Predict Post-shelter Outcomes for Dogs

Analysis of Long Beach Shelter Data from 2017-2024

Executive Summary

Problem:

- Improve outcomes for dogs in shelters by building a prediction model able to be implemented on intake

Findings:

- Gradient boost classification model achieved F1 scores >=64 and AUC >=80 for each outcome type
- Sex and intake condition are most influential features in prediction

Problem Statement

- Shelters handle huge numbers of intakes each year
- Over 1 million dogs have entered shelters in the US in the first half of 2024
- 12% of the 1.5 million dogs that have left shelters in the first half of 2024 had non-live outcomes
- Goals:
  - Identify features that predict outcome type
  - Create model to predict outcome type

Related Work

- Behavior, age, and appearance are common features associated with post-shelter outcomes
- Intake condition predicted outcomes for senior dogs
- Logistic regression and gradient boosting were common methods for analysis

Preprocessing

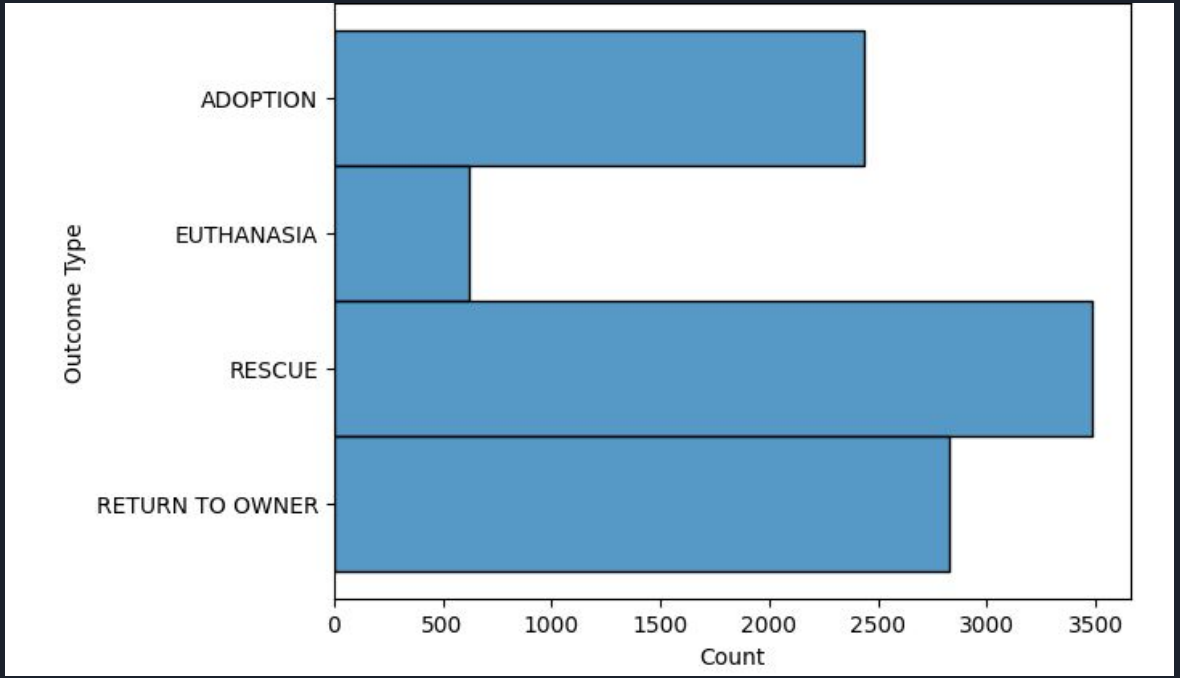
- Long Beach Animal Care Services dataset: 23 features and 9679 observations
- Preprocessing:
  - Remove unnecessary features
  - Identify and impute missing values and errors
  - Add features (age on intake, time in shelter, age category)
  - Visualize feature relationships
  - Summary statistics

Preprocessing

	Feature	Values
0	Primary Color	[BROWN, BLACK, RED, WHITE, SILVER, GOLD, TRICOLOR]
1	Secondary Color	[WHITE, TRICOLOR, BROWN, BLACK, SILVER, GOLD, RED, None]
2	Sex	[Spayed, Male, Female, Neutered]
3	Intake Subtype	[OTC, FIELD, EVICTION, PUB SAFETY, POLICE, CRUELTY, BITE, BORN@SHELT, ABANDON, HOSPITAL, OWNER DIED, RESCUE, EMERGENCY]
4	Intake Condition	[NORMAL, ILL/INJURED MILD, BEHAVIOR MODERATE, ILL/INJURED SEVERE, ILL/INJURED MOD, I/I REPORT, BEHAVIOR MODERATE, AGED, BEHAVIOR SEVERE, UNDER AGE/WEIGHT, BEHAVIOR MILD, WELFARE SEIZURES]
5	Intake Type	[OWNER SURRENDER, STRAY, RETURN, CONFISCATE, WELFARE SEIZED, QUARANTINE, SAFE KEEP]
6	Outcome Type	[ADOPTION, EUTHANASIA, RESCUE, RETURN TO OWNER]

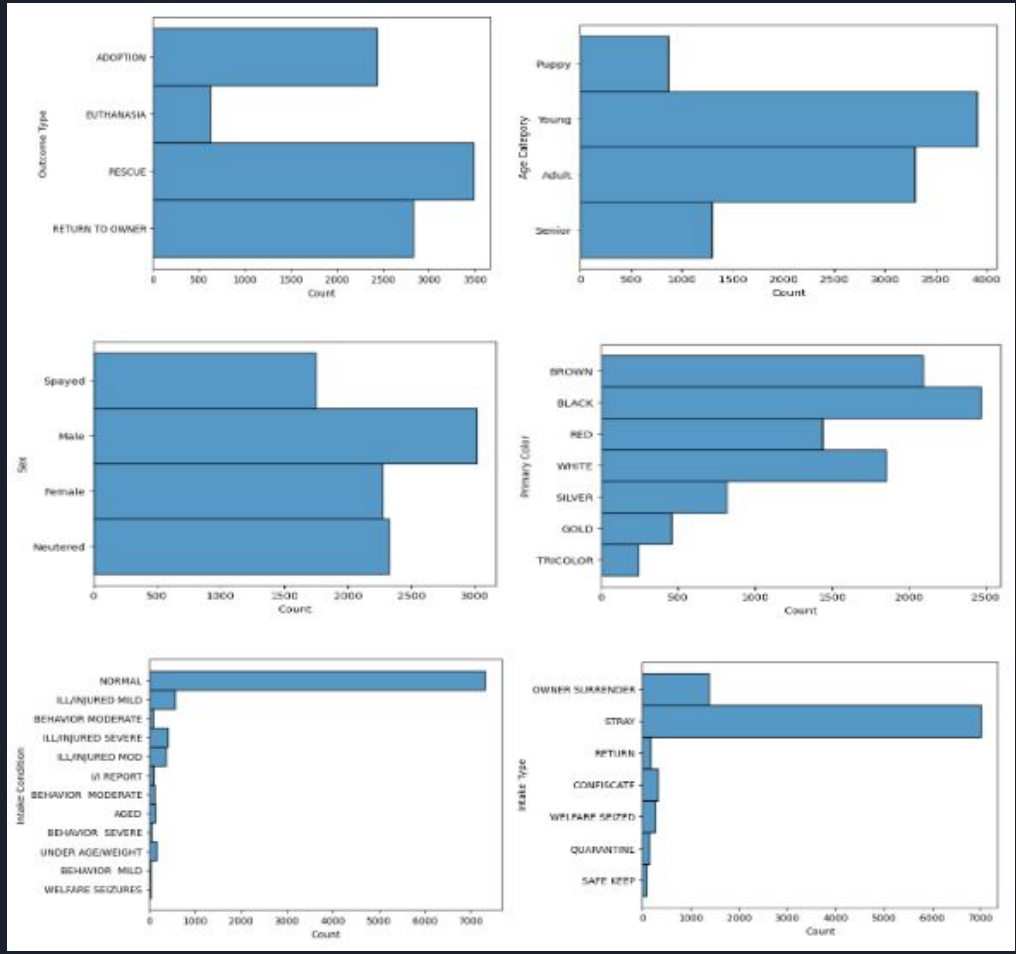
EDA

- Imbalanced dataset
- Addressed with upsampling the minority class (euthanasia)



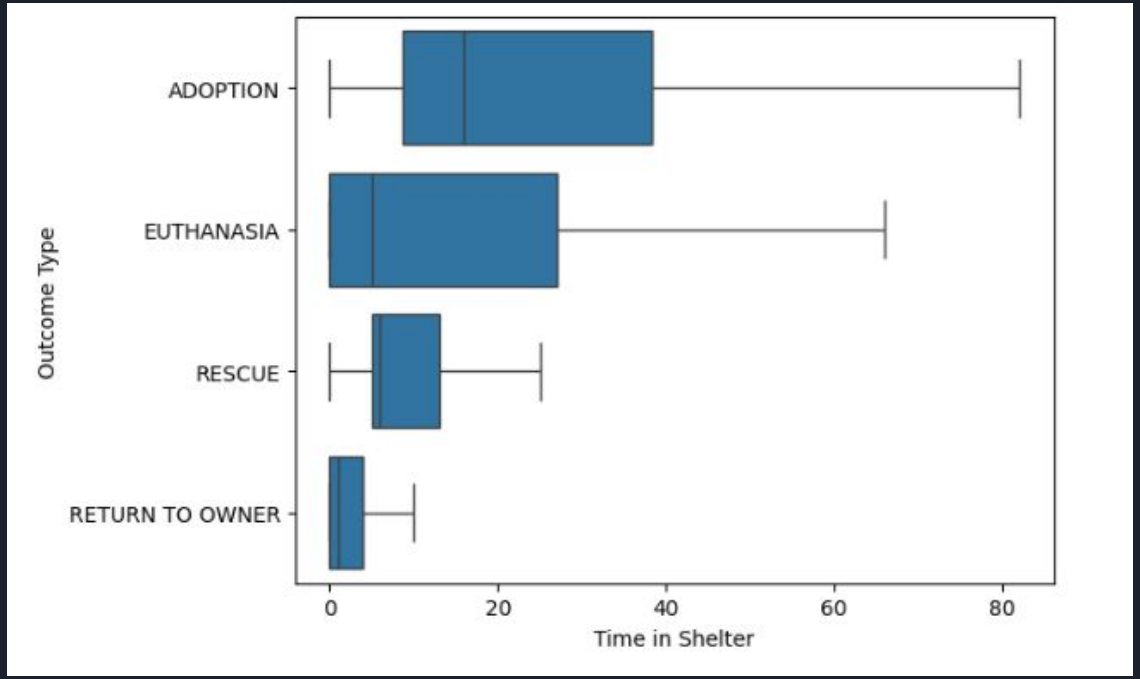
EDA

- Young and adult dogs made up 77% of the dataset
- 78% of dogs had a normal intake condition
- 75% of dogs came in as strays



EDA

- Return to owner has the shortest shelter stay
- Majority of dogs spent < 100 days in shelter



EDA

Two-way relative frequency table

Age Category	Puppy	Young	Adult	Senior	All
Outcome Type					
ADOPTION	0.24	0.33	0.24	0.11	0.26
EUTHANASIA	0.03	0.05	0.08	0.11	0.07
RESCUE	0.65	0.36	0.35	0.29	0.37
RETURN TO OWNER	0.07	0.27	0.33	0.49	0.30
All	1.00	1.00	1.00	1.00	1.00

Two-way relative frequency table

Sex	Female	Male	Neutered	Spayed	All
Outcome Type					
ADOPTION	0.10	0.08	0.47	0.50	0.26
EUTHANASIA	0.09	0.10	0.03	0.02	0.07
RESCUE	0.55	0.50	0.19	0.18	0.37
RETURN TO OWNER	0.27	0.32	0.31	0.31	0.30
All	1.00	1.00	1.00	1.00	1.00

Two-way relative frequency table

Intake Condition	AGED	BEHAVIOR MILD	BEHAVIOR MODERATE	BEHAVIOR SEVERE
Outcome Type				
ADOPTION	0.14	0.49	0.28	0.07
EUTHANASIA	0.10	0.15	0.17	0.52
RESCUE	0.30	0.24	0.32	0.23
RETURN TO OWNER	0.47	0.12	0.23	0.18
All	1.00	1.00	1.00	1.00

Intake Condition	BEHAVIOR MODERATE	I/I REPORT	ILL/INJURED MILD
Outcome Type			
ADOPTION	0.35	0.48	0.30
EUTHANASIA	0.15	0.10	0.07
RESCUE	0.27	0.26	0.37
RETURN TO OWNER	0.23	0.16	0.26
All	1.00	1.00	1.00

Intake Condition	ILL/INJURED MOD	ILL/INJURED SEVERE	NORMAL
Outcome Type			
ADOPTION	0.25	0.13	0.27
EUTHANASIA	0.17	0.40	0.03
RESCUE	0.38	0.31	0.37
RETURN TO OWNER	0.20	0.16	0.32
All	1.00	1.00	1.00

Intake Condition	UNDER AGE/WEIGHT	WELFARE SEIZURES	All
Outcome Type			
ADOPTION	0.11	0.17	0.26
EUTHANASIA	0.06	0.00	0.07
RESCUE	0.81	0.14	0.37
RETURN TO OWNER	0.02	0.69	0.30
All	1.00	1.00	1.00

Notable Results from Relative Frequency Tables

- Most frequent outcome for spayed and neutered dogs is adoption
- Most frequent outcome for intact dogs is rescue
- Most frequent outcome for dogs with severe behaviors or with severe illness or injury is euthanasia
- Seniors are less likely than average to be adopted

EDA: Return to Owner

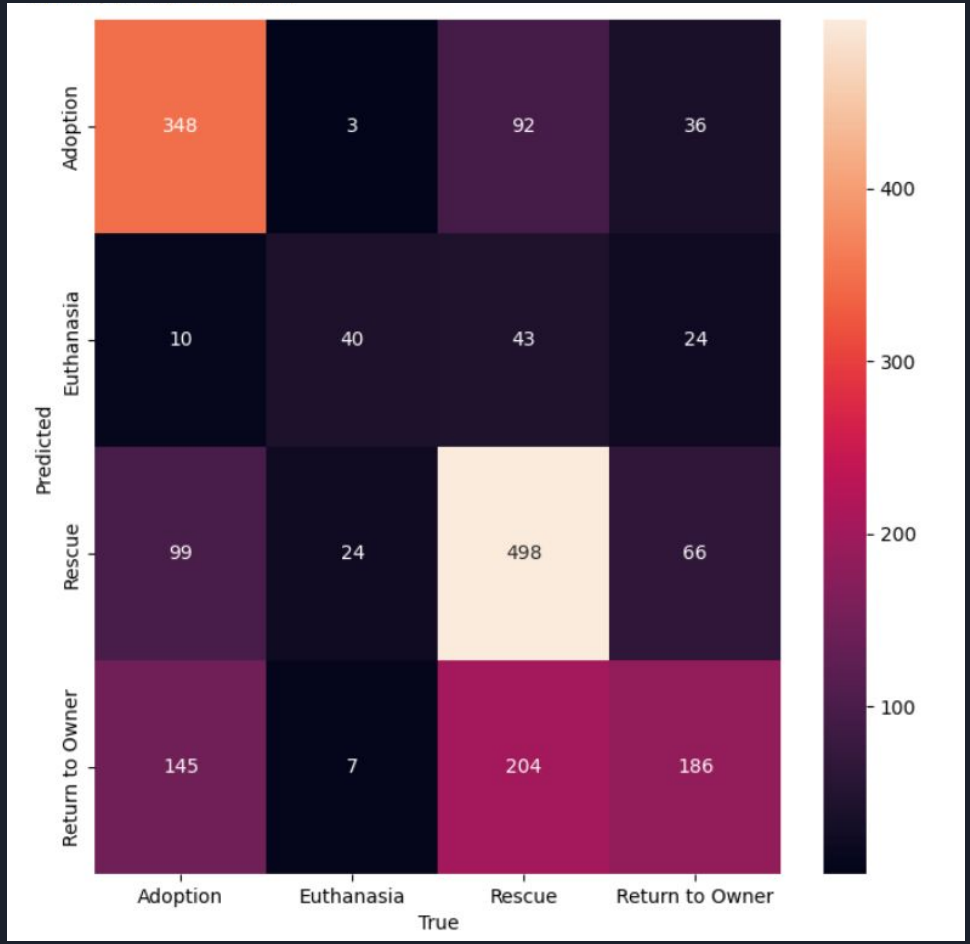
- Shelters have a stray hold period to allow owners to claim lost dogs (typically 3-10 days)
  - Dogs on stray hold cannot be adopted, euthanized or transferred to rescues
- 79% of dogs with a return to owner outcome entered the shelter as strays
- Median time in shelter for dogs with a return to owner outcome was 1 day

Proposed Work

- Questions:
  - Which features predict positive or negative post-shelter outcomes?
  - Can features available at intake predict outcomes?
- Proposed methods:
  - Gradient boosting
  - Random forest
  - AdaBoost

Alterations to Plan

- Dropped return to owner outcome
  - Poor performance
  - Limited utility in prediction due to short shelter stay

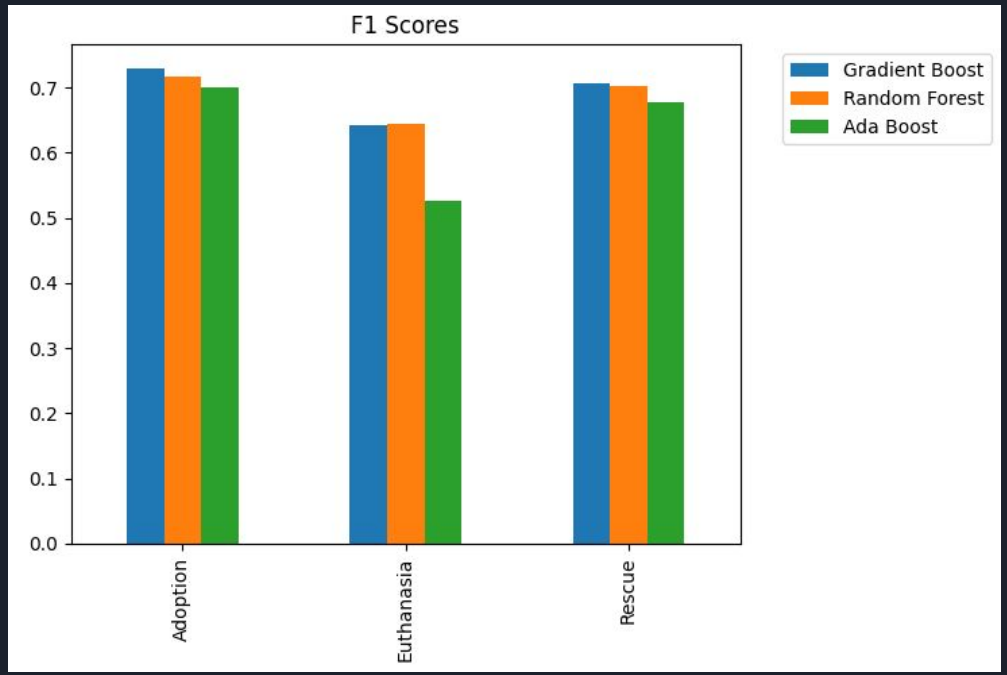


Evaluation

- Primary metric: F1-score and AUC
  - Balance precision and recall
- Method:
  - Cross-validation within models to optimize parameters
  - Compare F1-scores and AUC between models

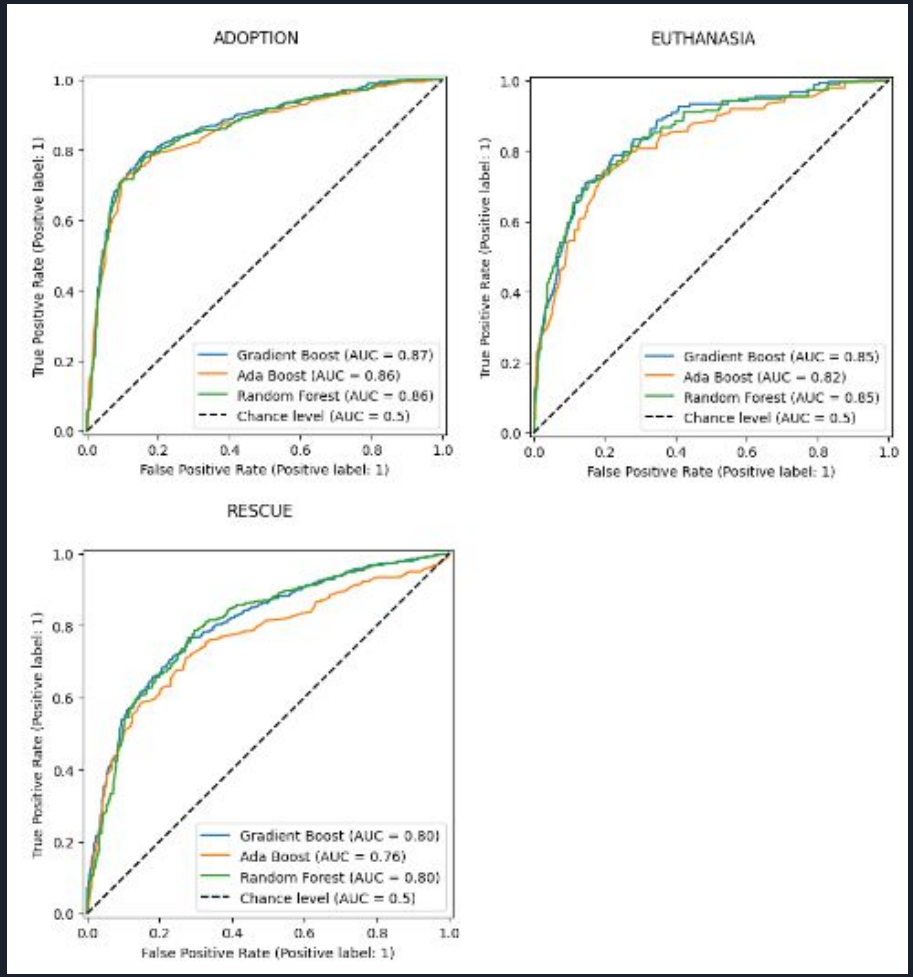
Evaluation: F1-score

- Gradient boost and random forest had similar performance
- All models performed worse predicting euthanasia outcomes



Evaluation: AUC

- Similar performance for adoption
- AdaBoost performed slightly worse for euthanasia and rescue



Feature Importance

Feature	Mean Permutation Importance	STD
Sex_Neutered	0.13	0.003724
Sex_Spayed	0.11	0.005842
Intake Condition_NORMAL	0.07	0.007705
Age on Intake	0.04	0.007729
Intake Type_STRAY	0.03	0.005278

Discussion

- Model successfully predicts outcomes based on information available at time of intake to the shelter
- Sex and intake condition are most influential features
- Enables data-driven decisions about dogs in the shelter

Future Directions

- Generalization to other shelters
- Feature engineering
  - Does sex matter or does spay/neuter status matter?
  - Can intake condition be simplified to normal vs abnormal?

Timeline

- Week 1: Project proposal and data preprocessing
- Week 2: Model building and evaluation
- Week 3: Interpretation and final report