Identify Vacant, Abandoned, and Disinvested (VAD) properties through Human-in-the-loop Machine Learning Model

Xiaofan Liang, Clio Andris, Brian Brainerd

FriendlyCitiesLab @ Georgia Tech

Georgia Tech

SAVANNAH
savannahga.gov

Why Should We Identify Vacant, Abandoned, and Disinvested (VAD) Properties?

A blighted/VAD property in Savannah

VAD properties are often associated with

Property Code Violation, Crime, Tax Delinquency, Vacancy, and Public Nuisances

City of Savannah’ VAD properties are growing

The number of properties with code violations has grown 160% from 2014 to 2019. 1404 properties have more than three years of tax delinquent history by 2019, likely due to abandonment and neglect.

VADs bleed public tax money

The funds are largely unrecoverable costs incurred for addressing overgrown grass, litter, illegal dumping, securing open structures and demolishing properties. It also includes lost property tax revenue.

VADs negatively impact marginalized community

Neighborhoods with a high presence of VAD properties are associated with **low school quality** (Sun et al, 2019), **high crime rates** (Branas, Bubin, & Guo, 2012), **higher male unemployment rate** (Appel et al, 2014), and **slower growth in property sales price** (Gilreath, 2013). These neighborhoods are also more likely to be home to **low-income and African American households** (Sun et al, 2019; Silverman et al, 2013) and suffer from declining home ownership and pessimistic perceptions of neighborhood trajectories

Why is identifying VAD properties difficult?

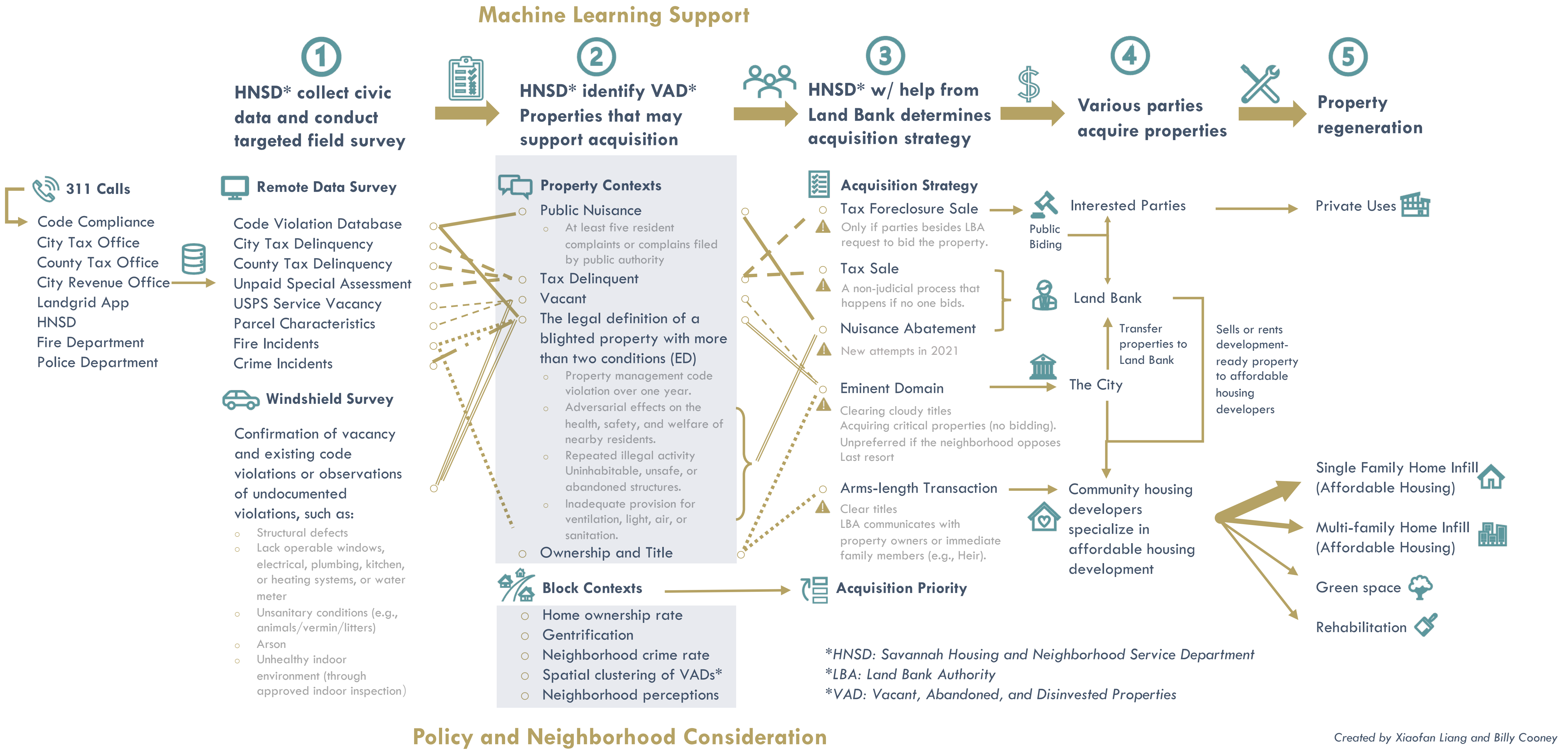
Many Factors Contribute to VADs for Acquisition

- **Parcel level:** Civic Data, Property Attributes
- **Neighborhood level:** Neighborhood value, Social Capital, Home ownership rate
- **City level:** Spatial clustering
- **Policy level:** Minority Neighborhood, Opportunity Zones in Housing Markets, Local Legal Requirements for Blight Acquisition.

Current Process is not Scalable and Biased

- Manual, relying on field survey
- Time-consuming
- Biased toward particular blocks or neighborhoods

Urban Property Regeneration
for Savannah, GA



Predict VAD Properties through Machine Learning

Current Process

HNSD collect civic data and conduct field surveys → HNSD identify contexts of properties that may support acquisition → HNSD w/ help from Land Bank determines acquisition strategy → Various parties acquire properties → Property regeneration

A Vision of Machine-Human Collaboration in a Spatial Decision Support System

Civic technology software pulls and prepares data from multiple civic data sources → Trained machine learning model predicts VAD candidates based on parcel-level data and give each candidate a VAD label → HNSD uses civic tech software interface to visualize the candidates and examine the spatial, policy, and market context for priority (e.g., gentrification, clustering, targeted blocks) → HNSD w/ help from Land Bank determines acquisition strategy

Citizen 311 calls

Community Inputs

Predicting VAD Properties through Human-in-the-loop Machine Learning Workflow

Iterate

1 Data Cleaning

Clean data from multiple civic data sources and create a parcel attribute table

2 QA/QC Data

Create VAD scores for each parcel and iterates with experts to filter out anomalies and targets outside of policy

3 Select Features

Iterate with experts to select variables to predict VADs and map out distribution of each input.

Iterate

7 Machine Learning

Train the model on labeled samples to predict VAD property, experiment on model choices, and analyze feature importance

6 Label Consistency

Fit decision trees to labeled samples to correct for inconsistent labels across multiple experts

5 Expert Labeling

Ask experts to label the sampled parcels as VAD candidates or not.

4 Sampling

User uncertainty and diversity sampling to filter 300 sample data from candidates for expert labeling

8 Model Robustness

Compare ML predicted VAD candidates with human found VAD properties in a target list and test how well the prediction still hold true in 2021

9 Bias Analysis

Assess where do ML outperform or underperform than human experts, as well as biases in input data

All data are up to 2019
VAD: Vacant, Abandoned, Disinvested Properties

4 Sampling – Pick Properties that Maximize Human Labeling Efficiency

Machine needs labeled data to learn prediction. However, we do not have a labeled dataset of VAD properties.

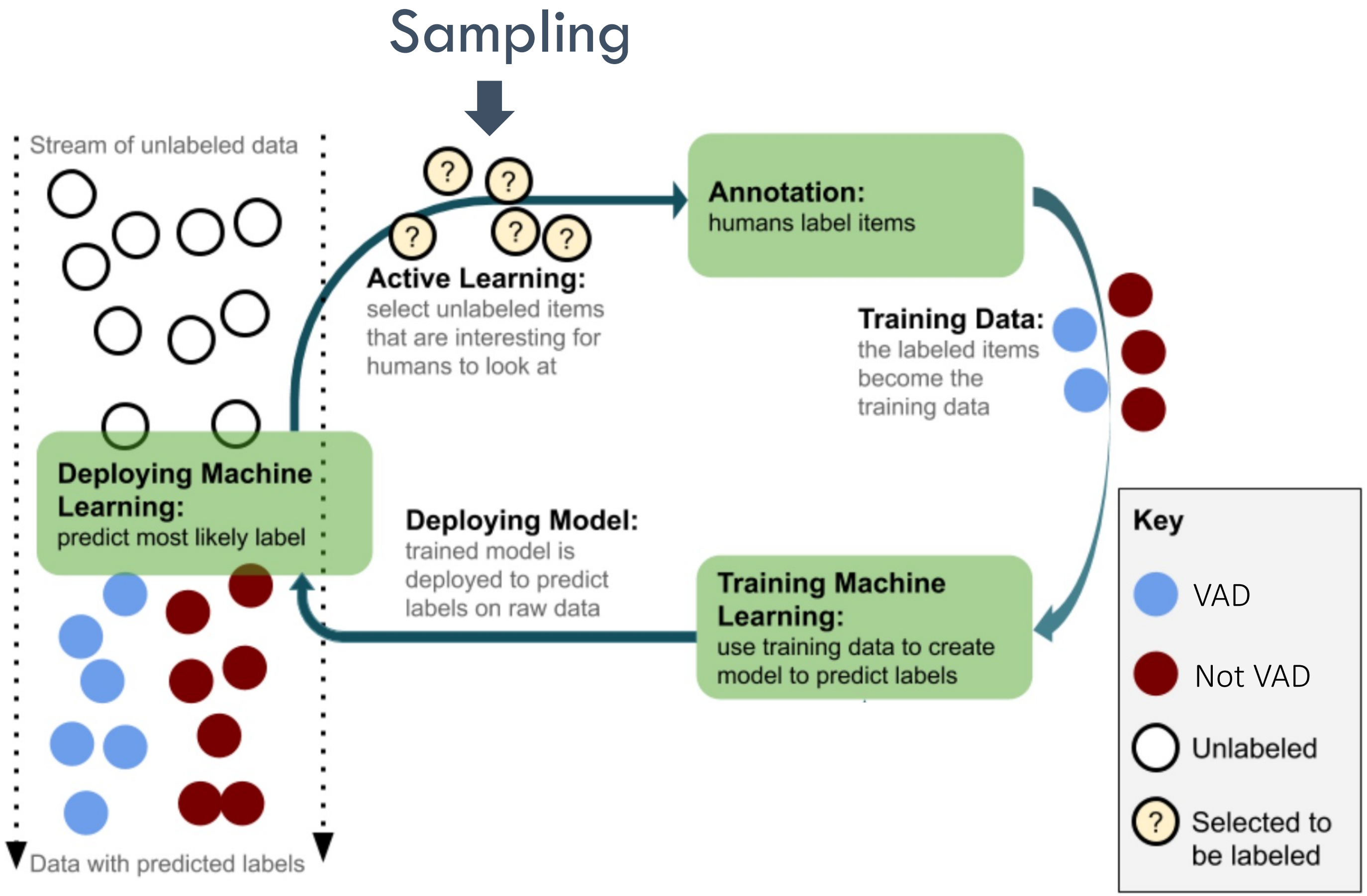
To maximize human labeling efficiency, we want to give housing experts samples that are diverse (in geography and characteristics) and difficult to judge.

We want to sample 300 properties and asked experts to label them as

- VAD
- Not VAD.

We deployed three rounds of sampling

- n=200: Uncertainty + Diversity Sampling
- n=100: Randomly sampled

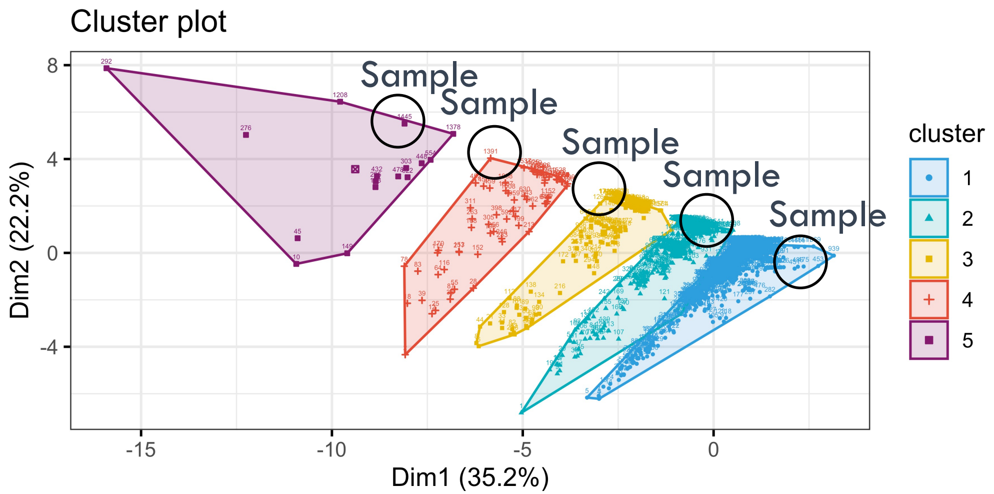


All data are up to 2019
VAD: Vacant, Abandoned, Disinvested Properties

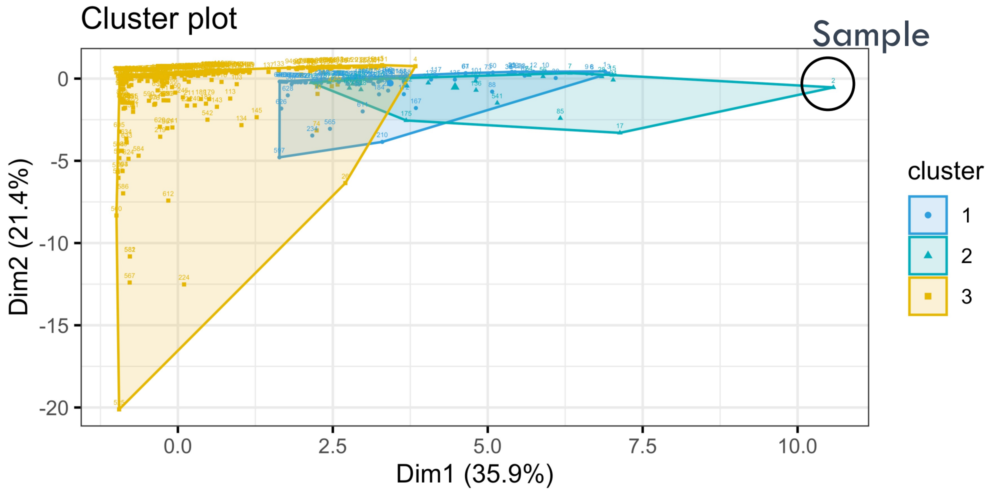
4 Sampling – Examples of Diversity Sampling

Samples represent different feature clusters

Structure

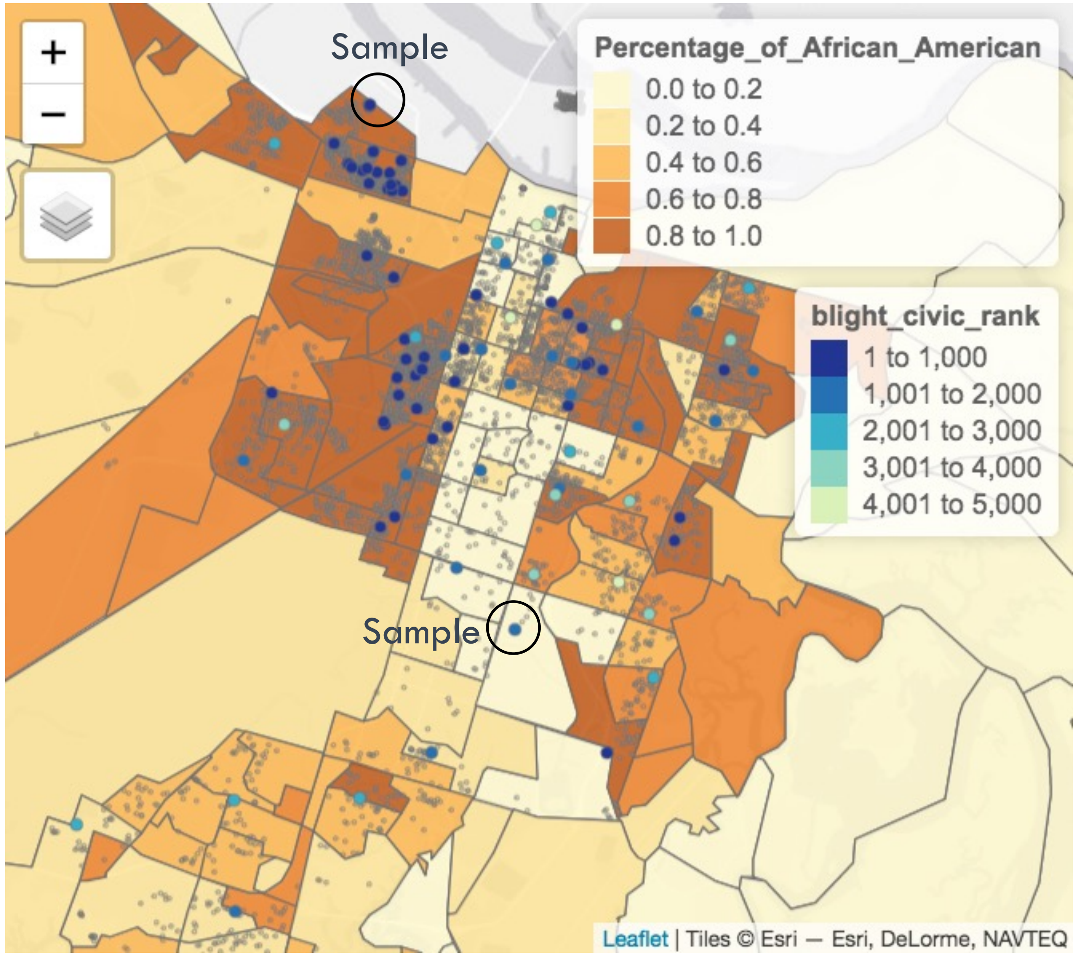


Land



K-means clustering on crime, code, tax, drug, vacancy

Samples distributed across neighborhoods

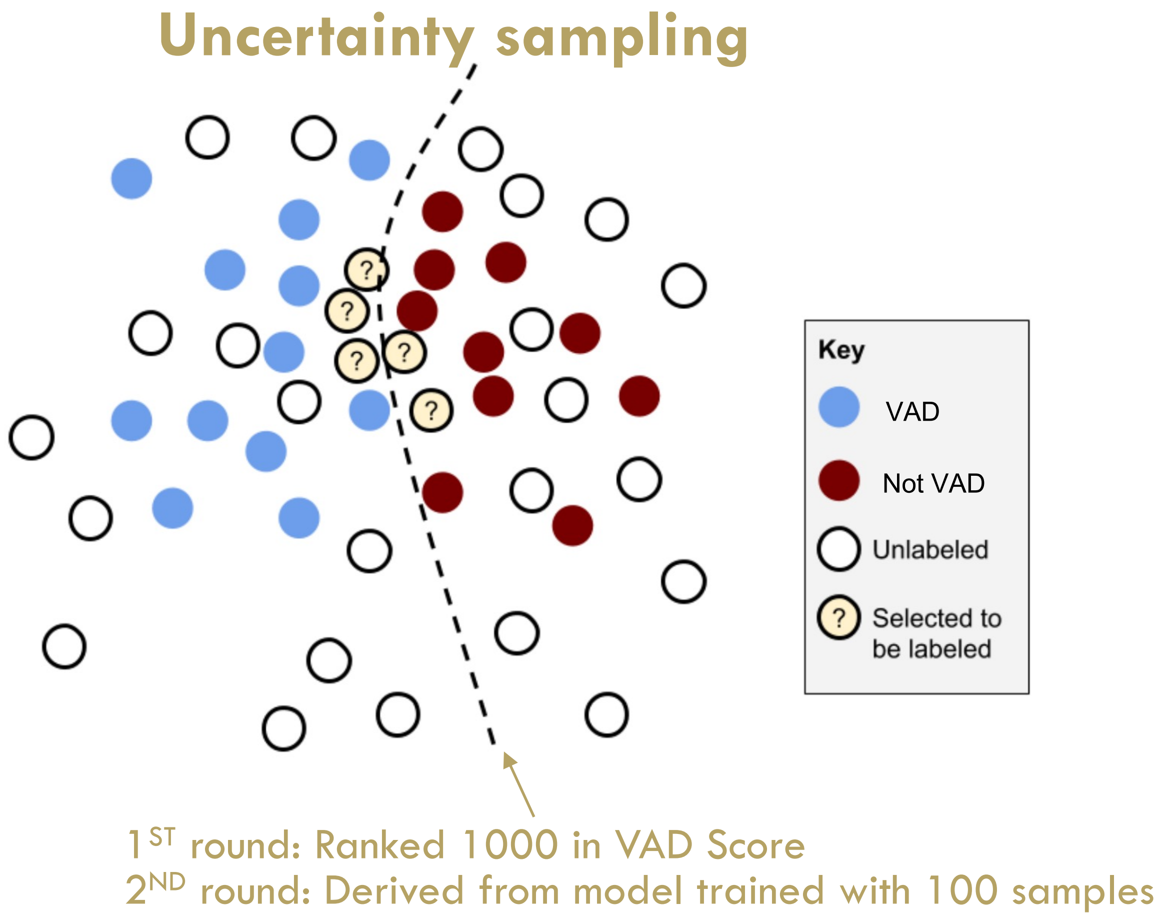


Selected samples distributed on the racial layer

4 Sampling – Examples of Uncertainty Sampling

Sample more properties that are not represented sufficiently, which will impact the results of cross validation, such as

- More Land (only 9 land labeled as NOT VAD in 1st round of sampling)
- More properties with rare events (drug crime and fire)
- More properties that have few instances in VAD type that are prone to errors
- More properties that are deemed "uncertain" by the ML model (trained with 1st round samples)



5 Expert Labeling – Excel Interface

One column contains information for one parcel

Civic Data

Parcel Attributes

	A	B	C
1	pid	2002710010	2005517003
2	crime_cnt_weighted	0	0
3	crime_cnt_by_2019	0	0
4	crime_cnt_recent_year	NA	NA
5	drug_n	0	0
6	drug_recent_yr	NA	NA
7	active_code_cnt_weighted	3.35	3.25
8	code_cnt_by_2019	13	6
9	active_code_cnt_by_2019	12	6
10	code_recent_year	2019	2019
11	fire_cnt_by_2019	0	0
12	fire_recent_year	NA	NA
13	total_Delin	1050.93	5333.79
14	tax_delin_years	3	2
15	Special_Assess_Tax	885	5333.79
16	avg_V	1	1
17	parcel_type	Structure	Structure
18	current_use_sub_type	SINGLE FAMILY	SINGLE FAMILY
19	land_size	0.07	0.16
20	year_built	1953	1923
21	sales_cnt	6	4
22	qualified_sales_cnt	0	0
23	unqualified_sales_cnt	6	4
24	sales_last_year	2019	2019
25	CAMA_TOTAL	19.6	48.5
26	Growth_5yrs_14_19	-0.51	0.62
27	neigh_median_CAMA_TOTAL	22.2	74
28	Label		
29	Comments	VAD	
30		Not VAD	

samples0618 | Dictionary | +

Experts label and write comments on each parcel

6 Label Consistency Across Multiple Experts

- Fit a decision tree to 300 labeled samples
- Visualize the decision tree
- Go through splits and see if they make sense
- Check bottom branches and branches that take multiple steps to split a small number of VAD property.
- Compare properties under the same sub-branch: if their conditions are similar but labeled differently, do they justify a correction?

For Example...

7 Machine Learning Drop-Column Feature Importance

	All Features				Reduced Features			
	200 samples		300 samples		200 samples		300 samples	
	Land	Structure	Land	Structure	Land	Structure	Land	Structure
Feature Importance Score and Rank								
Weighted Crime Count	**2.50 (R2)**	**5.07 (R1)**	**2.86 (R2)**	**3.59 (R2)**	**3.75 (R3)**	**6.70 (R1)**	**2.86 (R3)**	**4.62 (R2)**
Weighted Drug Count	-1.18*	-0.87*	0.95*	0.00*	2.57*	-1.70*	**3.81 (R2)**	-1.54
Weighted Active Code Cases	-1.18*	**1.74 (R3)**	0.00*	**2.05 (R4)**	**4.93 (R1)**	**3.37 (R3)**	**1.91 (R4)**	**1.03 (R3)**
Weighted Fire Count	-1.18	0.83	-2.86*	1.03*	NA	NA	NA	NA
Tax Delinquency & Delinquent Years	**3.53 (R1)**	**3.37 (R2)**	**8.57 (R1)**	**8.72 (R1)**	**4.85 (R2)**	**5.00 (R1)**	**13.33 (R1)**	**5.64 (R1)**
Special Assessment Tax Pct	-4.85*	0.87	0.95	**2.56 (R3)**	-3.60	1.63*	-2.86	**1.03 (R3)**
Vacancy Probability	-2.35*	0.91*	-0.95*	1.03*	NA	NA	NA	NA
Property Value	-3.60*	0.04*	-0.95*	0.51*	-2.43	-0.91*	0.00*	-1.54
Land Size	-3.53*	0.00*	-0.95*	-1.03	NA	NA	NA	NA
Qualified Sales	-3.60*	0.04*	0.95*	0.00*	NA	NA	NA	NA
Unqualified Sales	-3.60*	0.04	-0.95*	1.03*	NA	NA	NA	NA
Year Last Sold	-3.60*	0.04	-0.95*	-0.51*	NA	NA	NA	NA
Growth Rate	-3.60*	-0.87*	-0.95*	1.54*	NA	NA	NA	NA
Median Neigh PV	-3.60*	0.00*	0.00*	1.03*	NA	NA	NA	NA
Evaluation Metrics (%)								
Cross Validation Accuracy	87.94	90.65	91.43	87.69	91.62	91.45	95.24	87.69
OOB score	91.46	91.53	88.57	87.18	92.68	91.53	90.48	88.21

* Indicates that the number fluctuate above or below and equal to zero depending on the random state of the random forest algorithm and thus deemed uncertain. The unit of the number is %. R in parenthesis indicates the ranking of features that have stable contributes to at least 1% drop-column importance.

Model Settings

- All features include all feature in the feature column
- Reduced features only contain crime, drug, code, tax, SA tax, and PV
- 200 vs. 300 human labeled samples
- Land vs. Structure

Model Features

- Variables used to predict VAD properties

Drop-column importance value X on a feature means that

- model trained without this feature dropped the accuracy by X%.
- X > 0 means positive contribution to model accuracy
- X < 0 means negative contribution to model accuracy
- R indicates ranking
- Bolded highlight features that has stable contribution to model accuracy.

Value with * can be ignored: they are not stable and vary due to stochasticity.

- Random forest model involves bootstrapping (i.e., random sampling of rows) to form decision trees. Value is labeled with * if it fluctuates above or equal or below zero at different random state of random forest model.

Cross validation and OOB score are complementary measures of accuracy

- The value can be interpreted as X% of samples not being trained on are predicted correctly.

7 Feature Importance w/ Different Model Settings

Feature Importance Table

	All Features				Reduced Features			
	200 samples		300 samples		200 samples		300 samples	
	Land	Structure	Land	Structure	Land	Structure	Land	Structure
Feature Importance Score and Rank								
Weighted Crime Count	**2.50 (R2)**	**5.07 (R1)**	**2.86 (R2)**	**3.59 (R2)**	**3.75 (R3)**	**6.70 (R1)**	**2.86 (R3)**	**4.62 (R2)**
Weighted Drug Count	-1.18*	-0.87*	0.95*	0.00*	2.57*	-1.70*	**3.81 (R2)**	-1.54
Weighted Active Code Cases	-1.18*	**1.74 (R3)**	0.00*	**2.05 (R4)**	**4.93 (R1)**	**3.37 (R3)**	**1.91 (R4)**	**1.03 (R3)**
Weighted Fire Count	-1.18	0.83	-2.86*	1.03*	NA	NA	NA	NA
Tax Delinquency & Delinquent Years	**3.53 (R1)**	**3.37 (R2)**	**8.57 (R1)**	**8.72 (R1)**	**4.85 (R2)**	**5.00 (R1)**	**13.33 (R1)**	**5.64 (R1)**
Special Assessment Tax Pct	-4.85*	0.87	0.95	**2.56 (R3)**	-3.60	1.63*	-2.86	**1.03 (R3)**
Vacancy Probability	-2.35*	0.91*	-0.95*	1.03*	NA	NA	NA	NA
Property Value	-3.60*	0.04*	-0.95*	0.51*	-2.43	-0.91*	0.00*	-1.54
Land Size	-3.53*	0.00*	-0.95*	-1.03	NA	NA	NA	NA
Qualified Sales	-3.60*	0.04*	0.95*	0.00*	NA	NA	NA	NA
Unqualified Sales	-3.60*	0.04	-0.95*	1.03*	NA	NA	NA	NA
Year Last Sold	-3.60*	0.04	-0.95*	-0.51*	NA	NA	NA	NA
Growth Rate	-3.60*	-0.87*	-0.95*	1.54*	NA	NA	NA	NA
Median Neigh PV	-3.60*	0.00*	0.00*	1.03*	NA	NA	NA	NA
Evaluation Metrics (%)								
Cross Validation Accuracy	87.94	90.65	91.43	87.69	91.62	91.45	95.24	87.69
OOB score	91.46	91.53	88.57	87.18	92.68	91.53	90.48	88.21

* Indicates that the number fluctuate above or below zero depending on the random state of the random forest algorithm and thus deemed uncertain. The unit of the number is %. R in parenthesis indicates the ranking of features that have stable contributes to at least 1% drop-column importance.

Insight #1

Tax, Crime, Code ++++

- Overall, **tax delinquency and delinquent years** is **the most important** variable (or combo variable) in predicting VAD properties across Land and Structure, **second with Crime,** and **third with Code Cases.**

7 Feature Importance w/ Different Model Settings

Feature Importance Table

	All Features				Reduced Features			
	200 samples		300 samples		200 samples		300 samples	
	Land	Structure	Land	Structure	Land	Structure	Land	Structure
Feature Importance Score and Rank								
Weighted Crime Count	**2.50 (R2)**	**5.07 (R1)**	**2.86 (R2)**	**3.59 (R2)**	**3.75 (R3)**	**6.70 (R1)**	**2.86 (R3)**	**4.62 (R2)**
Weighted Drug Count	-1.18*	-0.87*	0.95*	0.00*	2.57*	-1.70*	**3.81 (R2)**	-1.54
Weighted Active Code Cases	-1.18*	**1.74 (R3)**	0.00*	**2.05 (R4)**	**4.93 (R1)**	**3.37 (R3)**	**1.91 (R4)**	**1.03 (R3)**
Weighted Fire Count	-1.18	0.83	-2.86*	1.03*	NA	NA	NA	NA
Tax Delinquency & Delinquent Years	**3.53 (R1)**	**3.37 (R2)**	**8.57 (R1)**	**8.72 (R1)**	**4.85 (R2)**	**5.00 (R1)**	**13.33 (R1)**	**5.64 (R1)**
Special Assessment Tax Pct	-4.85*	0.87	0.95	**2.56 (R3)**	-3.60	1.63*	-2.86	**1.03 (R3)**
Vacancy Probability	-2.35*	0.91*	-0.95*	1.03*	NA	NA	NA	NA
Property Value	-3.60*	0.04*	-0.95*	0.51*	-2.43	-0.91*	0.00*	-1.54
Land Size	-3.53*	0.00*	-0.95*	-1.03	NA	NA	NA	NA
Qualified Sales	-3.60*	0.04*	0.95*	0.00*	NA	NA	NA	NA
Unqualified Sales	-3.60*	0.04	-0.95*	1.03*	NA	NA	NA	NA
Year Last Sold	-3.60*	0.04	-0.95*	-0.51*	NA	NA	NA	NA
Growth Rate	-3.60*	-0.87*	-0.95*	1.54*	NA	NA	NA	NA
Median Neigh PV	-3.60*	0.00*	0.00*	1.03*	NA	NA	NA	NA
Evaluation Metrics (%)								
Cross Validation Accuracy	87.94	90.65	91.43	87.69	91.62	91.45	95.24	87.69
OOB score	91.46	91.53	88.57	87.18	92.68	91.53	90.48	88.21

* Indicates that the number fluctuate above or below zero depending on the random state of the random forest algorithm and thus deemed uncertain. The unit of the number is %. R in parenthesis indicates the ranking of features that have stable contributes to at least 1% drop-column importance.

Insight #2, #3

Increase labeled samples from 200 to 300 increase accuracy slightly for Land but not Structure

- The model may be slightly overfitted on 200 samples for structure. Overall, the accuracy seems to be stable without the need of additional samples.

Reducing features slightly improve model accuracy

- It is possible that other features provide more noise to prediction than helpful information. CV becomes more important on Land after reducing features.

7 Final Model

Feature Importance Table

	All Features				Reduced Features			
	200 samples		300 samples		200 samples		300 samples	
	Land	Structure	Land	Structure	Land	Structure	Land	Structure
Feature Importance Score and Rank								
Weighted Crime Count	**2.50 (R2)**	**5.07 (R1)**	**2.86 (R2)**	**3.59 (R2)**	**3.75 (R3)**	**6.70 (R1)**	**2.86 (R3)**	**4.62 (R2)**
Weighted Drug Count	-1.18*	-0.87*	0.95*	0.00*	2.57*	-1.70*	**3.81 (R2)**	-1.54
Weighted Active Code Cases	-1.18*	**1.74 (R3)**	0.00*	**2.05 (R4)**	**4.93 (R1)**	**3.37 (R3)**	**1.91 (R4)**	**1.03 (R3)**
Weighted Fire Count	-1.18	0.83	-2.86*	1.03*	NA	NA	NA	NA
Tax Delinquency & Delinquent Years	**3.53 (R1)**	**3.37 (R2)**	**8.57 (R1)**	**8.72 (R1)**	**4.85 (R2)**	**5.00 (R1)**	**13.33 (R1)**	**5.64 (R1)**
Special Assessment Tax Pct	-4.85*	0.87	0.95	**2.56 (R3)**	-3.60	1.63*	-2.86	**1.03 (R3)**
Vacancy Probability	-2.35*	0.91*	-0.95*	1.03*	NA	NA	NA	NA
Property Value	-3.60*	0.04*	-0.95*	0.51*	-2.43	-0.91*	0.00*	-1.54
Land Size	-3.53*	0.00*	-0.95*	-1.03	NA	NA	NA	NA
Qualified Sales	-3.60*	0.04*	0.95*	0.00*	NA	NA	NA	NA
Unqualified Sales	-3.60*	0.04	-0.95*	1.03*	NA	NA	NA	NA
Year Last Sold	-3.60*	0.04	-0.95*	-0.51*	NA	NA	NA	NA
Growth Rate	-3.60*	-0.87*	-0.95*	1.54*	NA	NA	NA	NA
Median Neigh PV	-3.60*	0.00*	0.00*	1.03*	NA	NA	NA	NA
Evaluation Metrics (%)								
Cross Validation Accuracy	87.94	90.65	91.43	87.69	91.62	91.45	95.24	87.69
OOB score	91.46	91.53	88.57	87.18	92.68	91.53	90.48	88.21

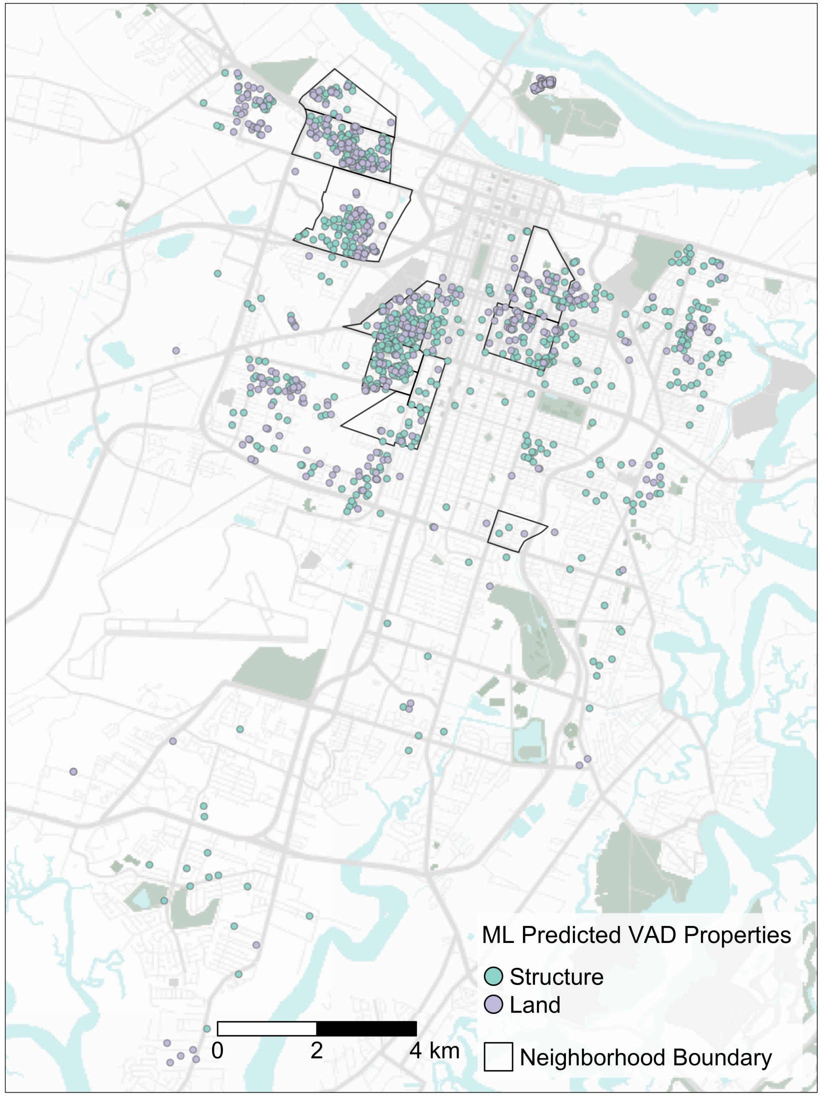
* Indicates that the number fluctuate above or below zero depending on the random state of the random forest algorithm and thus deemed uncertain. The unit of the number is %. R in parenthesis indicates the ranking of features that have stable contributes to at least 1% drop-column importance.

We adopted this model for generating prediction labels for properties

8 Robustness: ML prediction vs. Human target list

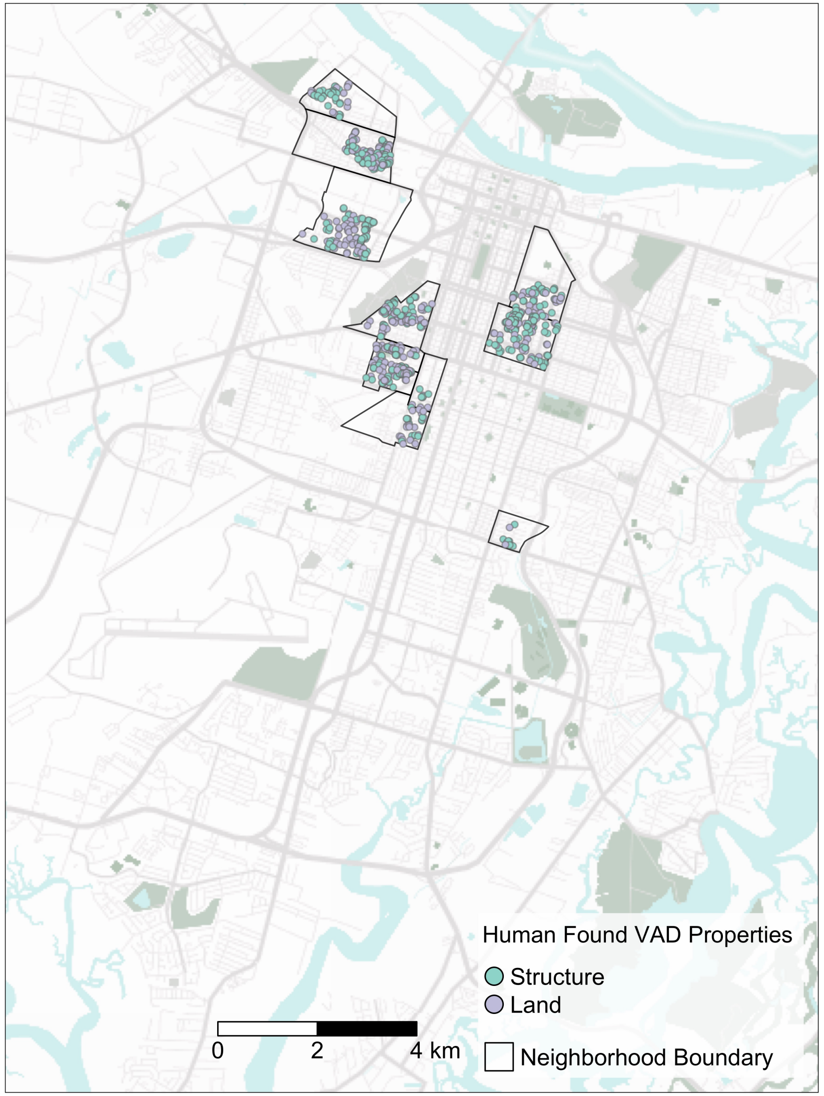
ML prediction

More spatially diverse



n=1234

Concentrate in clustered neighborhoods



n=693

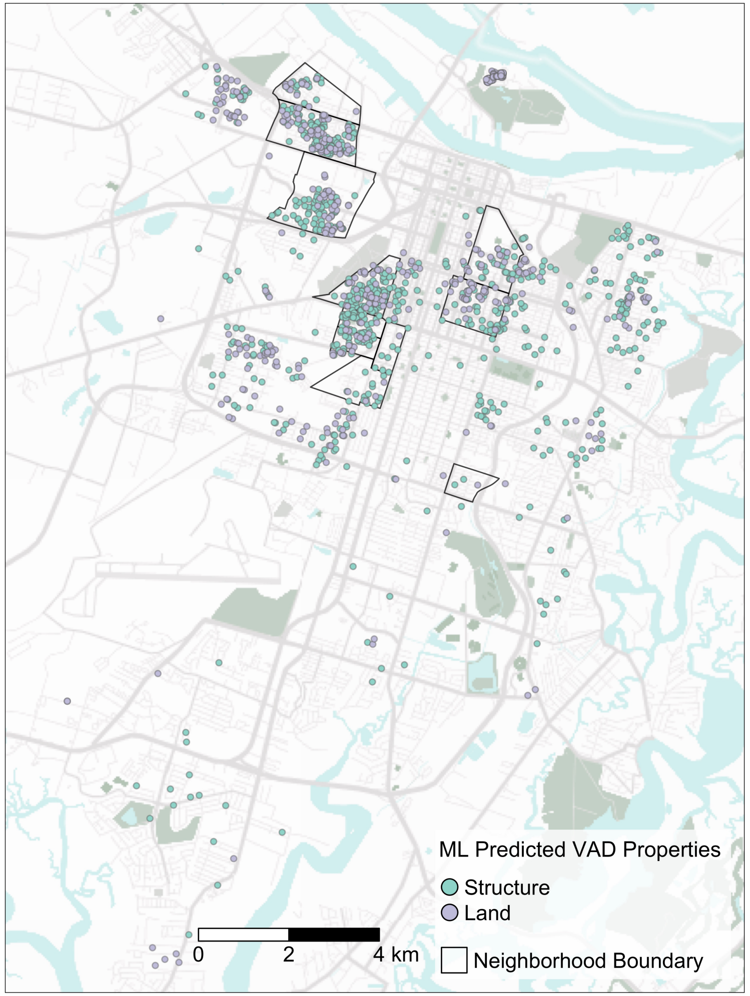
Human Found

286 (41%) did not even meet basic requirements to be ML candidates: no vacancy, and no CV, Fire, TD, or drug records, and/or have flood risk. They may be listed as target due to visual vacancy or data not caught up with parcel's current conditions.

Excluding those 286 parcels, **72%** of human found targets are also predicted as VAD by ML

8 Robustness – ML prediction vs. Human target list

ML predicted



n=1234

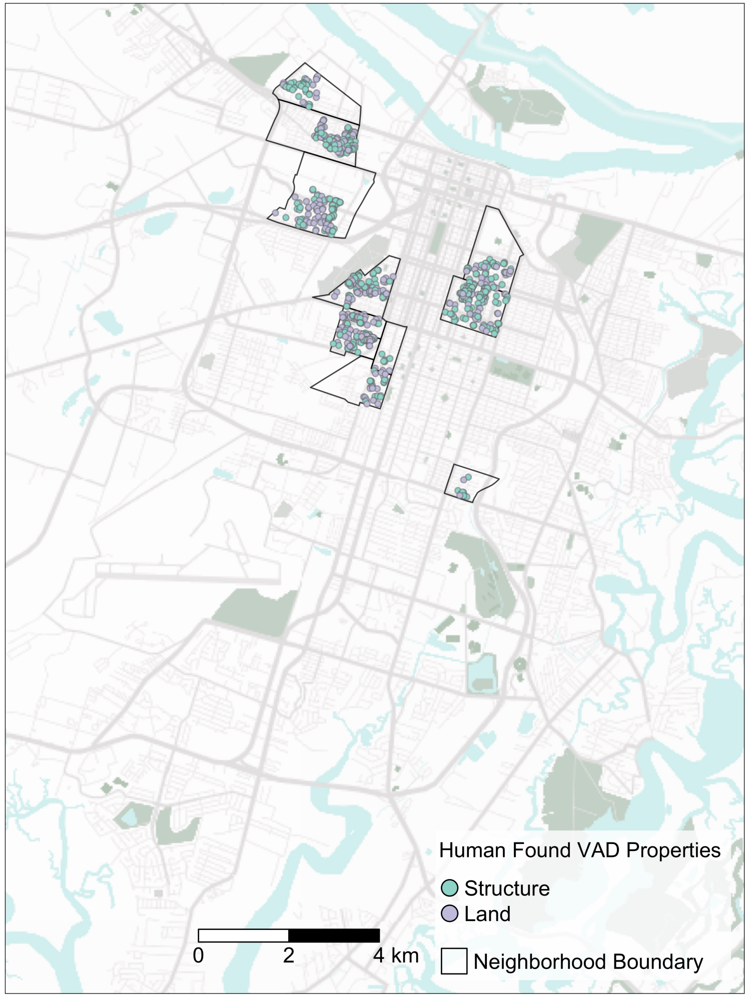
Land:		**Land:**
CR: **8.1%**	>	CR: 6%
TD: **100%**		TD: 67.5%
CV: **29.7%**		CV: 24.4%
LPV: 91%	<	LPV: **97.5%**
N=468		N=197
Structure:		**Structure:**
CR: **44.6%**	>	CR: 27%
TD: **98%**		TD: 88%
CV: 33.8%	<	CV: **43.3%**
LPV: 68%		LPV: **72%**
N= 766		N=210

Human generated list features higher percentages of **property with CV and LPV** in Structure (LPV only for Land).

This may be because that windshield survey focused more on visual cues (e.g., CV) and LPV neighborhoods.

CR: crime
TD: tax delinquency
CV: code violations
LPV: low property values

Human generated target list



n=693
(only 407 that meet minimum candidate requirements are being compared)

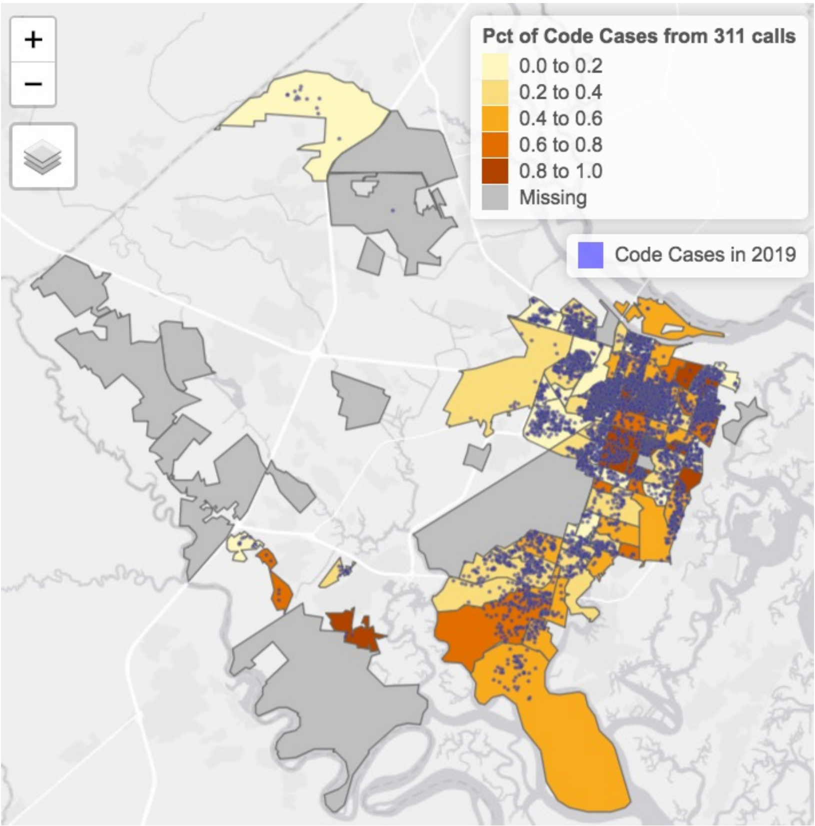
8 Robustness: Prediction in 2019 still true in 2021?

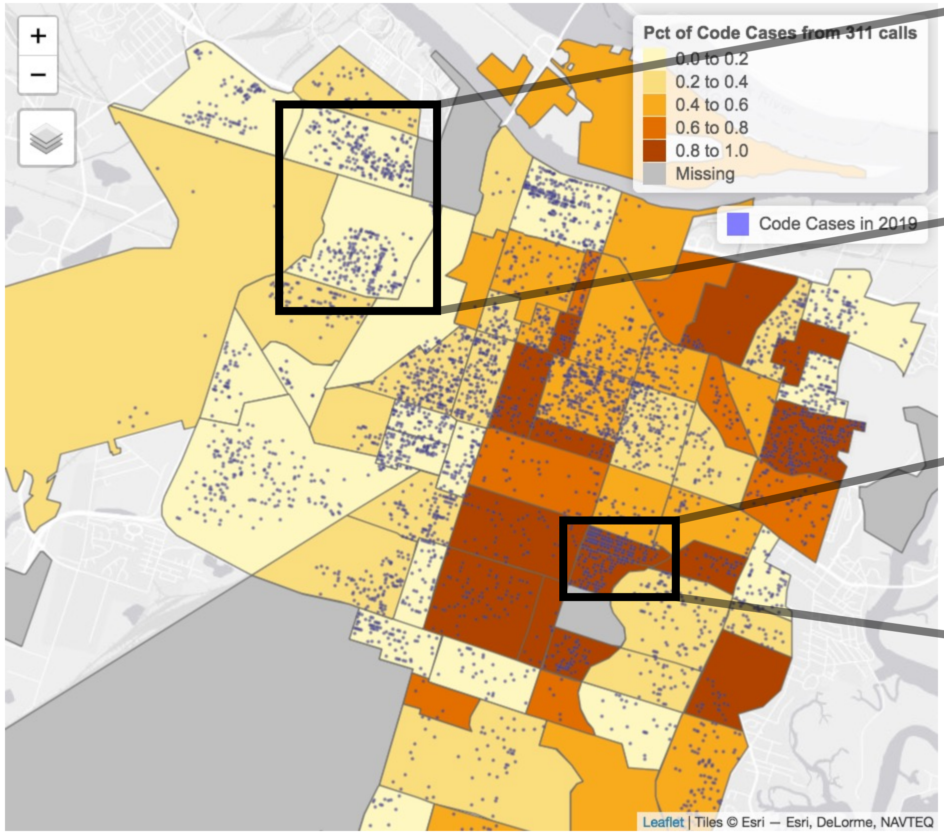
Our prediction is based on 2019 data, and now is 2021. How many of the predicted labels are still correct?

- We found that **66%** of predicted label in these 100 random samples are still correct in 2021.
- **20%** change from From Not VAD to VAD: properties may have evolving conditions in two years that qualify them to be VAD, or due to mislabeling that can be corrected with more contexts.
- **13%** change from VAD to Not VAD: properties may have already been intervened (e.g., went through tax sales), or due to mislabeling that can be corrected with more contexts.

9 Bias: Are there neighborhoods where data are less reliable?

Low pct of code cases from SAV311 means that residents do not actively provide leads for code compliance officer to find active code violations. These neighborhoods are particularly vulnerable in data bias, relying on code compliance officers to proactively survey the neighborhoods (if not, then the problems are underestimated).





<20% of code cases in West Savannah and Carver Heights come from 311 calls

>80% of code cases at Edgemere and Sackville come from 311 calls.

Conclusions

We produced …

- A machine learning model that reaches 87-95% accuracy at predicting VAD land or structures
- A human-in-the-loop process that involves
  - Regular presentation/feedback
  - Active sampling
  - Human labeling
  - Cross-check label consistency
  - Interpretable ML outcomes
  - Robustness comparison with human targets
  - Bias identification
- This model will be deployed and integrated with a third-party civic technology company's visual dashboard to form a spatial decision support system to guide property regeneration efforts in Savannah.

We learned that …

- Human-in-the-loop is important -> TRUST
- Tax, crime, and code are important in identifying VAD properties
- ML can generate more spatially diverse predictions that focus more on tax and crime compared with human
- Constant data integrations are required to keep the prediction up to date.
- Some neighborhoods are more susceptible to data bias due to low participation in 311 calls

Questions/Feedback can contact xiaofan.l@gatech.edu