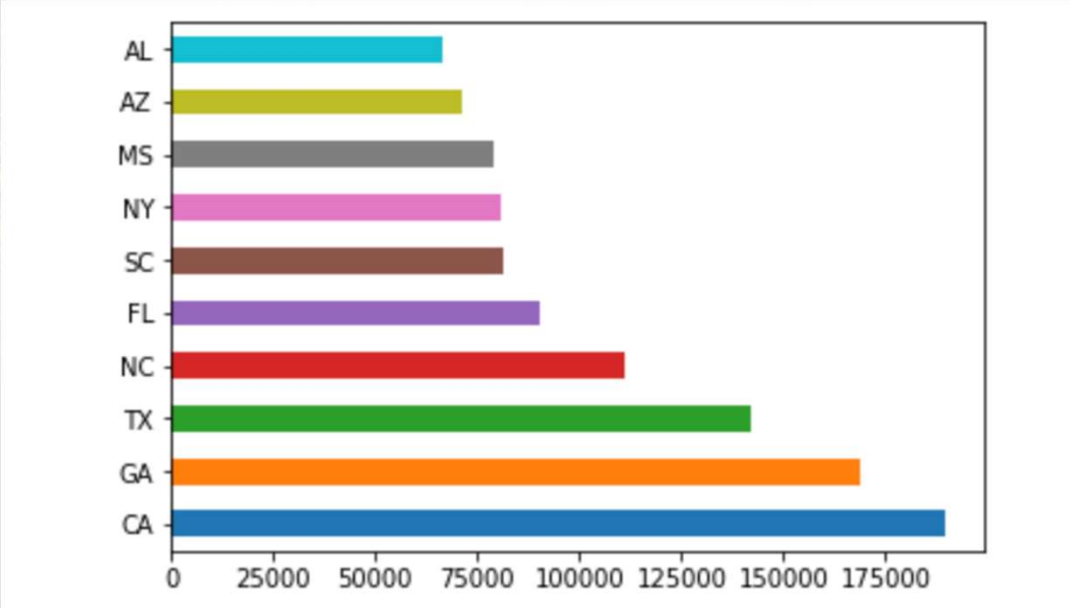


# EDA OCCURRENCES

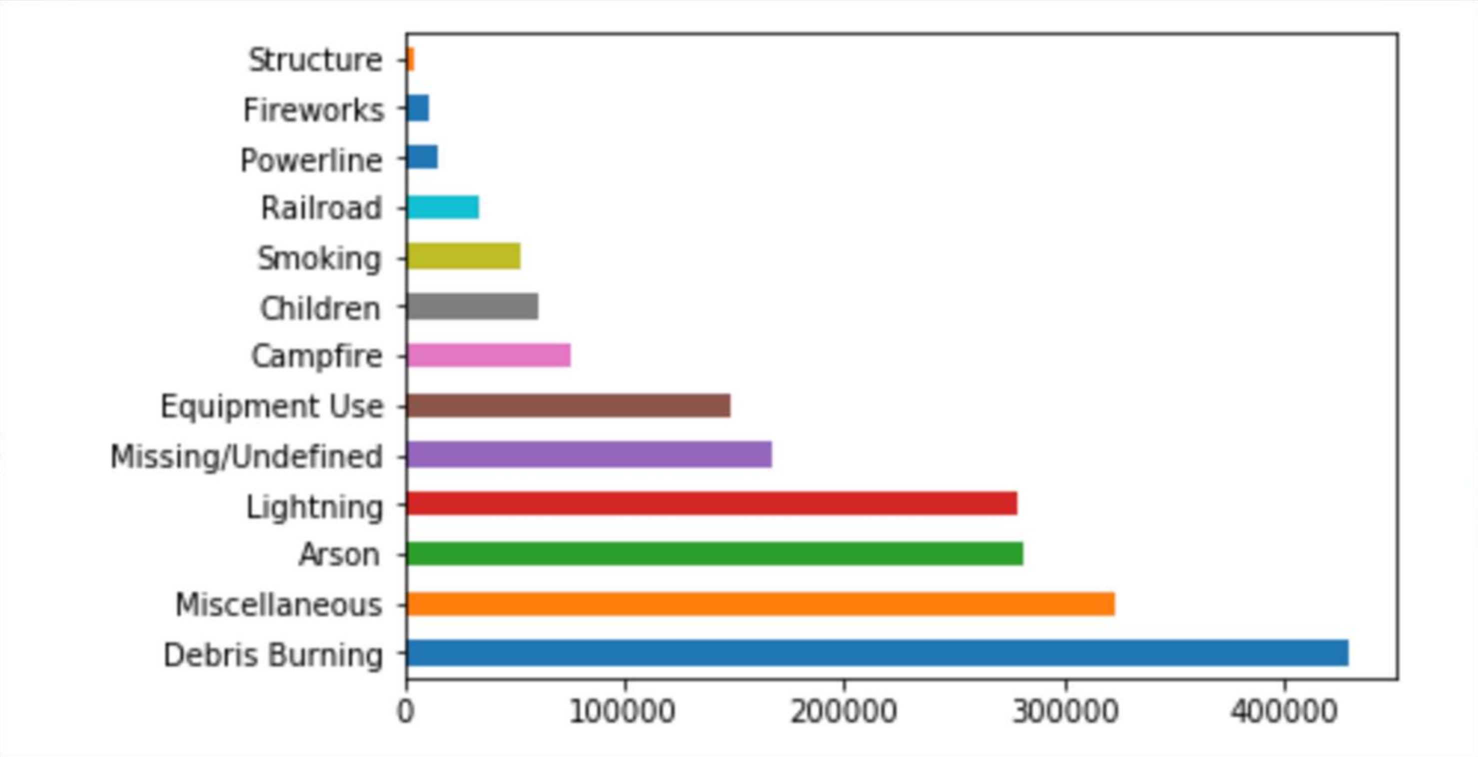
## Fire Occurrences by longitude and latitude




## Fire Occurrences Frequency by State

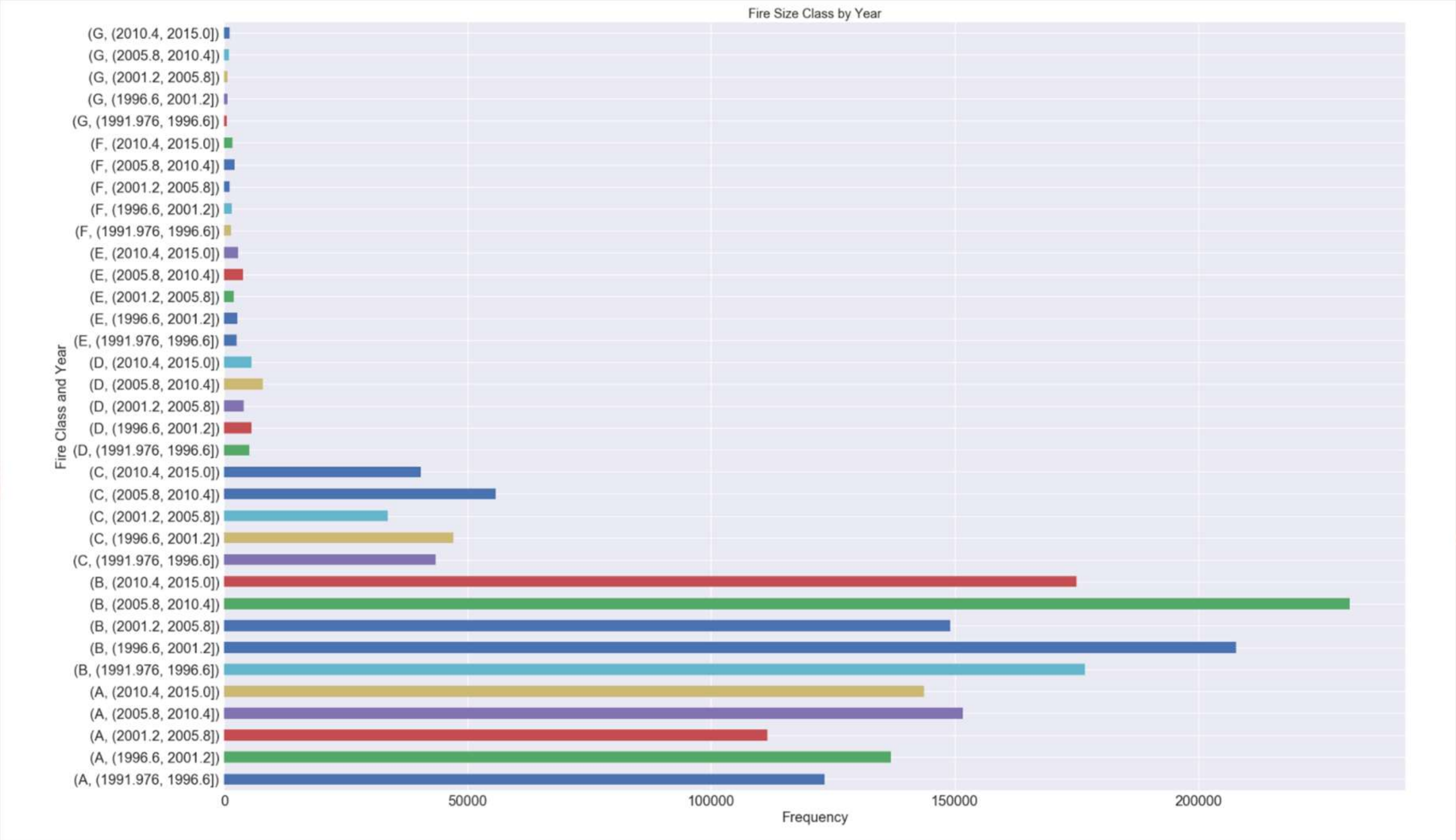


# EDA CAUSES

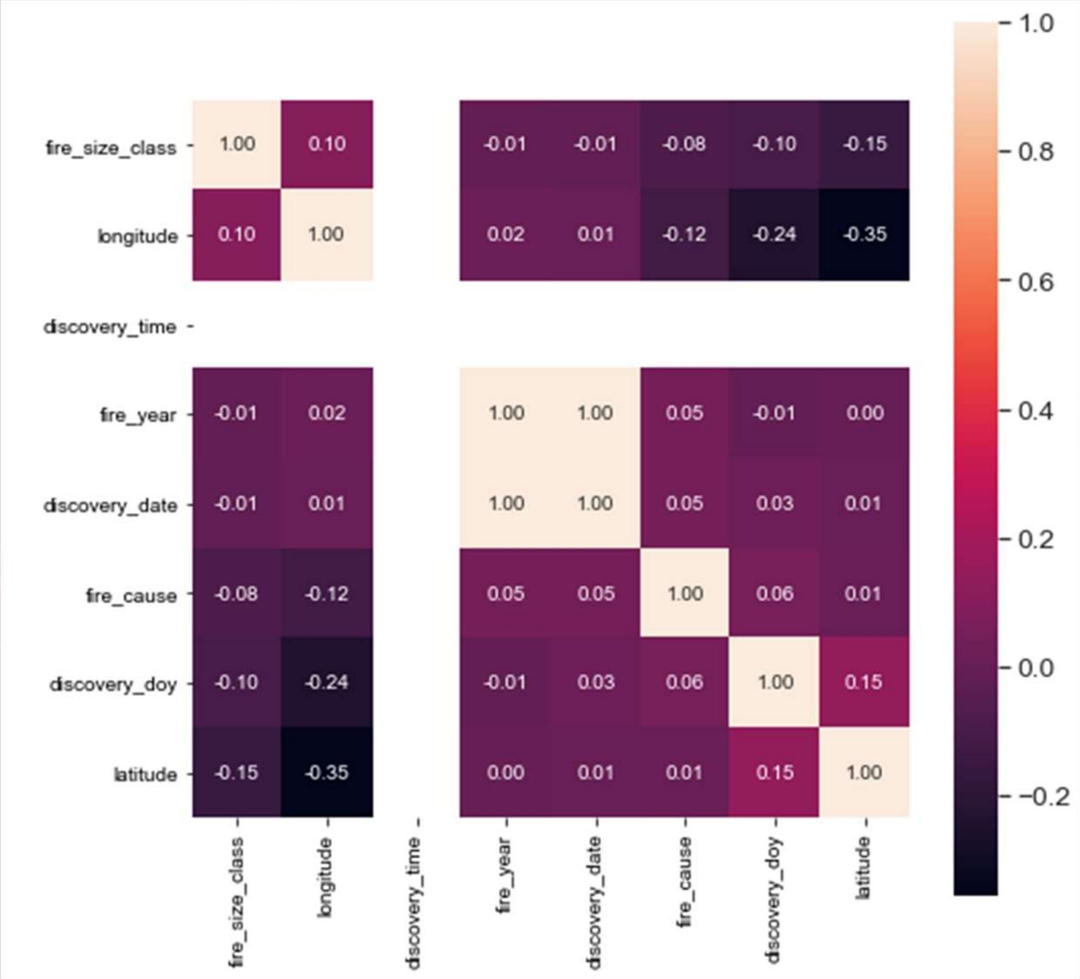


Structure
Fireworks
Powerline
Railroad
Smoking
Children
Campfire
Equipment Use
Missing/Undefined
Lightning
Arson
Miscellaneous
Debris Burning
0
100000
200000
300000
400000

# EDA FIRE CLASS BY YEAR



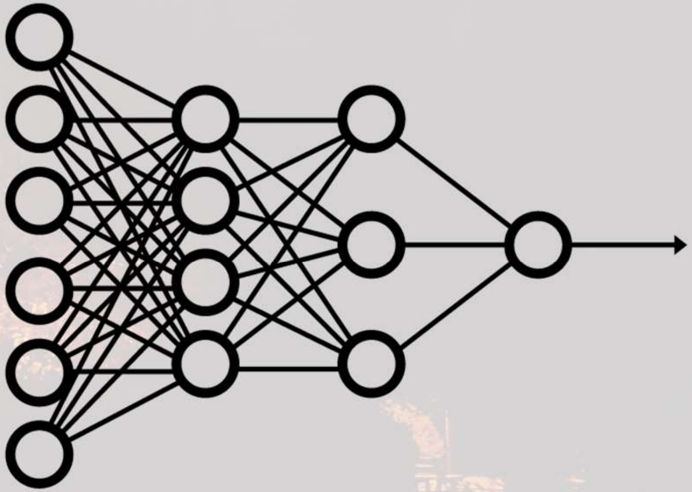
# EDA HEAT MAP



fire_size_class
longitude
discovery_time
fire_year
discovery_date
fire_cause
discovery_doy
latitude
1.00 0.10 -0.01 -0.01 -0.08 -0.10 -0.15
0.10 1.00 0.02 0.01 -0.12 -0.24 -0.35
-0.01 0.02 1.00 1.00 0.05 -0.01 0.00
-0.01 0.01 1.00 1.00 0.05 0.03 0.01
-0.08 -0.12 0.05 0.05 1.00 0.06 0.01
-0.10 -0.24 -0.01 0.03 0.06 1.00 0.15
-0.15 -0.35 0.00 0.01 0.01 0.15 1.00
1.0
0.8
0.6
0.4
0.2
0.0
−0.2

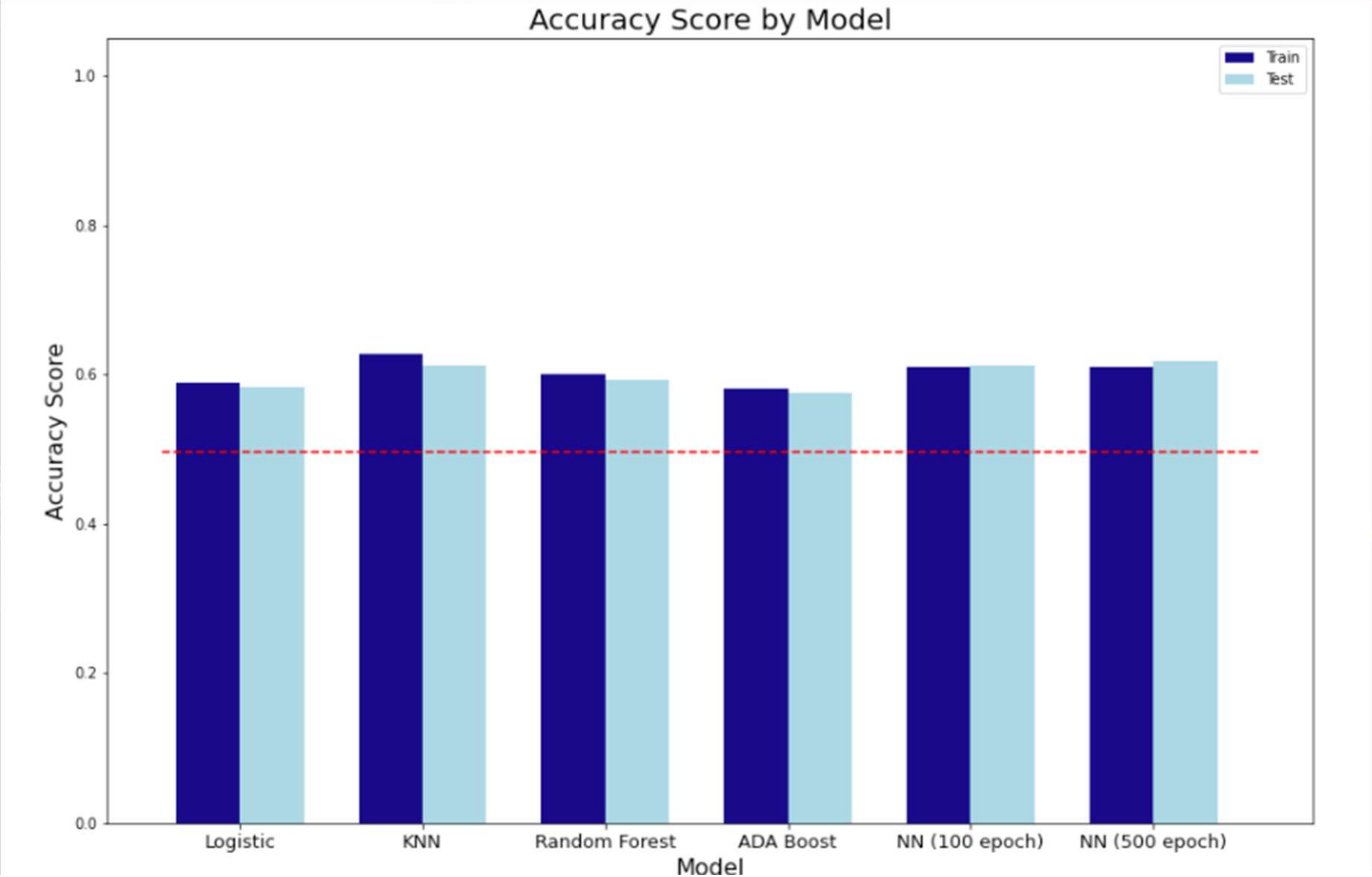
# FRESH APPROACH

- We created and fit five models:
  - 1: Logistic Regression
  - 2: ADAboost
  - 3: kNN
  - 4: Random Forest
  - 5: Feed-forward Neural Network



- We augmented our initial NN with additional hidden layers and tuned its hyperparameters

# BETTER RESULTS



# CONCLUSION

1. Neural Network was most effective model predicting the fire-size class—with train and test accuracy of 61 & 62 percent

2. This dataset is far better suited to a classification problem than a regression one, with a <u>significant</u> boost in model results—from < 0.05 r2 scores to approximately ~60 percent accuracy scores

3. Our results are comparable to others that have used this dataset: telling us our accuracy results are the best you can achieve when using the dataset as is

# RECOMMENDATIONS

- Historical weather data was only collected for 2015 given time/financial constraints. We recommend it is collected for all years (1992–2015) and integrated into the NN model
- There are significant class-distribution imbalances in this dataset. It would be beneficial to implement weight classes to compensate for the imbalance in data points

# QUESTIONS

Wayne Chan
Mariam Javed
Shawn Syms

Photo credit: Michal Balog | Unsplash