Predicting the Adoptability of Shelter Animals: A Classification Analysis

Nolan Vibhakar, Audra Cornick, Angela Yandrofski, Rowan Murphy

Table of contents

Group 2 . . . 1
Members: . . . 1

Summary **1**

Introduction **2**

Methods and Results **4**
Exploratory Analysis Visualizations . . . 6
Feature Engineering . . . 12
Classification Analysis . . . 12

Discussion **14**

References **14**

Group 2

Members:

- Audra Cornick
- Angela Yandrofski
- Nolan Vibhakar
- Rowan Murphy

Summary

The largest problems animal shelters are facing in North America is a lack of space and people to care for the large number of animals they receive everyday. Each day across the United States in 2023, 13,000 animals entered shelters with many more still being on the streets as strays ("Here's What's Happening in Animal Shelters Across the u.s." n.d.). Additionally many animals received by shelters have health conditions that need immediate attention or could potentially spread to other animals, which require more resources to ensure swift treatment of these animals and measures to prevent the spread of sickness within these facilities. It is important that shelters are able to make quick decisions about where they dedicate their limited resources and optimize the adoption process to ensure as many animals are able to find home as possible.

Our project uses animal shelter data from Lydia Gibson (2025) to predict whether or not an animal will be adopted based on their age, sex, animal type, intake condition, intake type, and length of stay at shelter. Using predictive analysis in R, we create a k-nearest neighbor model to predict if an animal is going to be adopted based on the features listed above. This analysis aims to help enhance adoption rates, so shelters can recognize which animals are unlikely to get adopted and tailor their outreach and marketing strategies to specifically promote the adoption of those animals. Additionally, better understanding which animals are at a higher risk of not being adopted allows shelters to allocate resources more effectively, by providing additional care and resources to help those that are unlikely to be adopted.

Introduction

Animal shelters provide a valuable service for both homeless animals and prospective pet owners, but more and more they are failing the animals due to being over capacity and having a lack of resources. The Shelter Animals Count Database describes the root of the problem as "more animals are entering shelters than are leaving them" (Burch, n.d.). Due to this, many animals must be gotten rid of before they even have a chance to be adopted. According to "Saving Companion Animals from Inhumane Conditons" (2018) over 4000 animals are euthanised in US shelters per day. This is an animal treatment crisis, as shelters struggle to find animals new homes. While many people are interested in adopting from shelters, they aren't doing it at high enough rates compared to animal abandonment. For the sake of the animals and shelters' continued success in their mission, It is important to figure out how best to address this issue.

A report from "2023 State of Shelter Adoption Report" (n.d.) says that one of the biggest obstacles in the way of pet adoption is guardians' unrealistic expectations. By examining the data on what prospective adopters look for in a pet, we can assess how shelters can address these issues. While certain qualities might be more desirable, shelters can focus their efforts on aiding those animals that might otherwise be overlooked. "Effective Tips to Boost

Adoptions: Reduce Barriers & Improve Experience" (n.d.) suggests streamlining adoption processes, better matchmaking, and enhanced marketing are all valuable ways to increase adoption in shelters. All of these can be aided by analysis of adoption data. By figuring out what characteristics in animals are most relevant to adoption, shelters can focus their marketing efforts and resources into helping the animals that need it most. In our project, we will attempt to predict whether an animal will be adopted or not based on different descriptive factors of the animal, so that shelters are better equipped to find homes for as many animals as possible.

Research Question: Can we predict whether an animal will be adopted based on certain factors?

We used data from the Long Beach Animal Shelter, which contains:

Table 1: Retrieved from Lydia Gibson (2025)

variable	class	description
animal_id	character	Unique identification for each animal.
animal_name	character	Name of the Animal (Blank value means name not known). Animals with "*" are given by shelter staff.
animal_type	factor	Species name of the animal.
primary_color	factor	The predominant color of the animal.
secondary_color	factor	Additional coloring, less predominant than the primary color.
sex	factor	Altered Sex of the animal.
dob	date	Date of Birth (if blank, DOB unknown).
intake_date	date	Date on which Animal was brought to the shelter .
intake_condition	factor	Condition of animal at intake.
intake_type	factor	The reason for intake such as stray capture, wildlife captures, adopted but returned, owner surrendered etc.
intake_subtype	factor	The method or secondary manner in which the animal was admitted to the shelter.
reason_for_intake	factor	The reason an owner surrendered their animal.
outcome_date	date	Exit or Outcome date such as date of adoption or date animal died.
crossing	character	Intersection/Cross street of intake or capture.
jurisdiction	factor	Geographical jurisdiction of where an animal originated.
outcome_type	factor	Outcome associated with animal - adopted, died, euthanized etc.

variable	class	description
outcome_subtype	factor	Secondary manner in which the animal left the shelter, usually used to identify which program, group, or other data useful in measuring program efficiency.
latitude	double	The latitude of the crossing.
longitude	double	The longitude of the crossing.
outcome_is_dead	logical	Whether animal is dead at outcome.
was_outcome_alive	logical	Whether animal was alive at outcome.
geopoint	character	Latitude and longitude of crossing.

Methods and Results

To clean the data, we decided to only select the variables `animal_type`, `primary_color`, `sex`, `dob`, `intake_condition`, `intake_type`, and `outcome_type` as we believe these variables would be the most relevant to helping us predict whether an animal is likely to be adopted or not. We also made sure to omit all the NA values in order to no skew the data analysis.

Additionally, because the `outcome_type` variable had so many different variations of results, we decided to group them into only 2 outcomes (Adopted vs Not Adopted) as that is what we are most interested in. So, we decided to create another variable called `outcome_group` to add our analysis of predicting adoptability.

For our summarization of data, we decided to take a detailed look at the various outcome types (Table 3), and we also decided to show how many animals are adopted vs not adopted (Table 2) to see if there are any biases, and if there is enough data for both outcomes. Additionally, we decided to show a more detailed look at the `intake_condition` variable (Table 4).

```
Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

```
Warning in attr(x, "format"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

Table 2: Breakdown of outcome_group

data_cols	Freq
Adopted	12327
Not Adopted	13684

```
Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

```
Warning in attr(x, "format"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

Table 3: Summary of Outcome Types

data_cols	Freq
adoption	6156
community cat	334
died	585
disposal	45
duplicate	27
euthanasia	4148
foster	10
foster to adopt	164
homefirst	86
missing	54
rescue	5911
return to owner	3107
return to rescue	46
return to wild habitat	146
shelter, neuter, return	896
transfer	4128
transport	133
trap, neuter, release	35

```
Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

```
Warning in attr(x, "format"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

Table 4: Summary of Intake Condition Variable

data_cols	Freq
aged	155
behavior mild	68
behavior moderate	124
behavior severe	68
feral	278
fractious	634
i/i report	224
ill mild	1040
ill moderatete	759
ill severe	991
injured mild	710
injured moderate	780
injured severe	1393
intakeexam	1
normal	12762
under age/weight	5964
welfare seizures	60

Exploratory Analysis Visualizations

Explanation of Relevance: Based on Figure 1 and Figure 2, we can see a high adoptability rate if the animal is livestock, a rabbit, a guinea pig, a reptile, or a cat. This is relevant to our question as it allows us to see which animal types are the most likely to get adopted overall in comparison to some that may be returned to the wild or euthanized. However, it's important to note that when we look at the Animal Type vs Outcome type visualization, a large proportion of animals get returned to owners and to the wild.

Explanation of Relevance: Based on Figure 4 we can see there is a clear relationship between animals that are severely ill and/or injured, feral, have welfare seizures, and have severe behavioral problems when they arrive at the shelter and not having a high adoptability rate later on. This can also be explained by looking at Figure 3, as we can see that there are high rates of ethunaisa for animals that have severe behavior or are severely injured at intake. These visualizations are relevant to our question because it shows that animals that need more care are less likely to get adopted, this could be attributed to the animal sadly passing away

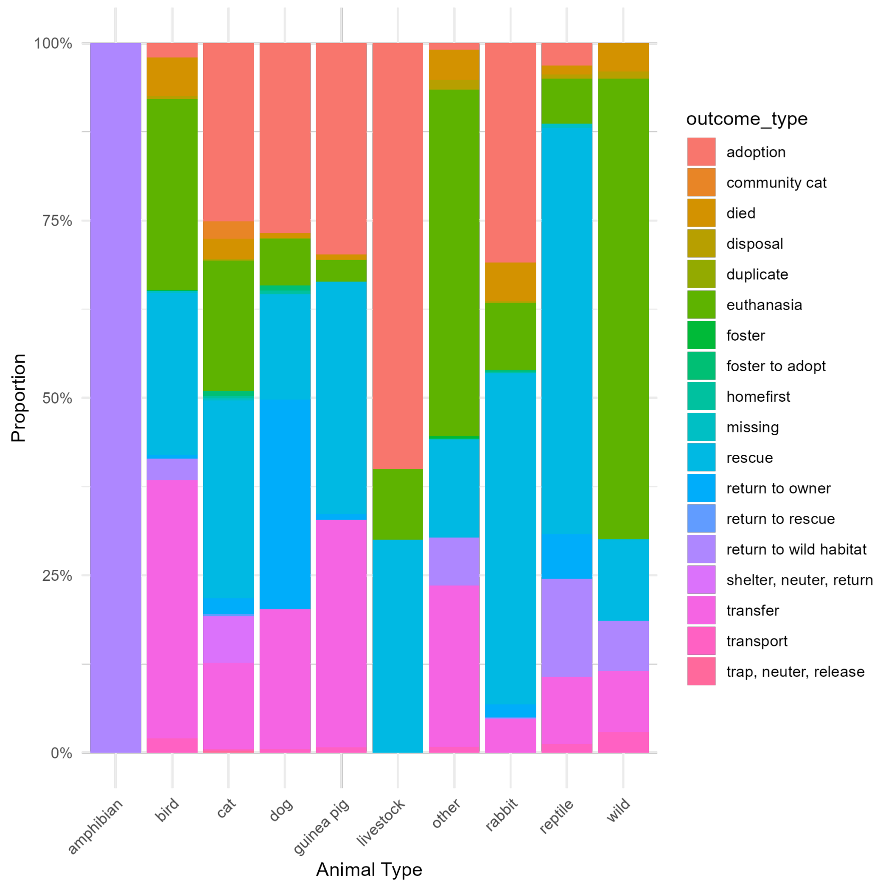

Figure 1: Animal Type vs Outcome Type

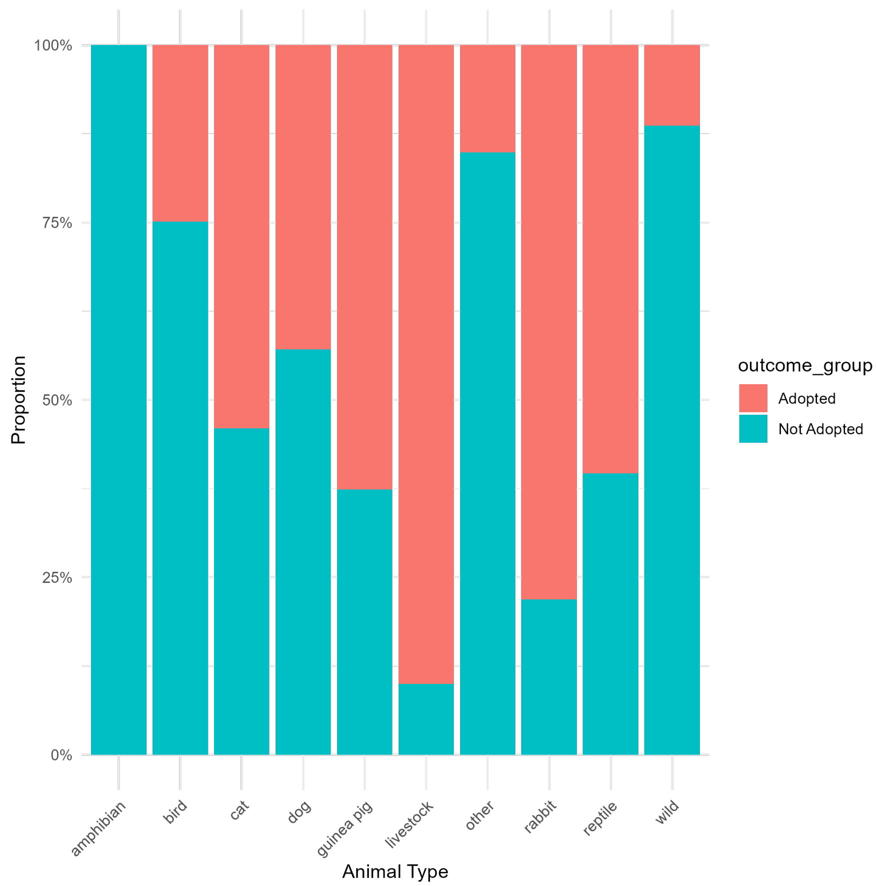

Figure 2: Animal Type vs Outcome Group (adopted / not adopted)

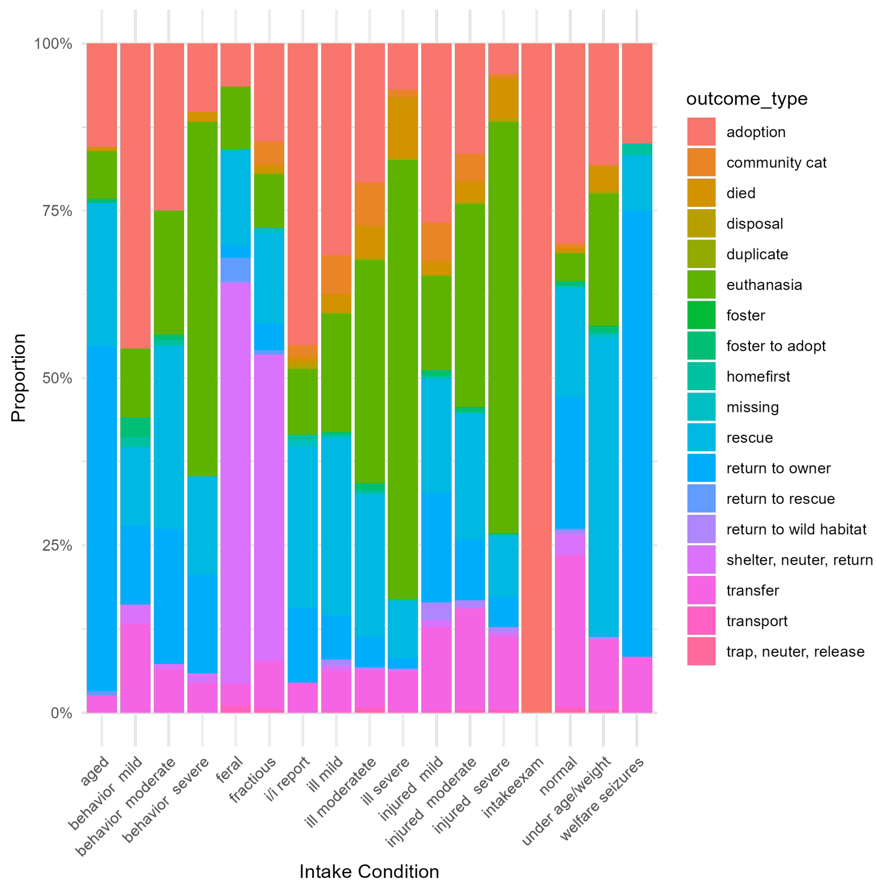

Figure 3: Intake Condition vs Outcome Type

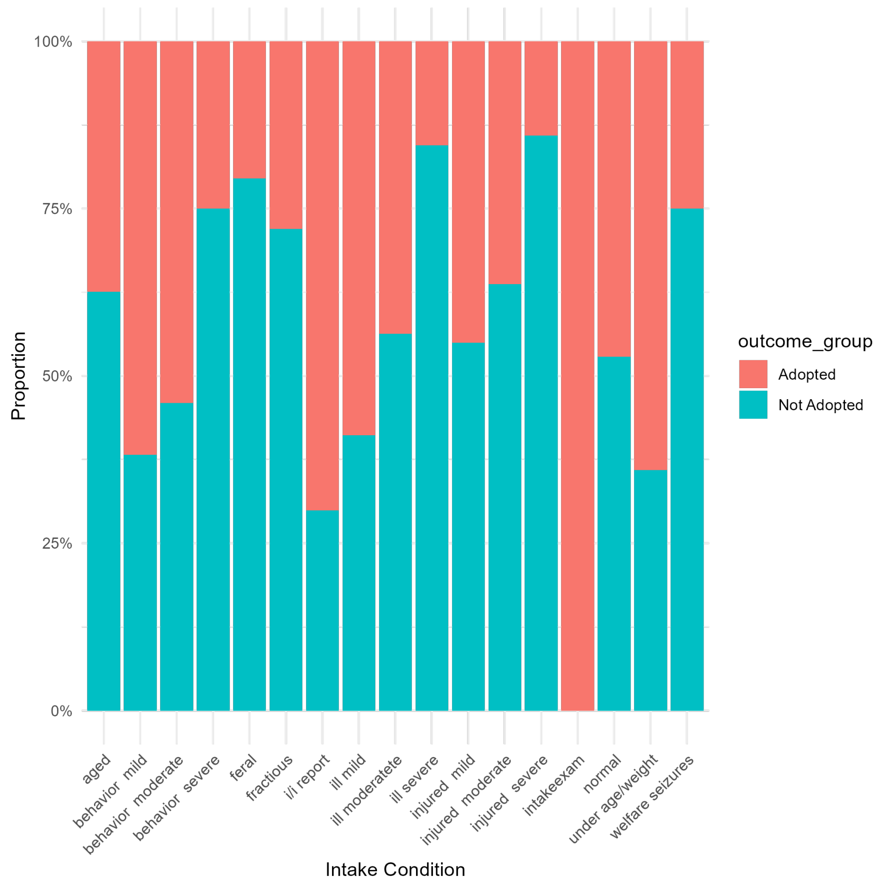

Figure 4: Intake Condition vs Outcome Group (adopted / not adopted)

while at the animal shelter or people being less likely to adopt animals that might need more care (whether thats due to financial concerns or otherwise). This is important as it shows that animal shelter could focus on helping these animals get adopted more often through appealing emotionally to the public or expanding their reach to 'market' the shelter animals to a wider audience that may be more willing to take in animals that require more care.

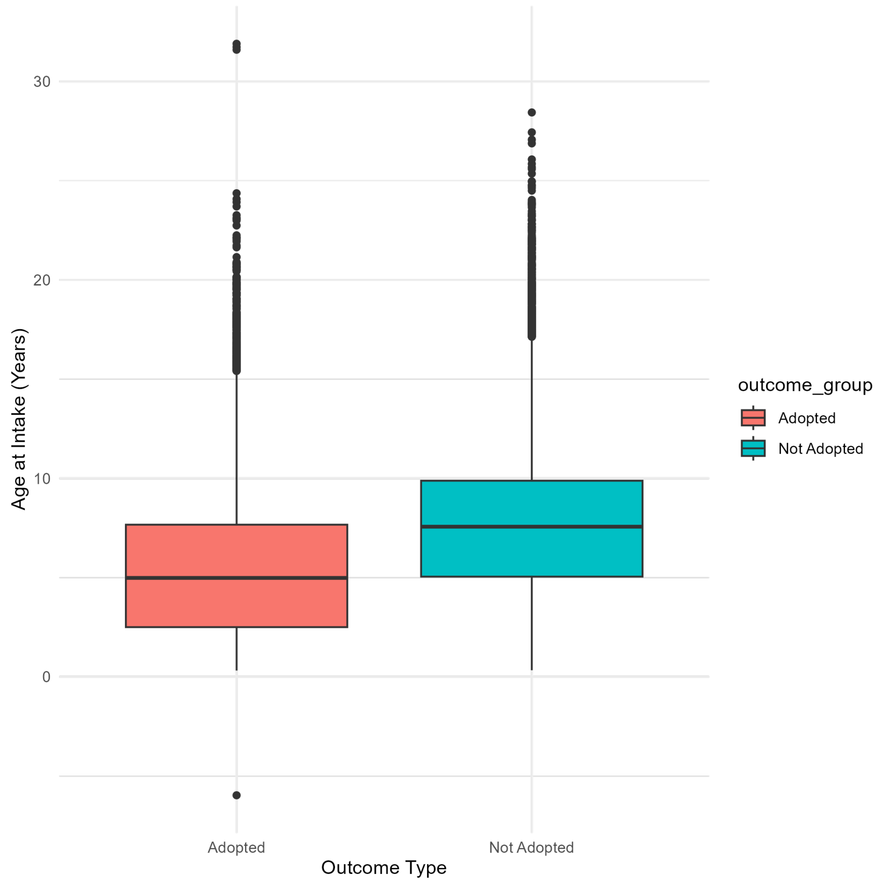

Figure 5: Age at Intake vs Outcome Group (adopted / not adopted)

Explanation of Relevance: Based on Figure 5, we can see that the median age for animals that end up being adopted is slightly lower than those that don't end up getting adopted. However, both groups have a large number of outliers for much older animals. The adopted group has a lower median age and a wider range toward younger ages. This is relevant because it shows that there is a clear preference for younger animals, and animals that are on the older side don't tend to be adopted as frequently. This means there could be an improvement in how the animal shelter uses its resources in promoting the adoption of older animals.

Feature Engineering

Studies have shown that the longer animals have been in a shelter, the less likely they are to be adopted. In order to utilize this knowledge as a feature we have created the `days_in_shelter` feature by finding the number of days they were in the shelter from their intake day to their outcome day.

Additionally, many pets are more likely to be adopted when they are young, with older animals being less likely to get adopted. If we were to just use their date of birth as the predictor it would not actually account for how old the animal is was when the left the shelter. To better represent that age we created the `age_in_days feature`, which calculates the number of days old the animal was one their outcome day using dob and outcome_date.

Classification Analysis

We currently have one large data set that contains all of our animal shelter observations. In order to be able to gauge the performance of our classification model we split it into two data set: training data and validation data.

We used the training data to build our model and tune our model and save our validation data to test the final accuracy of the model on new, unseen data. This gives us an estimate of how the model would perform when given unclassified, 'real-world' data.

Next, we have to build our classification model. We used K-Nearest Neighbors as our classification model as it doesn't have an specific assumption about data distribution, it is robust to noise data, it is easy to implement and tune, and it is an efficient model since it does not require a training phase.

To tune our model, we found the accuracy of the model using different numbers of neighbors (k) from 1 to 25. The higher the k value, the more complex our model becomes, so it is important to ensure that the k is not too high and causing the model to over fit to our training data, and not be general enough to accurately predict the training data.

Using Figure 6, we found that the best number of neighbors (k) for our classification is 5. This is the point where any significant improvement in accuracy begins to plateau and choosing a k value that is larger could potentially lead to over fitting the model. Using this k, we fitted our model on the training data and generated predictions based on our test data.

```
Warning in attr(x, "align"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

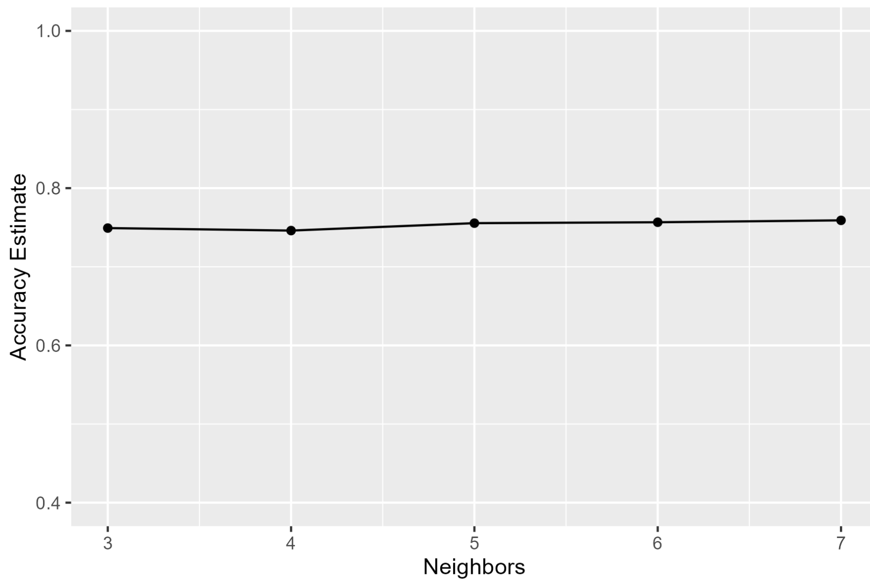

Figure 6: Elbow Plot of various neighbor numbers

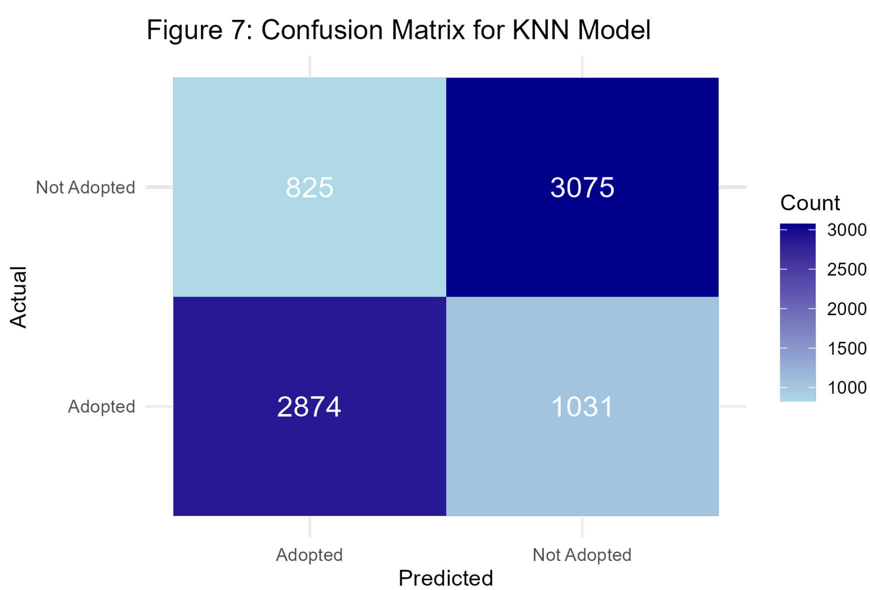

Figure 7: Confusion Matrix for our model

```
Warning in attr(x, "format"): 'xfun::attr()' is deprecated.
Use 'xfun::attr2()' instead.
See help("Deprecated")
```

Table 5: Summary statistics about our model

.metric	.estimator	.estimate
accuracy	binary	0.7622037
precision	binary	0.7359795
recall	binary	0.7769667

To visualize our results, we created a confusion matrix (Figure 7) to assess our model's performance and some summary statistics (Table 5) to assess our model's accuracy.

Discussion

With a final accuracy of around 0.76, we can fairly confidently state that we were able to successfully predict whether or not an animal would be adopted, at least to an extent. Unfortunately, one of the shortcomings of our chosen analysis model is the lack of feature importances. K nearest neighbours is a non-parametric model, meaning that it does not use coefficients to make its predictions, and as such it is unclear which features in this model contributed the most to this final result. However, what we have produced still serves as a valuable tool for shelter owners as a pre-trained model designed to predict whether or not an animal will be adopted. Overall, we are glad to see that our model functions well as it currently exists, but room for improvement still remains. An accuracy of 0.76 is good, but not fantastic, and we would be able to gain even more information by utilizing a model that would allow us to determine which specific factors contribute to an animal being adopted or not. These are important questions to consider in the future, and can be easily developed with further analysis of this data.

References

"2023 State of Shelter Adoption Report." n.d. Hill's Pet Nutrition. https://theaawa.org/wp-content/uploads/2023/11/hills-pet-nutrition-2023-state-of-shelter-adoption-report.pdf.

Burch, Kelly. n.d. "Pet Euthanasia Rates Are at a 3-Year High, but Experts Say There Are Ways to Help Besides Adoption." Shelter Animals Count. https://www.shelteranimalscount.org/pet-euthanasia-rates-are-at-a-3-year-high-but-experts-say-there-are-ways-to-help-besides-adoption/.

"Effective Tips to Boost Adoptions: Reduce Barriers & Improve Experience." n.d. ASPCApro. https://www.aspcapro.org/resource/effective-tips-boost-adoptions-reduce-barriers-improve-experience.

"Here's What's Happening in Animal Shelters Across the u.s." n.d. Best Friends: Save Them All. https://bestfriends.org/no-kill-2025/animal-welfare-statistics#:~:text=Each%20day%20in%202023%20across,dogs%20and%20cats%20last%20year.

Lydia Gibson, City of Long Beach Animal Care Services. 2025. "Long Beach Animal Shelter." Open Data. https://github.com/rfordatascience/tidytuesday/tree/main/data/2025/2025-03-04.

"Saving Companion Animals from Inhumane Conditons." 2018. Rescue Paws foundation. https://www.ourcauseforpaws.org/blogs/news/saving-companion-animals-from-inhumane-conditons?gad_source=1&gclid=Cj0KCQiA2oW-BhC2ARIsADSIAWq1F3Lnd2Skw3YVgVHO5
1YCFx22YWl-eJ9mrUaJ4HX_dxgdrOYaAu1YEALw_wcB.