Dog Adoption Modeling

Predictive modeling for dog adoption at the Austin Animal Center

Dieter, Joe, Daniel, & Lauren

Problem background

The Austin Animal Center runs the largest municipal animal shelter in the US, providing shelter to more than 18,000 animals every year.

Challenge

Animal shelters have limited resources and need to make decisions around which dogs to market heavily to their communities. They can also benefit from more accurate predictions of their future shelter capacity (i.e. how many animals they must care for at a given time).

Can we predict how long a dog will be at the AAC after intake?

Ellie

3 months old
Beagle
Female, not spayed

George

14 years old
Labrador
Male, neutered
Hip problems

Goose

1.5 years old
Pitbull Terrier
Male, not neutered

Marshmallow

6 years old
Mixed breed
Female, spayed
Tri-pawed

How long will each of these dogs will be at AAC after intake?

Framing the predictive analytics problem:

— — —

Data:

Historical intake & outcome data from Austin Animal Shelter from 2013-Present
(filtered to dogs only, and only those that have had an outcome).

Information:

Predicted length of time spent at AAC from intake to outcome.

Decision:

IF predicted length of time is greater than [threshold] THEN prioritize for increased resource use.

Advantage:

Better distribution of resources, decreased costs, informed planning, helping dogs find forever homes.

Data Exploration

Sample of data

Days at AAC	Purebreed	Blacklisted Breed	Color	Intake Date	Intake Weekday	Intake Month	Intake Time of Day	Found in Austin	Intake Type	Intake Condition	Sex	Reproductive Condition	Intake Age Years	Date Birth	Outcome Type
13	No	No	Multi	5/30/2017	Tuesday	May	Afternoon	Yes	Stray	Normal	Female	Fixed	1	5/30/2016	Adoption

1. Cleaned Data
    a. Removed observations where stay duration was misreported
    b. Removed observations where age was unknown
2. Early Insights
    a. Significant number of dogs stay at adoption center for less than 2 months
    b. Intake age, Blacklisted breeds, Intake condition

Analytics

Numeric

Linear Regression

Estimated length of stay at shelter

Best for interval values

Nearest Neighbor

Estimated length of stay at shelter

Best for numeric or classifier - we used it for numeric

Classifier

Classification Tree

Predict if still present at shelter after 2 weeks, 1 month, and 3 months

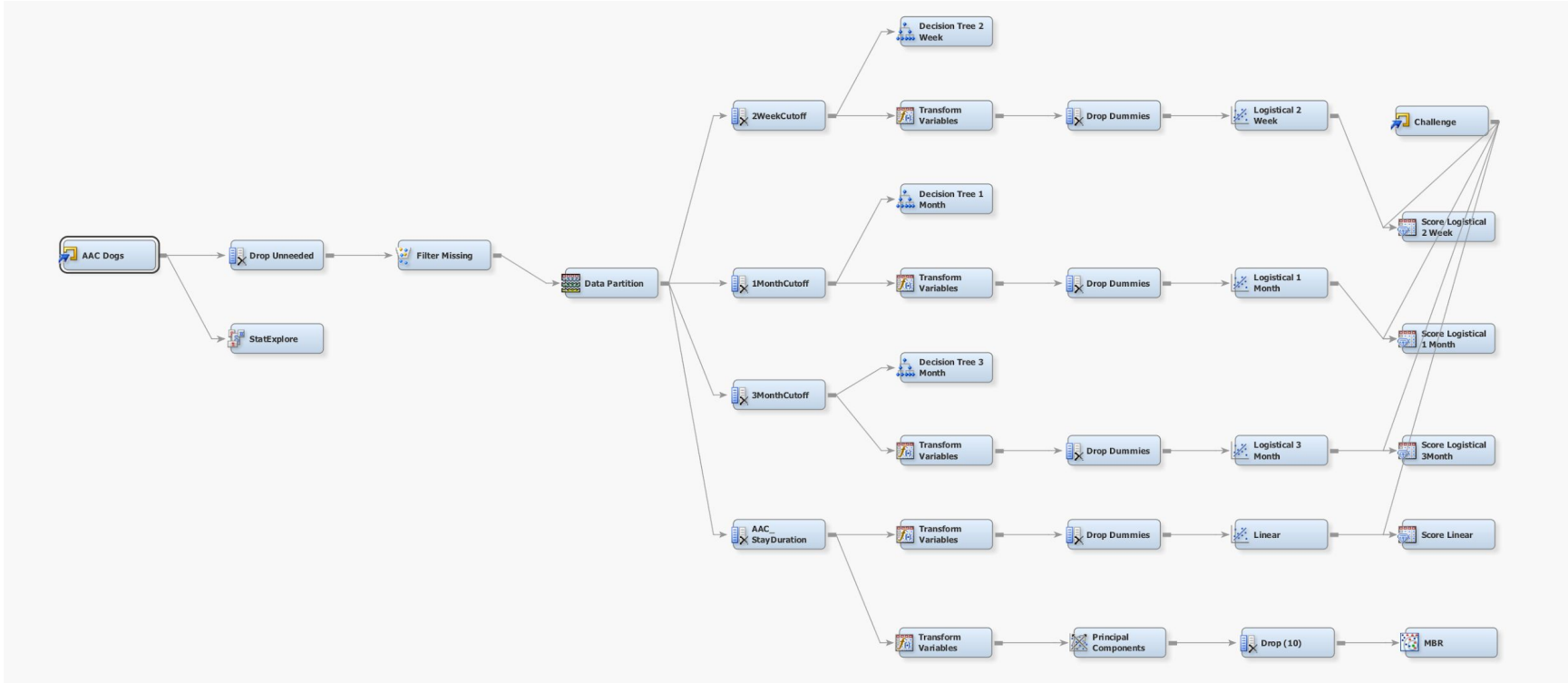
Best for understanding most influential inputs

Logistic Regression

Predict if still present at shelter after 2 weeks, 1 month, and 3 months

Best for understanding input contributions

Implementation



Implementation

Numeric

Linear Regression

Calculate stay duration of historical data

Standardize values

Drop dummy variables

Run regression

Nearest Neighbor

8 or more nearest neighbors gave roughly the same RMSE

Classifier

Classification Tree

Specified cut off

Max depth levels varied:
2 weeks - 3
1 month - 4
3 months - 6

Logistic Regression

Specified cut off

Standardized values

Drop dummy variables

Run regression

Results

Linear Regression

Intercept	0.0946
Intake Type - Public Assist	-0.2004
Intake Month - June	-0.0907
Found in Austin	-0.0611
Intake Type - Stray	-0.0607
Reproductive Condition - Fixed	-0.0456
Blacklisted Breed	-0.0373
Intake Age - Years	0.0374
Intake Weekday - Monday	0.0537
Intake Condition - Sick	0.1111

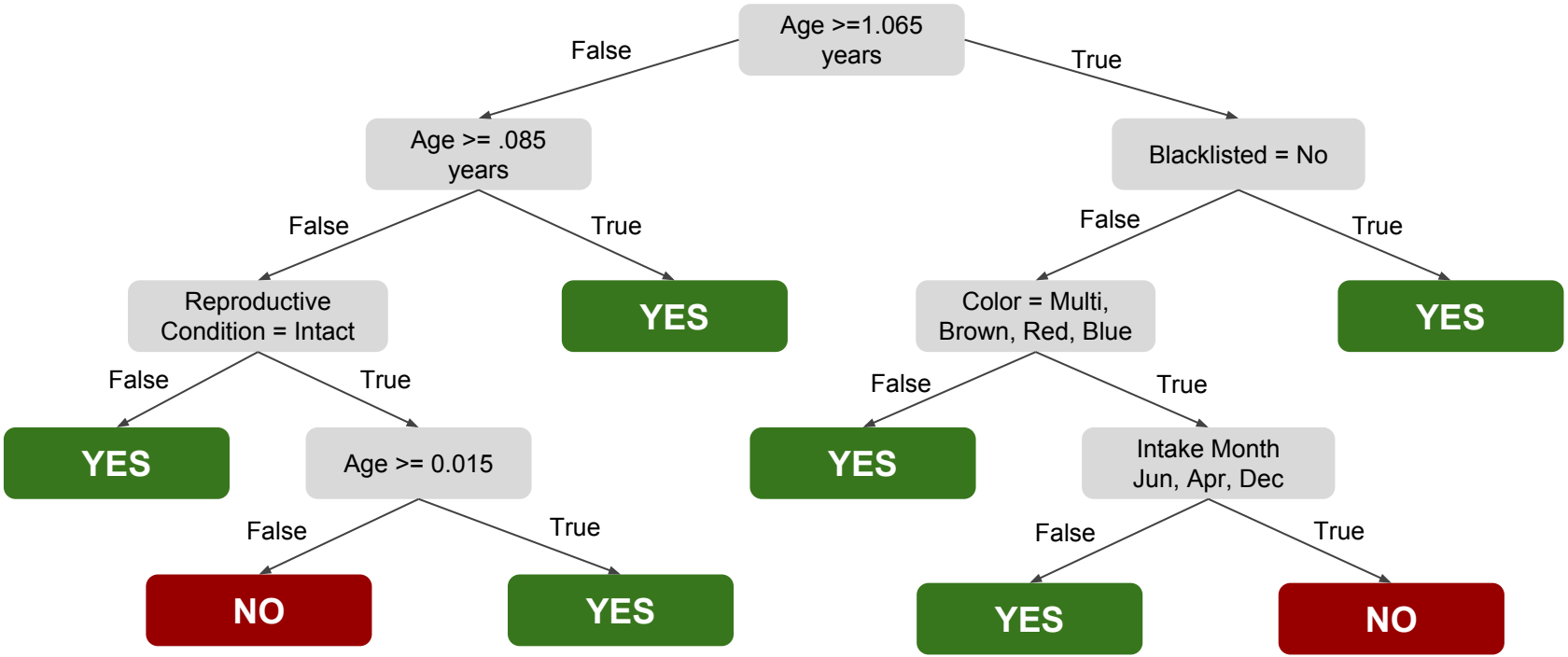
RMSE = .97

Logistic Regression

Misclassification Rate	Naive Predictions
2 week - 30%	2 week - 35%
1 month - 22%	1 month - 22%
3 month - 9%	3 month - 9%

Decision Tree:

Adopted within 1 month?



Misclassification Rate for 1 month - 21%

Insights

Numeric

Linear Regression	Nearest Neighbor
RMSE = 0.97	RMSE = 1.02
Shows contributions of inputs to total	Shows groupings of similar inputs
Granular	Coarse groups
Quick computation	Slow computation
Easy to Understand	Difficult to Understand

Classifier

Classification Tree	Logistic Regression
Misclassification Errors: 2 week - 30% 1 month - 21% 3 month - 9%	Misclassification Errors: 2 week - 30% 1 month - 22% 3 month - 9%
Handles missing values effectively	Cannot have missing values
Coarse groups	Granular
Shows groupings of similar inputs	Shows contributions of inputs to total

Ellie

3 months old
Beagle
Female, not spayed

George

14 years old
Labrador
Male, neutered
Hip problems

Goose

1.5 years old
Pitbull Terrier
Male, not neutered

Marshmallow

6 years old
Mixed breed
Female, spayed
Tri-pawed

How long will each of these dogs will be at AAC after intake?

Ellie

3 months old
Beagle
Female, not spayed

George

14 years old
Labrador
Male, neutered
Hip problems

Goose

1.5 years old
Pitbull Terrier
Male, not neutered

Marshmallow

6 years old
Mixed breed
Female, spayed
Tri-pawed

Animal_ID	Animal_type	Purebreed	Blacklisted_breed	Color	Intake_date	Intake_weekday	Intake_month	Intake_time_of_day	Found_in_Austin	Intake_type	Intake_condition	Sex	Reproductive Condition	Intake_age_years	Date_birth
1	Dog	Yes	No	Multi	4/17/18	Tuesday	April	Morning	Yes	Stray	Normal	Female	Intact	0.25	1/23/18
2	Dog	Yes	No	Black	4/17/18	Tuesday	April	Morning	Yes	Stray	Injured	Male	Fixed	14	4/23/04
3	Dog	No	Yes	Multi	4/17/18	Tuesday	April	Morning	Yes	Stray	Normal	Male	Intact	1.5	10/23/17
4	Dog	No	No	Multi	4/17/18	Tuesday	April	Morning	Yes	Stray	Injured	Female	Fixed	6	4/23/12

Ellie

3 months old
Beagle
Female, not spayed

Predicted: 35 days

George

14 years old
Labrador
Male, neutered
Hip problems

Predicted: 64 days

Goose

1.5 years old
Pitbull Terrier
Male, not neutered

Predicted: 45 days

Marshmallow

6 years old
Mixed breed
Female, spayed
Tri-pawed

Predicted: 53 days

Project challenges

- Raw AAC data was messy
- Pre-processing in Excel and SAS
- Creating calculated fields and modified fields out of string inputs
- Determining HOA ‘banned dogs’
- Designating ex-ante fields and point of use cases

Closing thoughts

- Many different types of organizations can benefit from predictive models
- Predictive models can save lives!
- How might we make our model useful for ‘on-the-ground’ (non-technical) shelter workers?