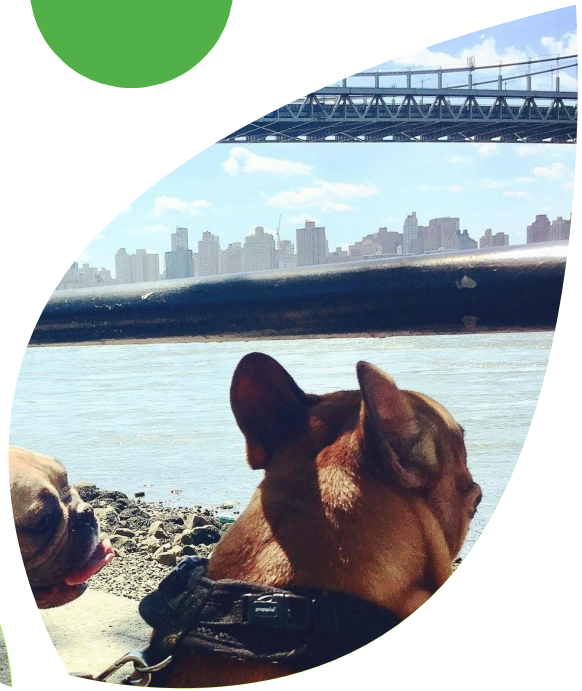


New York City Tree Health Classification

Benjamin Dean

Business Problem

- On which trees should efforts of the New York City Department of Parks and Recreation be concentrated?
- Which features most contribute to substandard tree health?

Introduction

Data cleaned and preprocessed before being passed to models

Decision Tree

Random Forest

eXtreme Gradient Boosting

Fit using default then tuned hyperparameters

Data

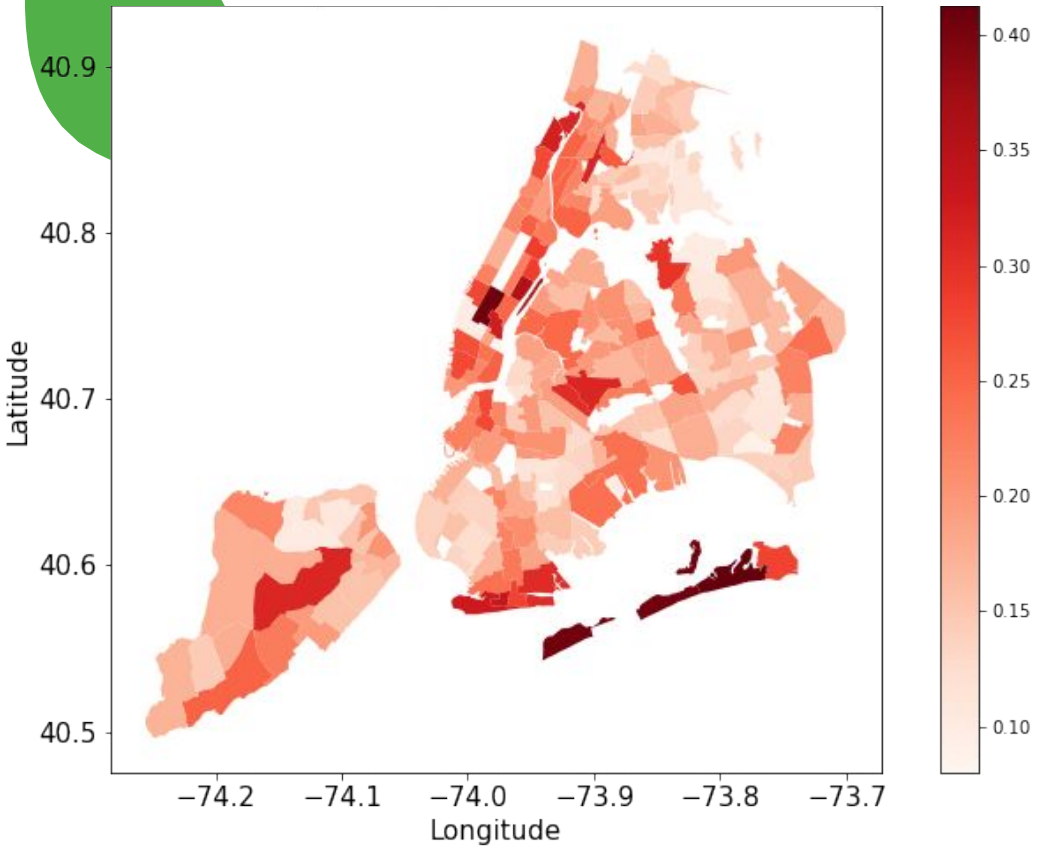
- 2015 Street Tree Census provided by NYC OpenData
- 683,788 curb adjacent trees

	health	tree_dbh	curb_loc	spc_latin	root_stone	trnk_light	steward	brch_shoe	borough	nta_name	latitude	longitude	census tract
65637	Good	12	OnCurb	Ulmus americana	No	No	None	No	Manhattan	Upper East Side-Carnegie Hill	40.774906	-73.965345	130.0
80378	Fair	33	OnCurb	Platanus x acerifolia	Yes	No	None	No	Brooklyn	Bay Ridge	40.630974	-74.035590	44.0
326730	Good	15	OnCurb	Gleditsia triacanthos var. inermis	Yes	No	None	No	Bronx	Co-op City	40.864951	-73.823924	302.0

- Target variable ‘health’ + 44 features

Libraries and Modules

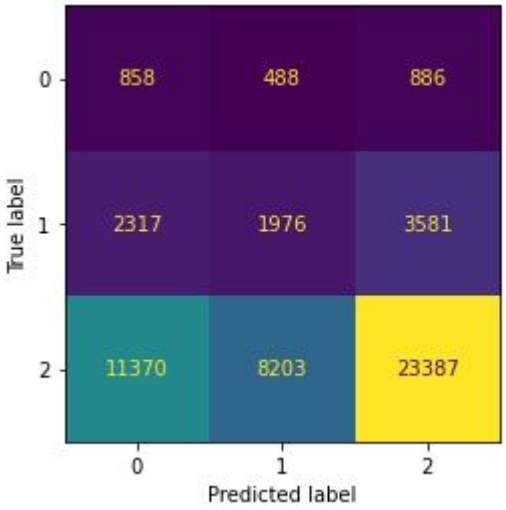
Percentage of Substandard Trees by Neighborhood



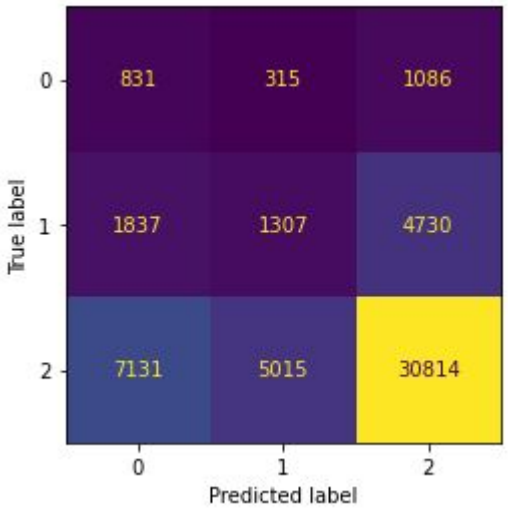
nta_name	percent_substandard	boro_name
Hammels-Arverne-Edgemere	0.412917	Queens
Breezy Point-Belle Harbor-Rockaway Park-Broad Channel	0.406119	Queens
Midtown-Midtown South	0.405357	Manhattan
Lenox Hill-Roosevelt Island	0.354402	Manhattan
West Brighton	0.334066	Brooklyn
Seagate-Coney Island	0.326500	Brooklyn
Murray Hill-Kips Bay	0.323299	Manhattan
Washington Heights North	0.320365	Manhattan
Marble Hill-Inwood	0.314030	Manhattan
Claremont-Bathgate	0.313946	Bronx

Decision Tree

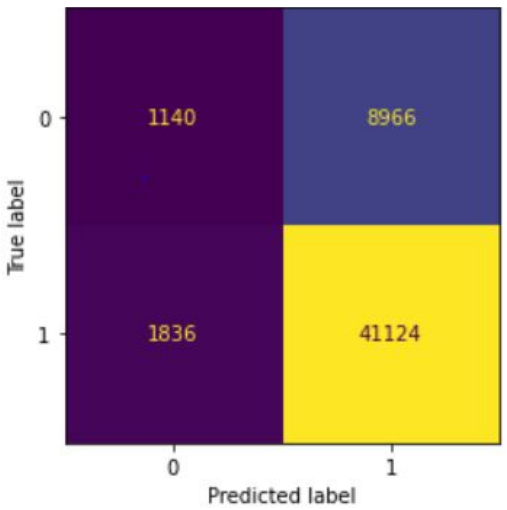
Balanced Ternary



Tuned Ternary

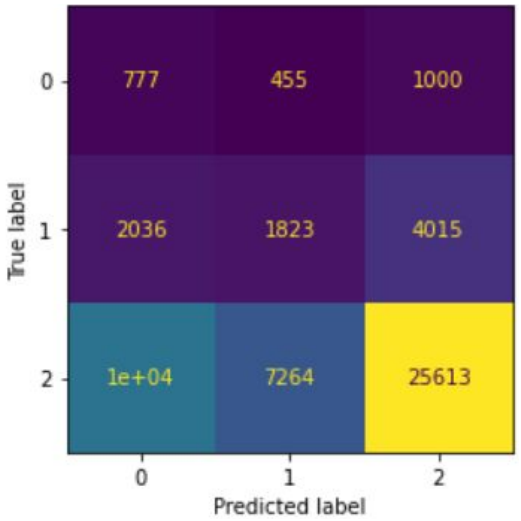


Baseline Binary

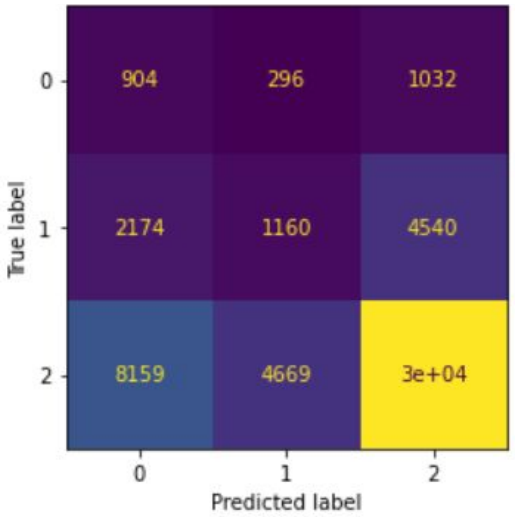


Random Forest

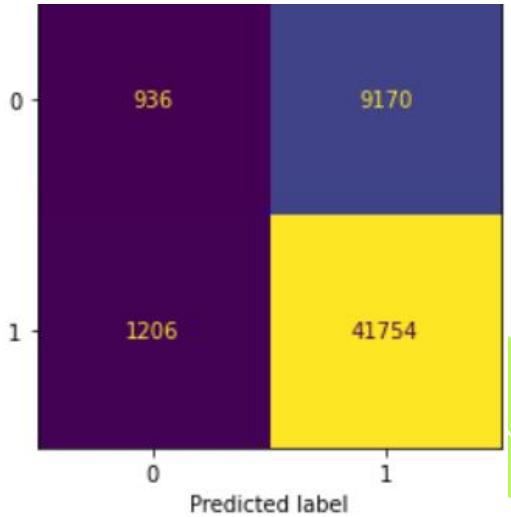
Balanced Ternary



Tuned Ternary

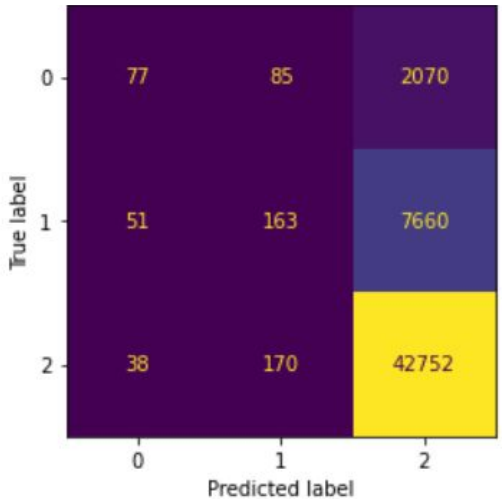


Baseline Binary

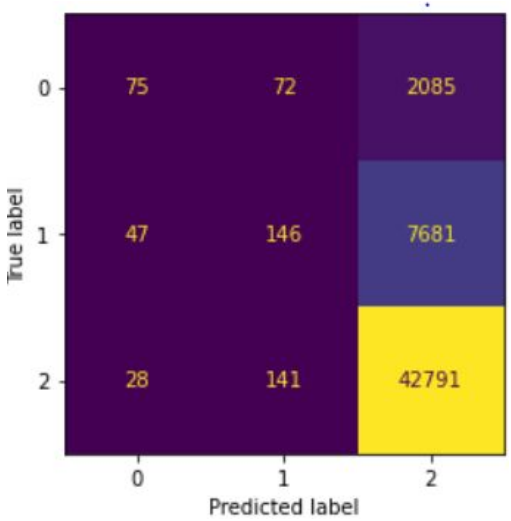


eXtreme Gradient Boost

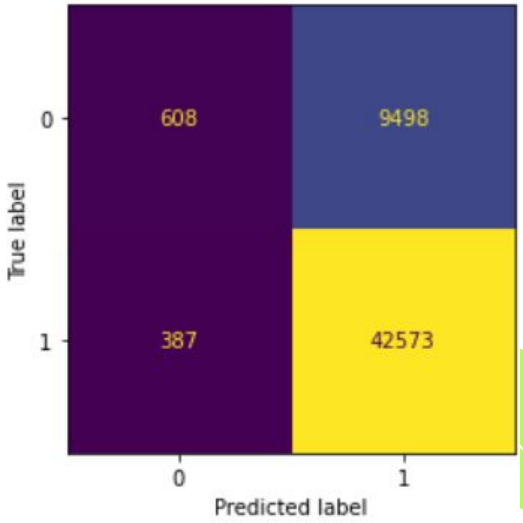
Balanced Ternary



Tuned Ternary

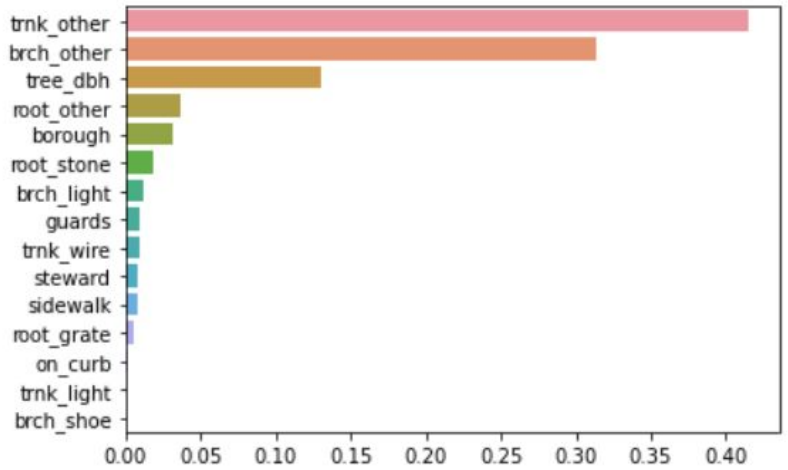


Baseline Binary



Conclusions and Actions

- Binary classifications perform better than ternary classifications
- Tuned XGBoost model produces most reliable predictions



- Prioritize trees with trunk problems (non-lights/rope/wires)
- Prioritize trees with branch problems (non-lights/shoes/wires)
- Allow trees to further mature before moving to street

Future Analysis

- Utilize Scikit-learn's pipeline module to more efficiently and cleanly work with data
- Difficulty with run time of programs led to limited executions of notebooks and reduced parameter searching

Thank you for reading

GitHub: https://github.com/Benjamin2817

Email: benjamin.a.dean.17@gmail.com

LinkedIn: TBA