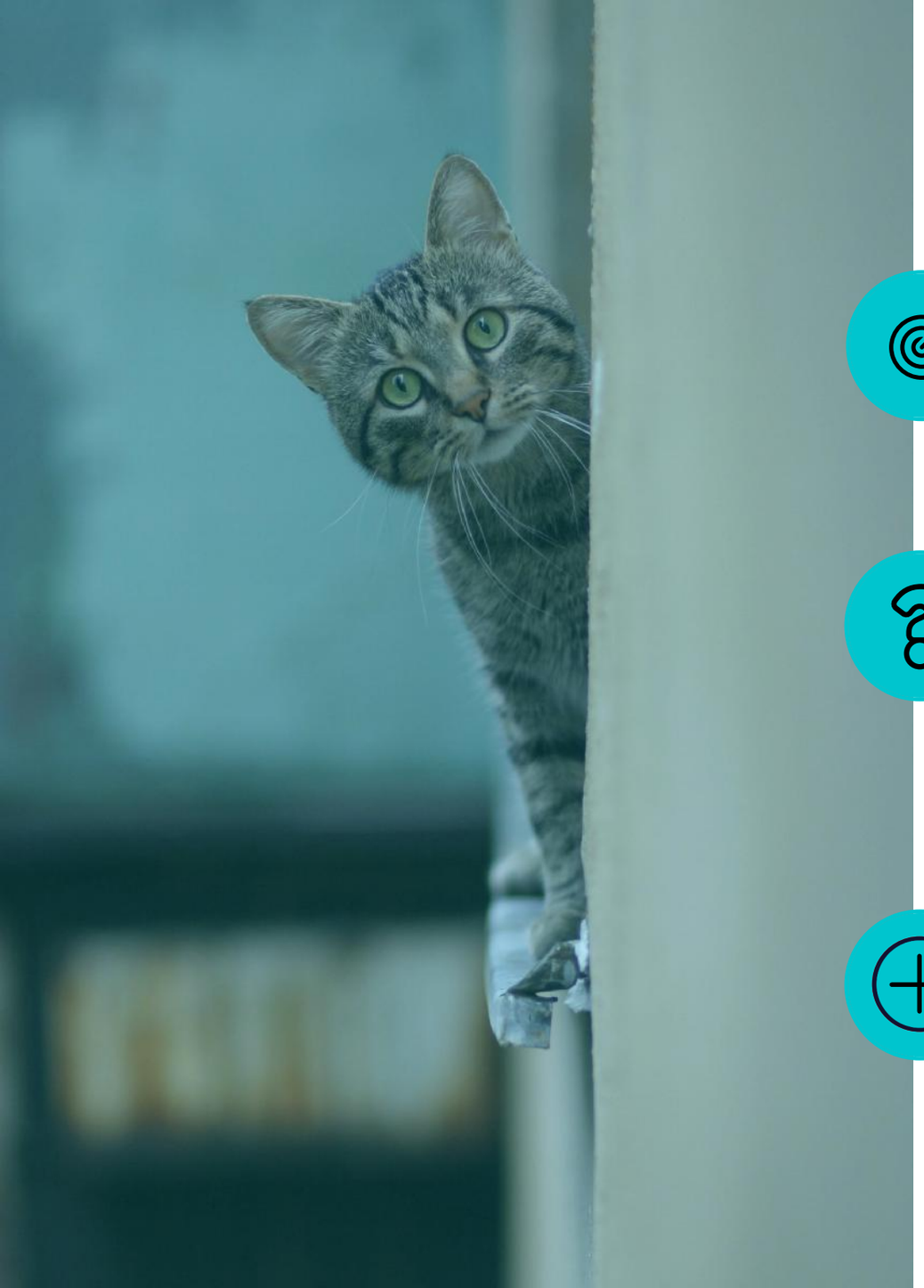
Predictive model of early pet adoption

Date:
06/10/2023

Presented by:
Ali Hijazy



BUSINESS PROBLEM

Objective

Predict the likelihood of quick pet adoptions

Why it matters

- Optimize pet profiles for faster adoption.
- Anticipate duration of pets' stays.
- Efficiently plan resources (food, space, care).
- Reduce animal suffering and euthanization rates.

Operational Benefits

- Improved profiling: Insights on what makes a profile attractive.
- Efficient Resource Allocation: Predictive info helps shelters prepare in advance.
- Foster System Support: Knowing which pets might take longer to adopt can encourage fostering.

DATASET

petfinder

- Type
- Age
- Breed
- Sex
- Color
- Size
- Fur
- Health
- Sterilized
- Vaccinated
- Dewormed
- Fee
- Photos
- Videos

Target variable:
Adoption Speed

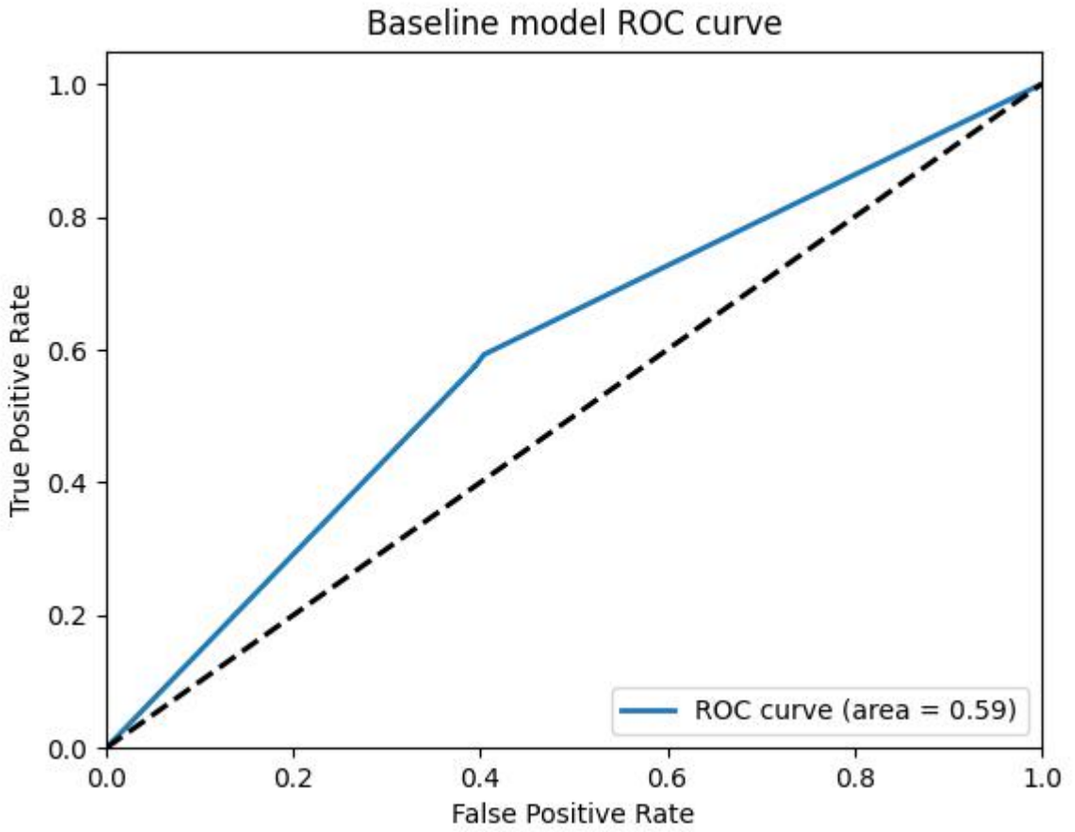
- 1: Adopted within a month: 7537
- 2: Not adopted within a month: 7456

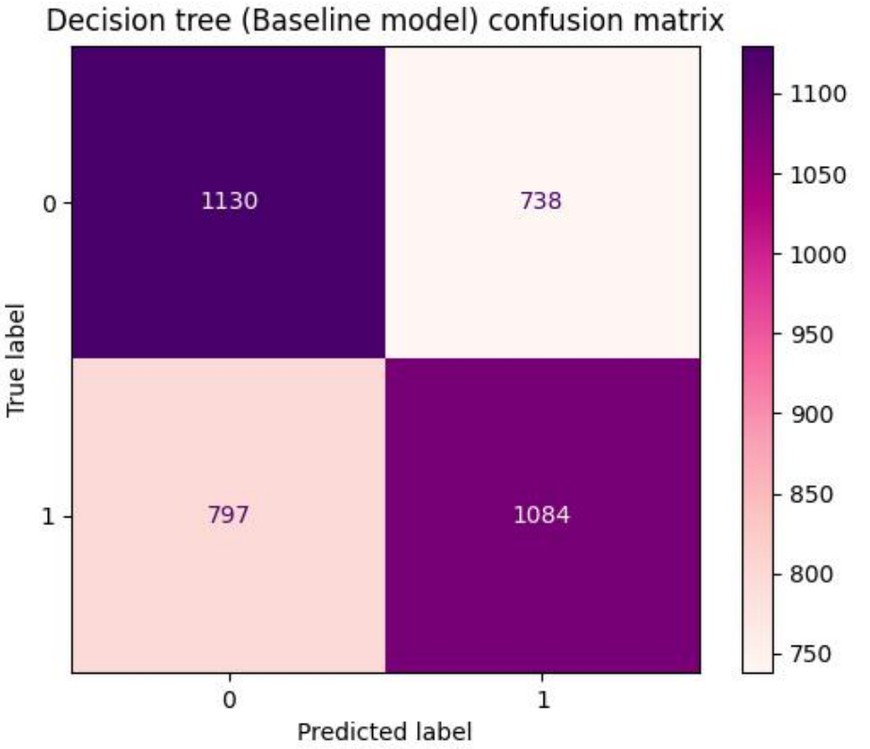
MODELS

Decision tree (Baseline model)

ACCURACY ON TRAIN DATA: 0.99

ACCURACY ON TEST DATA: 0.59



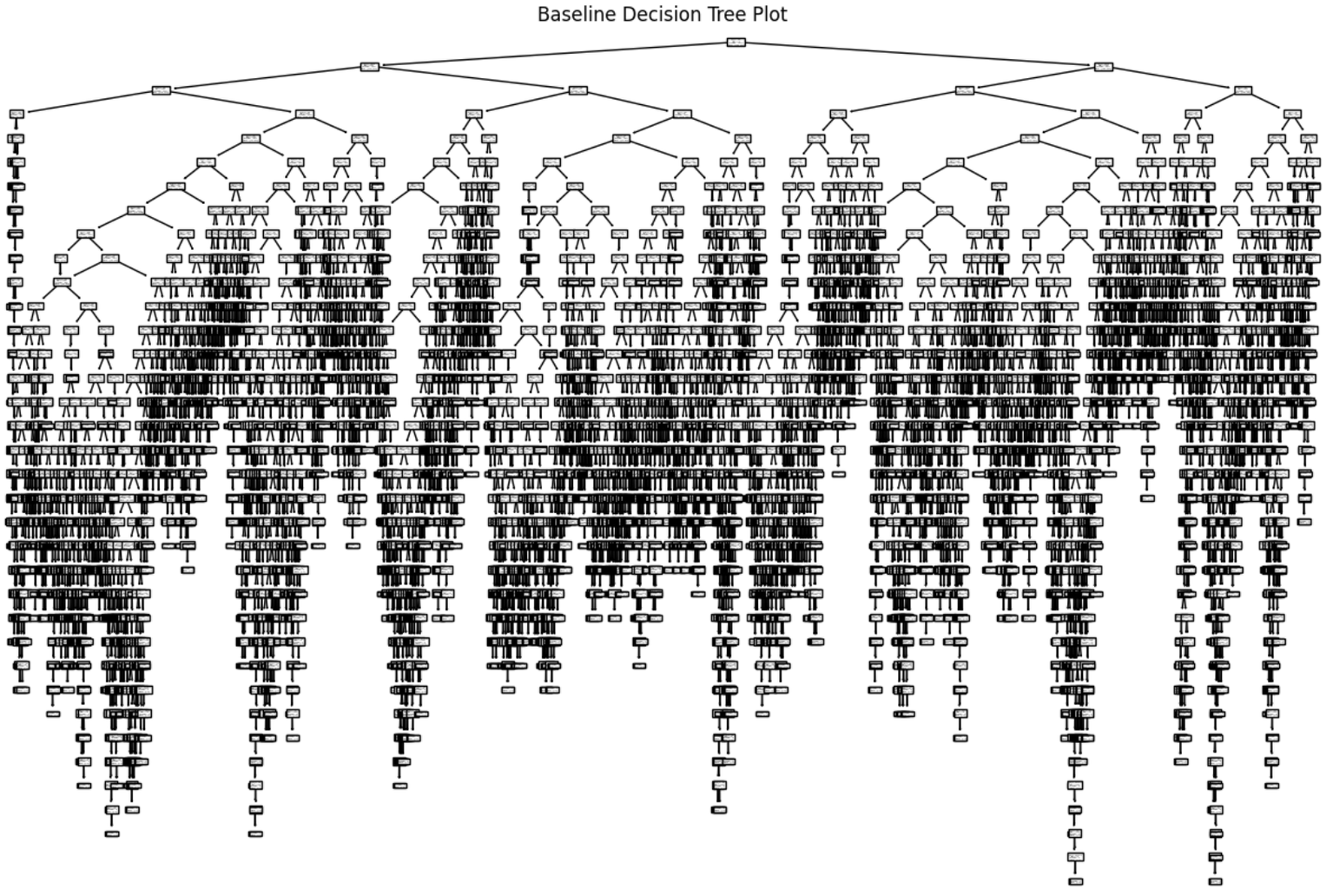


MODELS

Decision tree (Baseline model)

ACCURACY ON TRAIN DATA: 0.99

ACCURACY ON TEST DATA: 0.59

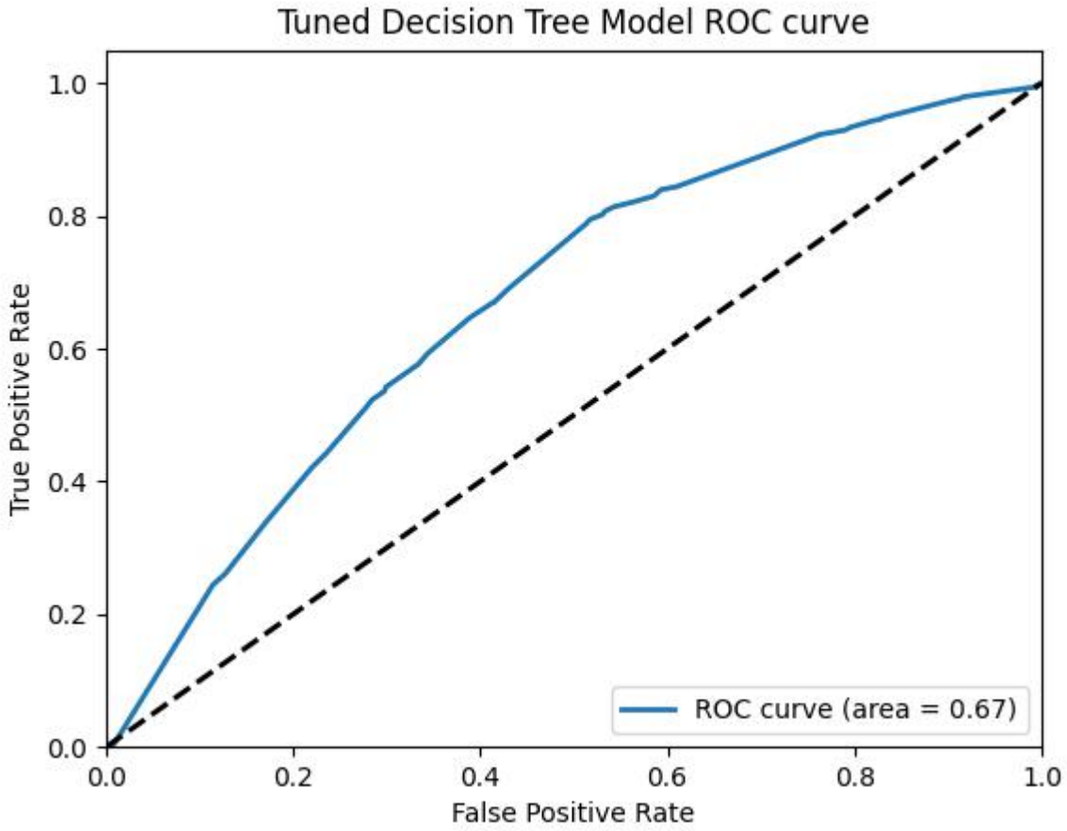


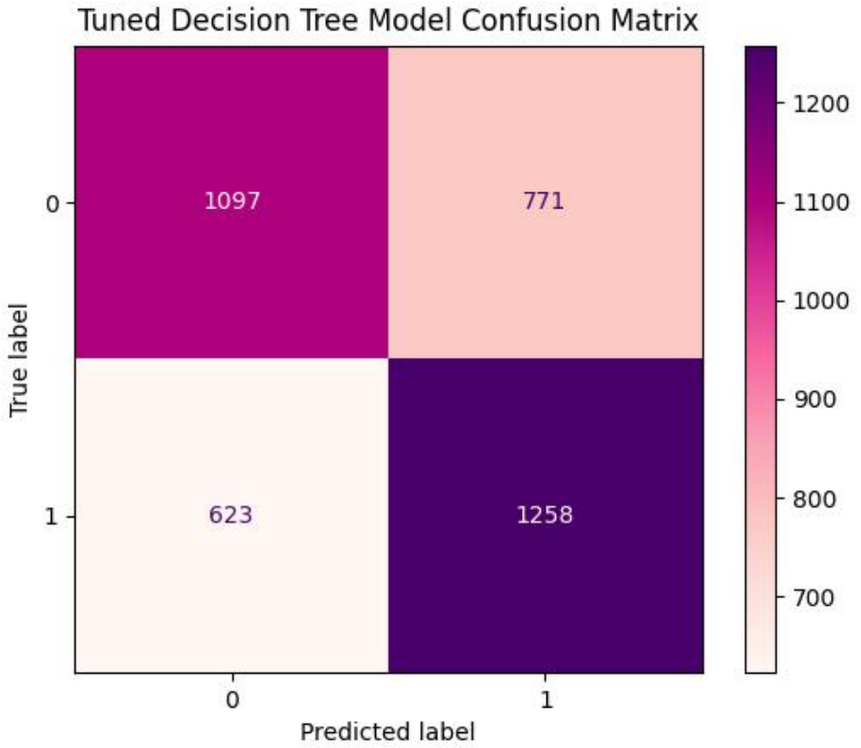
MODELS

Tuned decision tree

ACCURACY ON TRAIN DATA: 0.65
ACCURACY ON TEST DATA: 0.63

USING GRIDSEARCHCV
BEST HYPERPARAMETERS:
CRITERION: GINI
MAX_DEPTH: 6
MIN_SAMPLES_LEAF: 4
MIN_SAMPLES_SPLIT: 2
SPLITTER: BEST



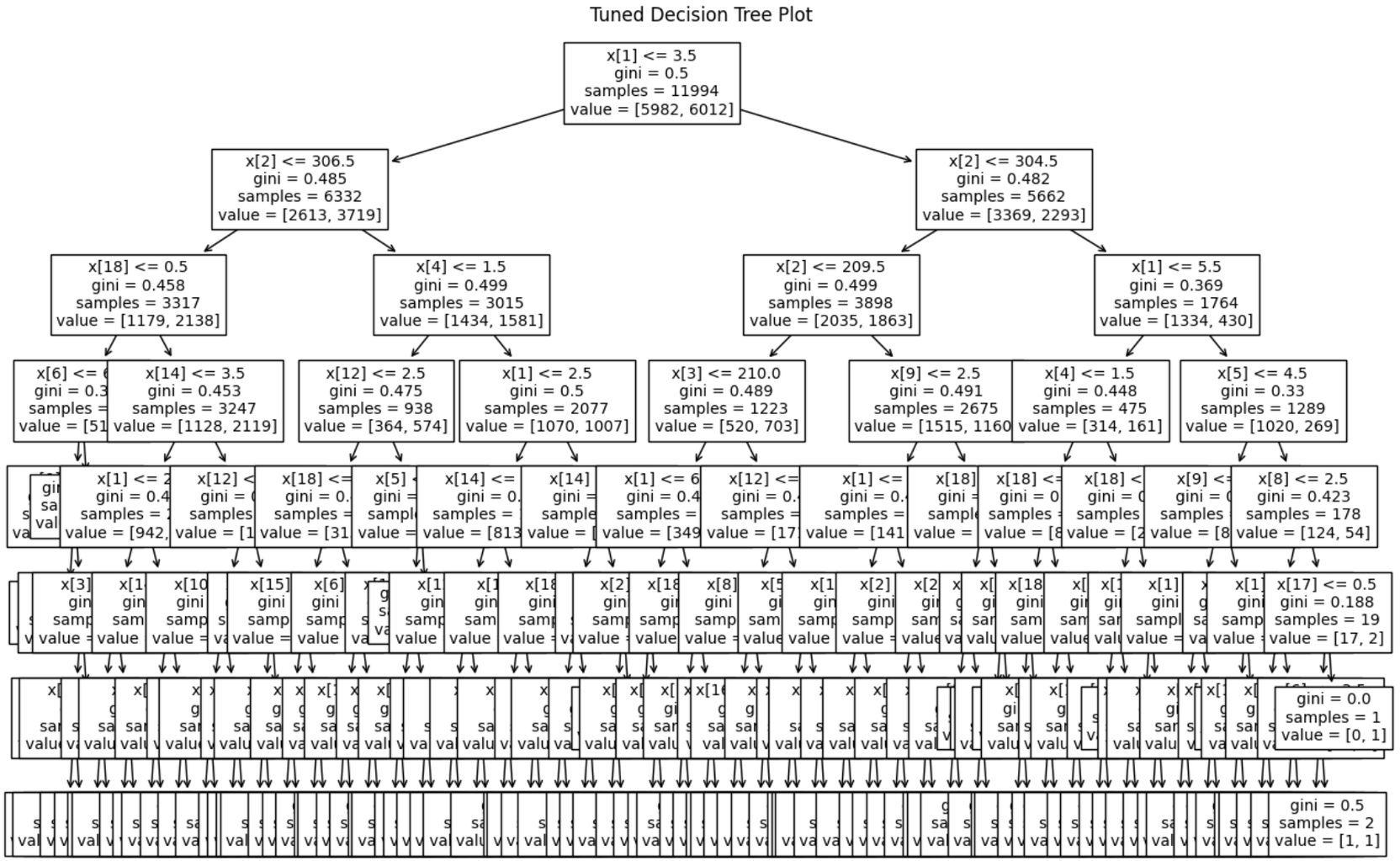


MODELS

Tuned decision tree

ACCURACY ON TRAIN DATA: 0.65

ACCURACY ON TEST DATA: 0.63

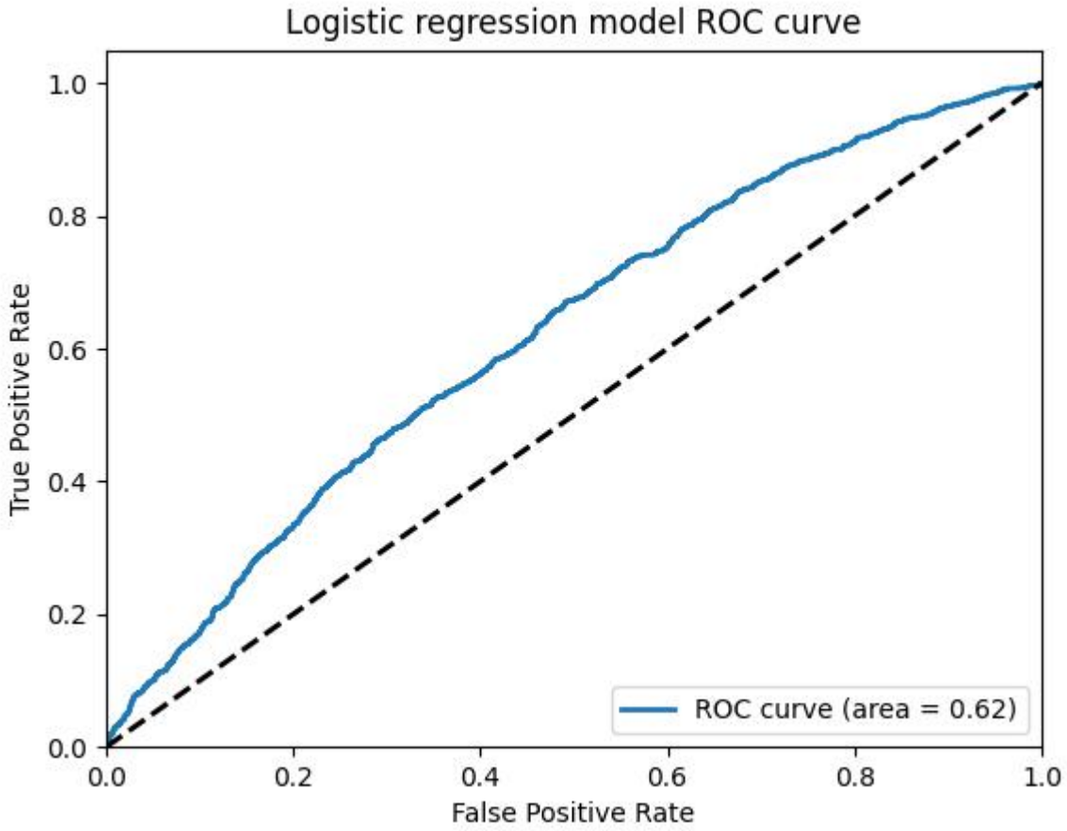


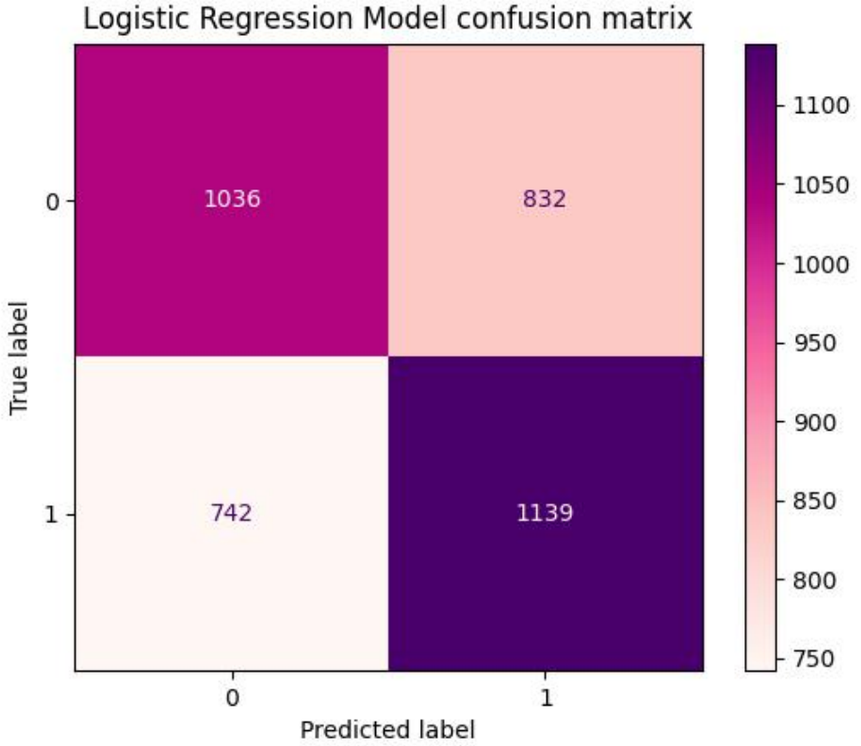
MODELS

Logistic regression

ACCURACY ON TRAIN DATA: 0.59

ACCURACY ON TEST DATA: 0.58





MODELS

Tuned logistic regression

ACCURACY ON TRAIN DATA: 0.59

ACCURACY ON TEST DATA: 0.59

USING GRIDSEARCHCV

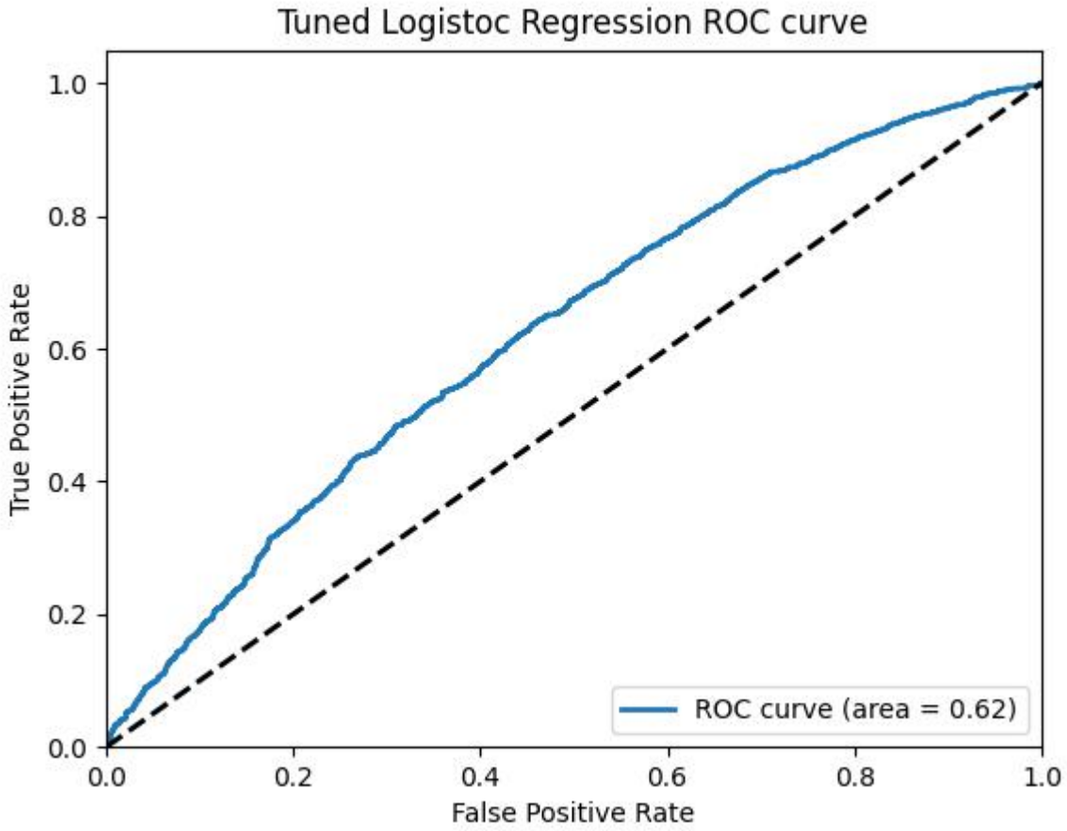
BEST HYPERPARAMETERS:

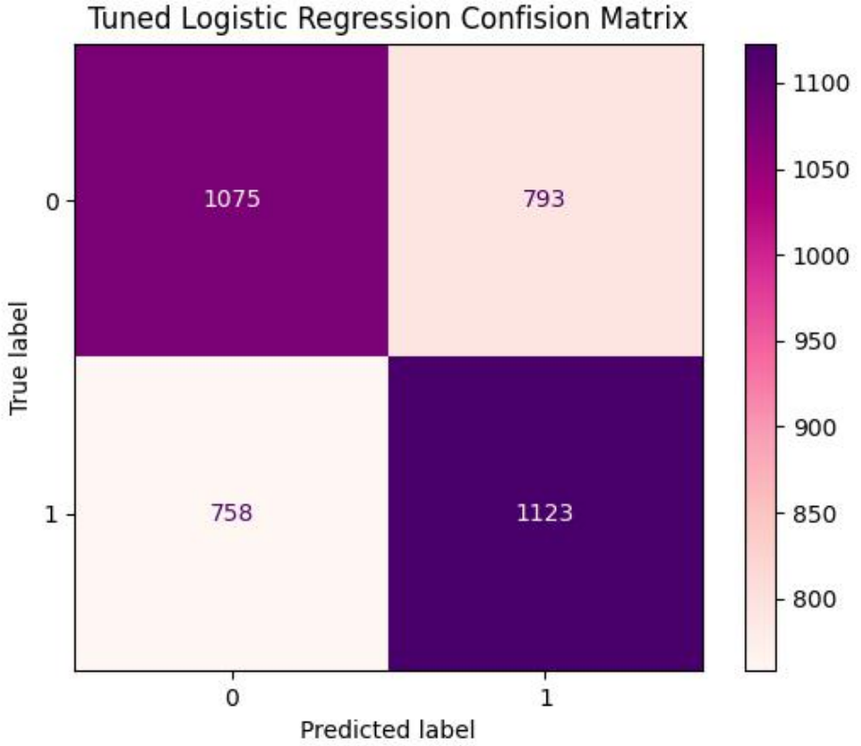
C: 0.01

PENALTY: L2

SOLVER: 'SAGA'

TOL: 1





MODELS

Refining the Decision Tree

	Feature	Importance
1	Age	0.301293
2	Breed1	0.280473
18	PhotoAmt	0.100192
14	Quantity	0.049627
3	Breed2	0.048646
12	Sterilized	0.034529
4	Gender	0.029830
15	Fee	0.027520
9	FurLength	0.027414
5	Color1	0.023370
8	MaturitySize	0.018921
10	Vaccinated	0.012139
11	Dewormed	0.008440
7	Color3	0.004149
0	Type	0.002889
13	Health	0.002820
16	State	0.002615
17	VideoAmt	0.001938

Refitting the tuned model using only these 10 columns

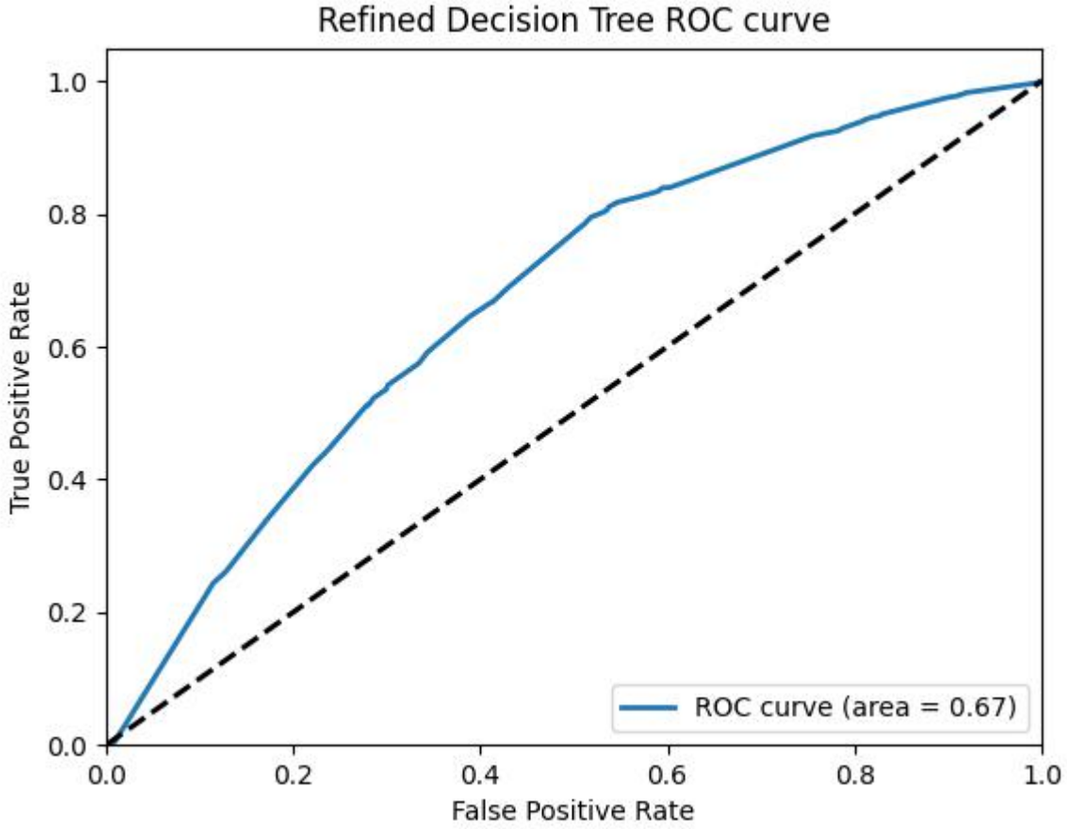
BEST HYPERPARAMETERS:
CRITERION: GINI
MAX_DEPTH: 6
MIN_SAMPLES_LEAF: 4
MIN_SAMPLES_SPLIT: 2
SPLITTER: BEST

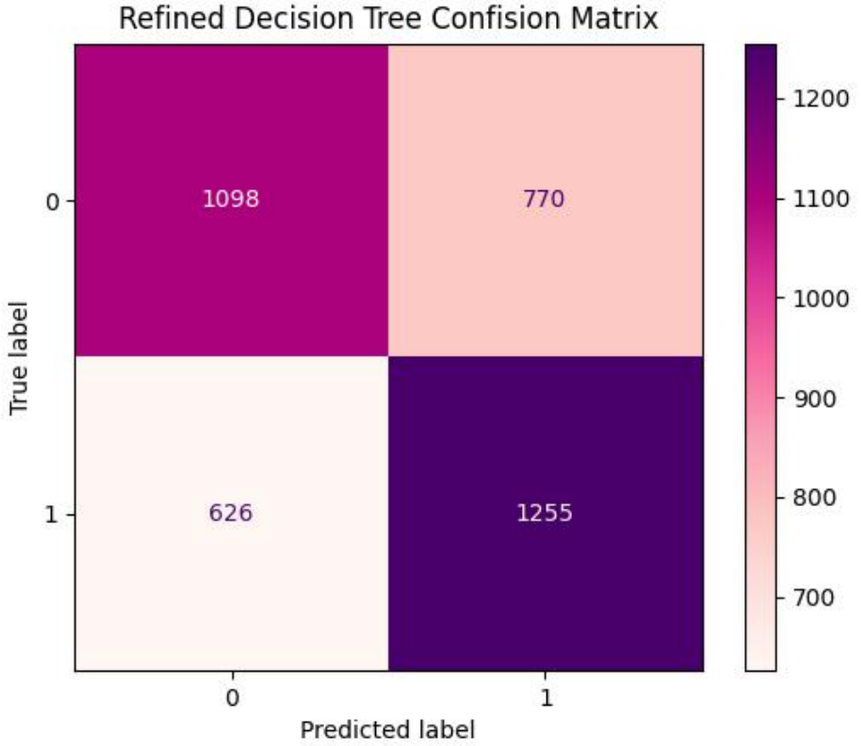
MODELS

Refined Decision Tree

ACCURACY ON TRAIN DATA: 0.65
ACCURACY ON TEST DATA: 0.63

BEST HYPERPARAMETERS:
CRITERION: GINI
MAX_DEPTH: 6
MIN_SAMPLES_LEAF: 4
MIN_SAMPLES_SPLIT: 2
SPLITTER: BEST





Best performing model

Logistic Regression	Tuned Logistic Regression	Decision Tree	Tuned Decision tree	Refined Decision Tree
Test: 0.58 **Train: 0.59**	**Test: 0.59** **Train: 0.59**	**Test: 0.59** **Train: 0.99**	**Test: 0.63** **Train: 0.65**	**Test: 0.63** **Train: 0.65**
Features used: 24	Features used: 24	Features used: 24	Features used: 24	Features used: 10

Github.com/alihijazy/

Linkedin.com/in/alihijazy/

THANK YOU