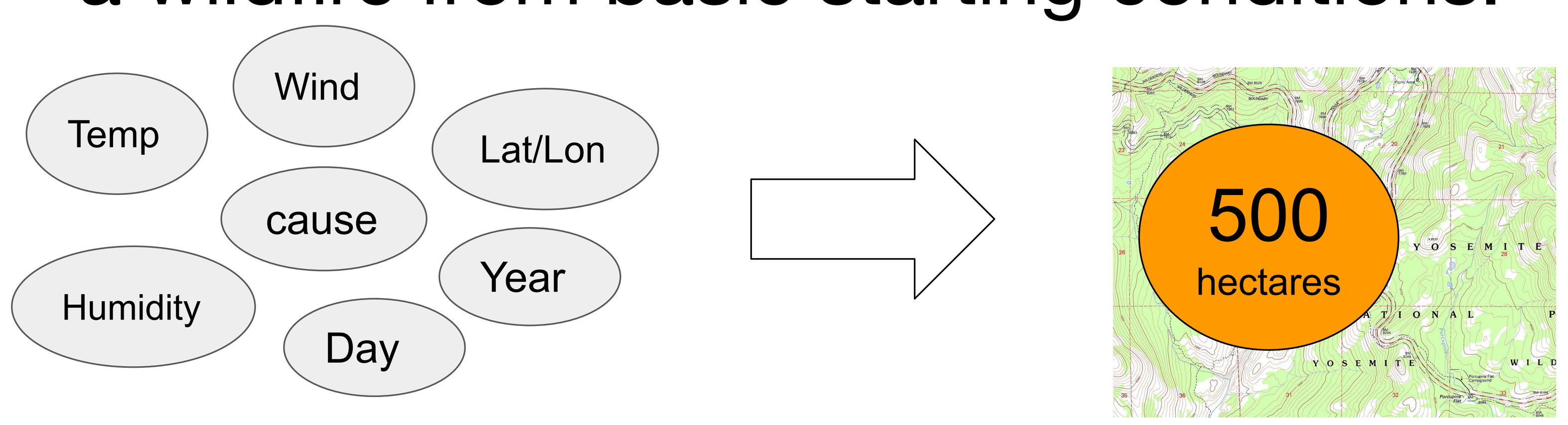
Ben Moore
(benmoore@stanford.edu)

Wildfire Burn-Area Prediction

Adam Stanford-Moore
(sasm@stanford.edu)

Abstract

- **Context:** Wildfires destroy billions of dollars worth of infrastructure. Difficult to predict
- **Purpose:** Design model to predict the size of a wildfire from basic starting conditions.



Fire starting conditions

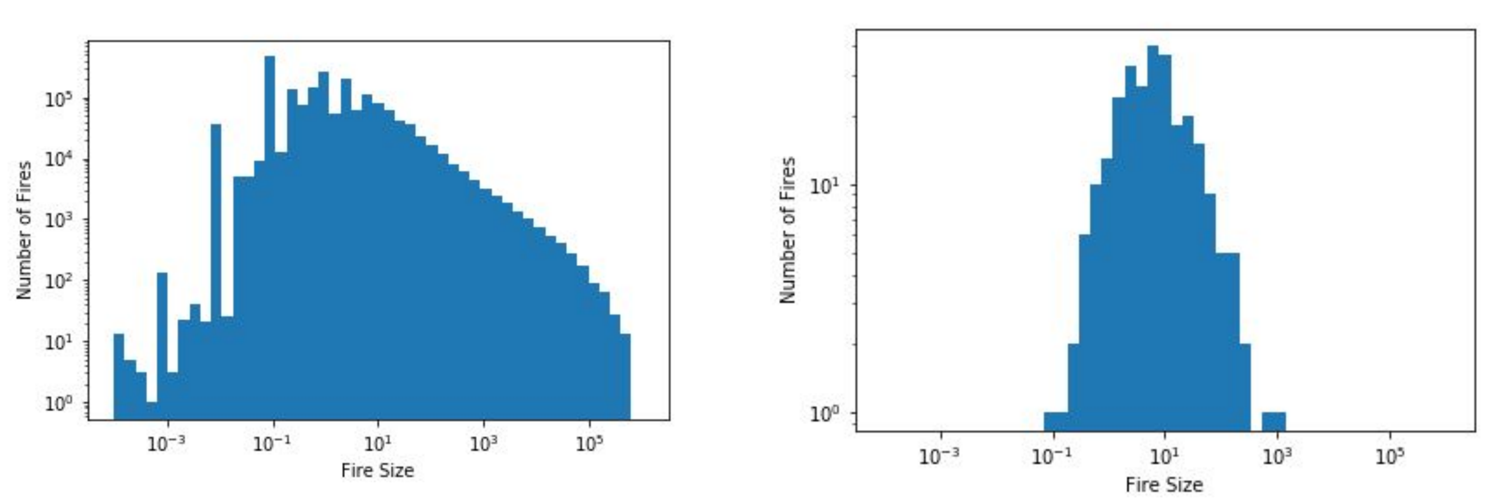
Predict Fire Area

Model/Results:

- Predict final area burned with neural network, SVM, and others
- Best prediction accuracy from SVM: Mean-Absolute-Percent-Error = **78%.**

Data

- 1.88 million historic US fires (Kaggle)[1]
  - Mostly tiny fires. No weather
- 600 fires from park in N. Portugal (UCI)[2]
  - Weather features, but Geo-specific, few samples in dataset, more balanced



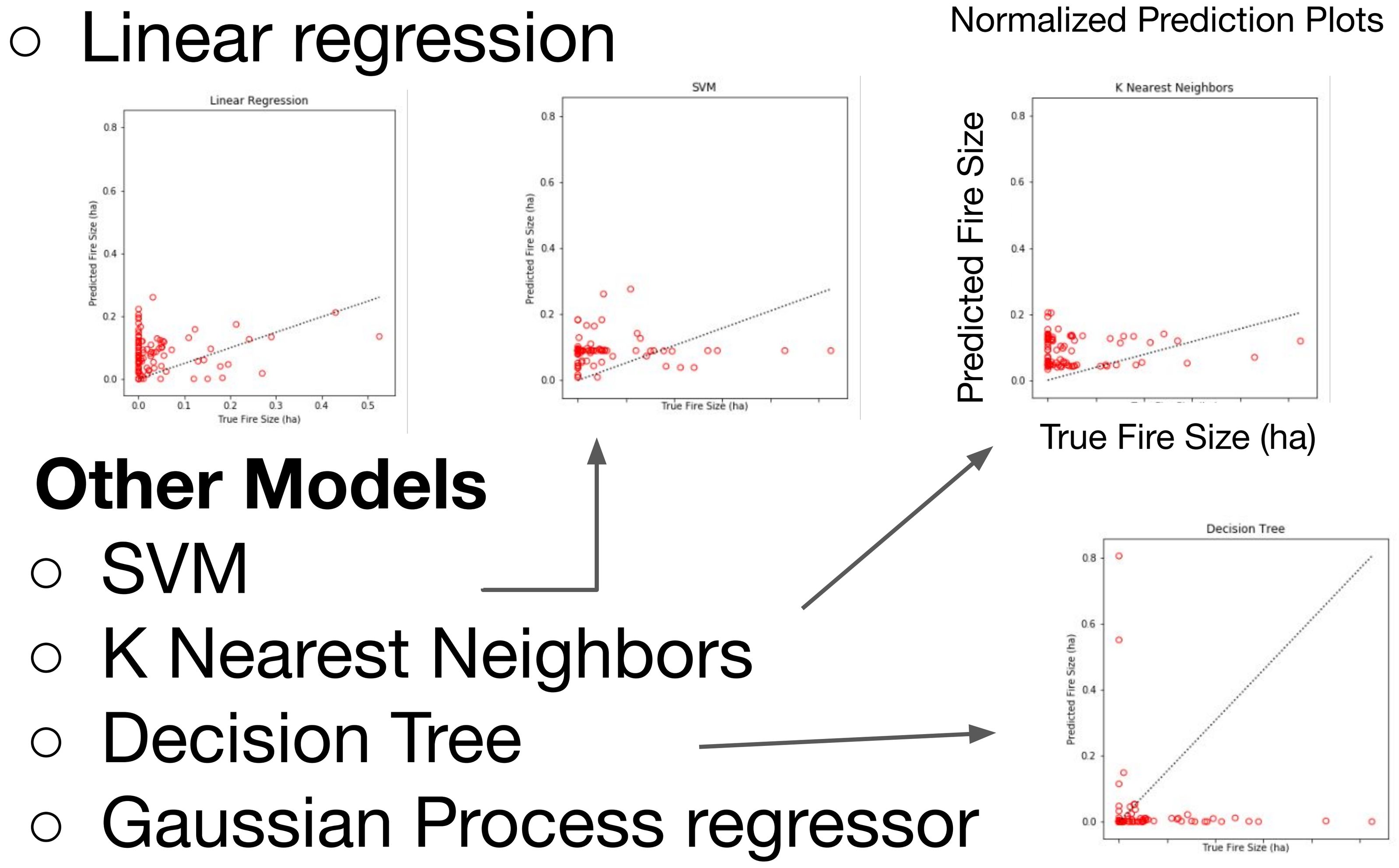
Kaggle dataset (left) swamped by tiny fires, UCI dataset (right) more balanced

[1] Short, Karen C. 2017. Spatial wildfire occurrence data for the United States, 1992-2015 [FPA_FOD_20170508]. 4th Edition. Fort Collins, CO: Forest Service Research Data Archive.
[2] UCI, machine learning repository, forest fires data set. https://archive.ics.uci.edu/ml/ datasets/forest+fires.

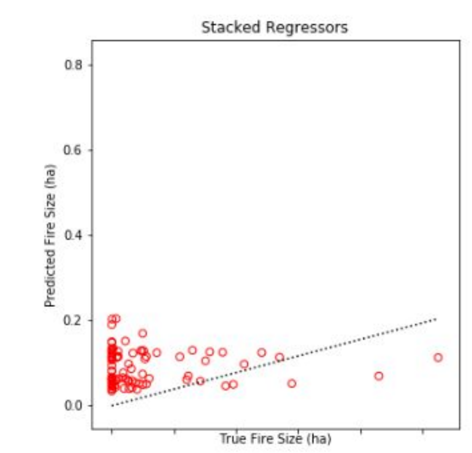
Models

- **Baseline**
  - Linear regression

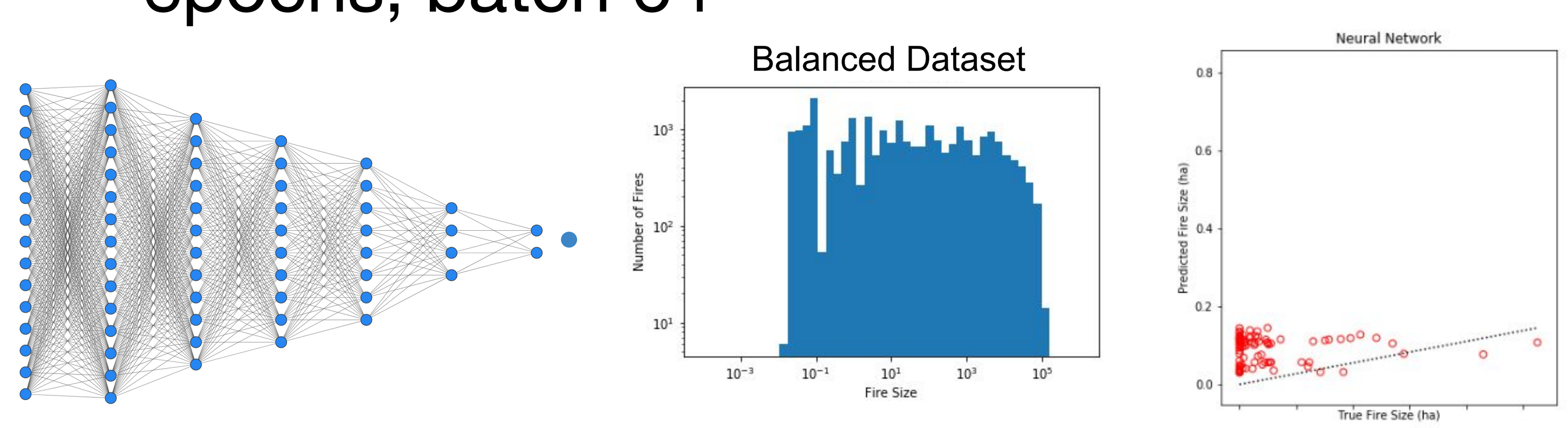
Normalized Prediction Plots



- **Other Models**
  - SVM
  - K Nearest Neighbors
  - Decision Tree
  - Gaussian Process regressor



- **Ensembling**
  - Stacked regressors
  - Random Forest
  - Adaboost
- **Neural Network**
  - Adam optimization, relu activation, 10 epochs, batch 64

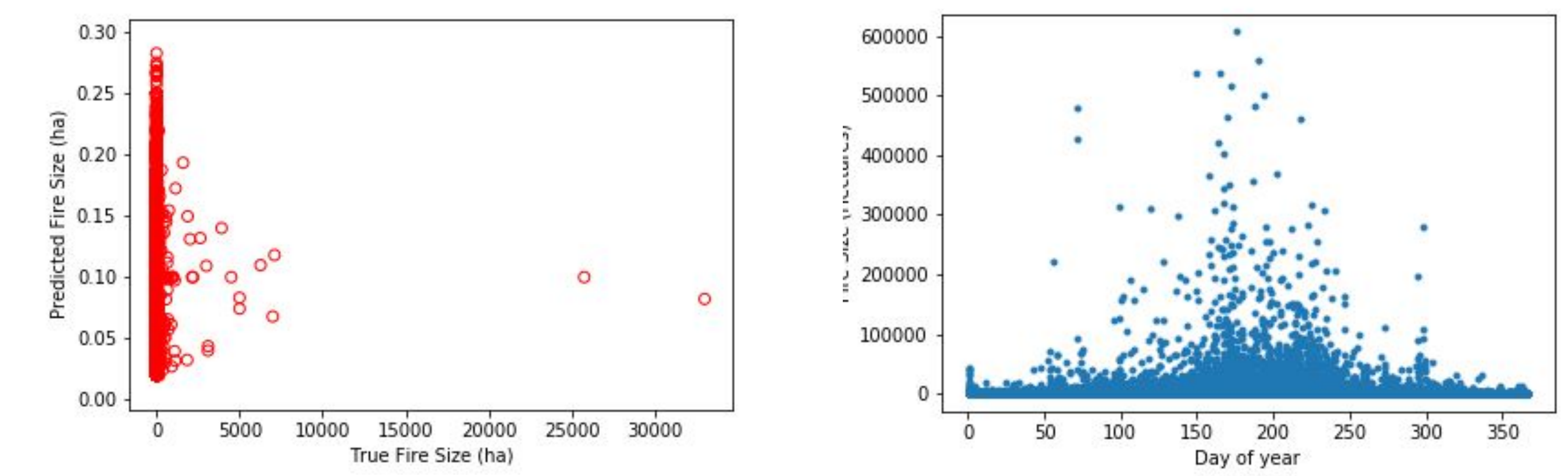


- **Model Tuning**
  - Learning rate/ Batch Size
  - SVM Kernel
  - Tree depth/ Neighbors
  - Loss functions
- **Evaluation Metrics**
  - Mean-Abs. Error (MAE)
  - Mean-Percent Err. (MAPE)
  - MAPE had systematically smaller predictions

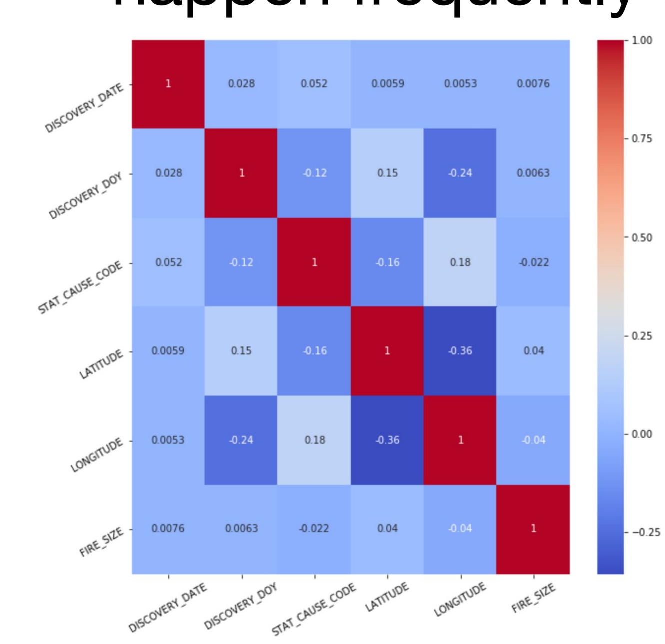
Results/Discussion

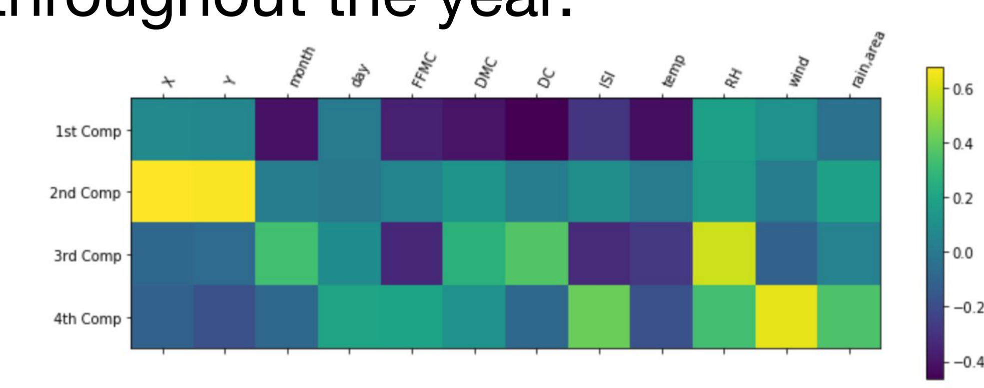
	Kaggle		Kaggle-Balanced		UCI	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
Linear Regression	171.12	82531.18	164.9	79895	15.547	869.165
SVM	52.25	50.91	35.57	41.0	6.334	77.711
Neural Network	87.29	82.51	36.12	43.3	8.264	347.628
K-nearest-neighbors	111.75	3190.5	36.91	43.39	15.53	639.88
Decision Tree	118.19	4119.8	36.9	42.9	31.5	995.8
Stacked Regressors	53	1326.9	36.92	43.39	9.45	64.54

- Balancing dataset improves performance
- In general, models perform worse on Kaggle due to data imbalance and fewer highly correlated features
- SVM, neural net, and stacked regressors perform best



- Large number of small fires makes prediction difficult: mega-fires occur mostly in summer, yet small fires happen frequently throughout the year.





PCA on UCI dataset. Strongest correlations with Long/Lat and weather features

Kaggle has weak correlations of features to fire size

Future Work

- Extend datasets to Include additional salient features
- Merge additional weather features with Kaggle dataset to tackle high variance issue
- Predict fire size class instead

Acknowledgements

We would like to thank our TA, Leo Mehr, for his advice and support throughout the project as well as the teaching staff for helpful comments and ideas.