Predicting the Adoption Rate of Shelter Pets

CS 573 Final Project

Mohit Gupta & Nandita Viswanath

Agenda

- ➔ Introduction
- ➔ Data Exploration
- ➔ Data Preprocessing
- ➔ Results and Observations

Agenda

- ➔ **Introduction**
- ➔ Data Exploration
- ➔ Data Preprocessing
- ➔ Results and Observations

Background

6.5 Million pets are abandoned at shelters every year.

Every year**, 1.5 million** of them are **euthanized.**

- ➔ Help shelters forecast the resources required for pet care.
- ➔ Identify key factors that affect the pet advertisement.
- ➔ Create awareness and reduce the number of pets that are euthanized.
- ➔ **Dataset:** Pet advertisement dataset provided by petfinder.my on Kaggle.

Agenda

- ➔ Introduction
- ➔ Data Exploration
- ➔ Data Preprocessing
- ➔ Results and Observations

Agenda

- ➔ Introduction
- ➔ **Data Exploration**
- ➔ Data Preprocessing
- ➔ Results and Observations

Data Exploration

- ➔ Understanding the Dataset
- ➔ Data Visualization

Data Exploration

- **Understanding the Dataset**
- Data Visualization

Understanding the Dataset

Dataset Features: (24 total)

- **Type** (cat or dog)
- **Gender, Age** (months)
- **Name, Advertisement Description**
- **Breed** (Breed ID1 & Breed ID2)
- **Color, Fur Length**
- **Maturity Size**
- **Medical Attributes** (vaccinated, sterilized, health condition, etc.)
- **Adoption Fee**

Class Label - Adoption Rate

- **0 -** Adopted within 24 hours.
- **1 -** Adopted within the first week.
- **2 -** Adopted between 1 week and 1 month.
- **3 -** Adopted between 1 and 2 months.
- **4 -** No adoption after 100 days.

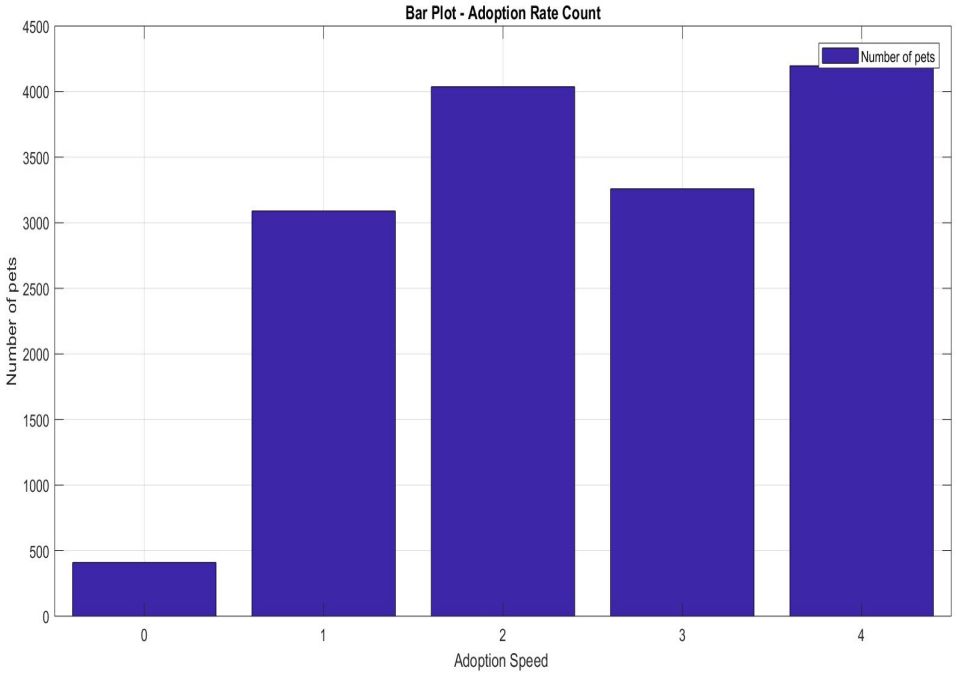
Data Exploration

- ➔ Understanding the Dataset
- ➔ Data Visualization

Data Exploration

- ➔ Understanding the Dataset
- ➔ **Data Visualization**

Bar Plot: Adoption Speed

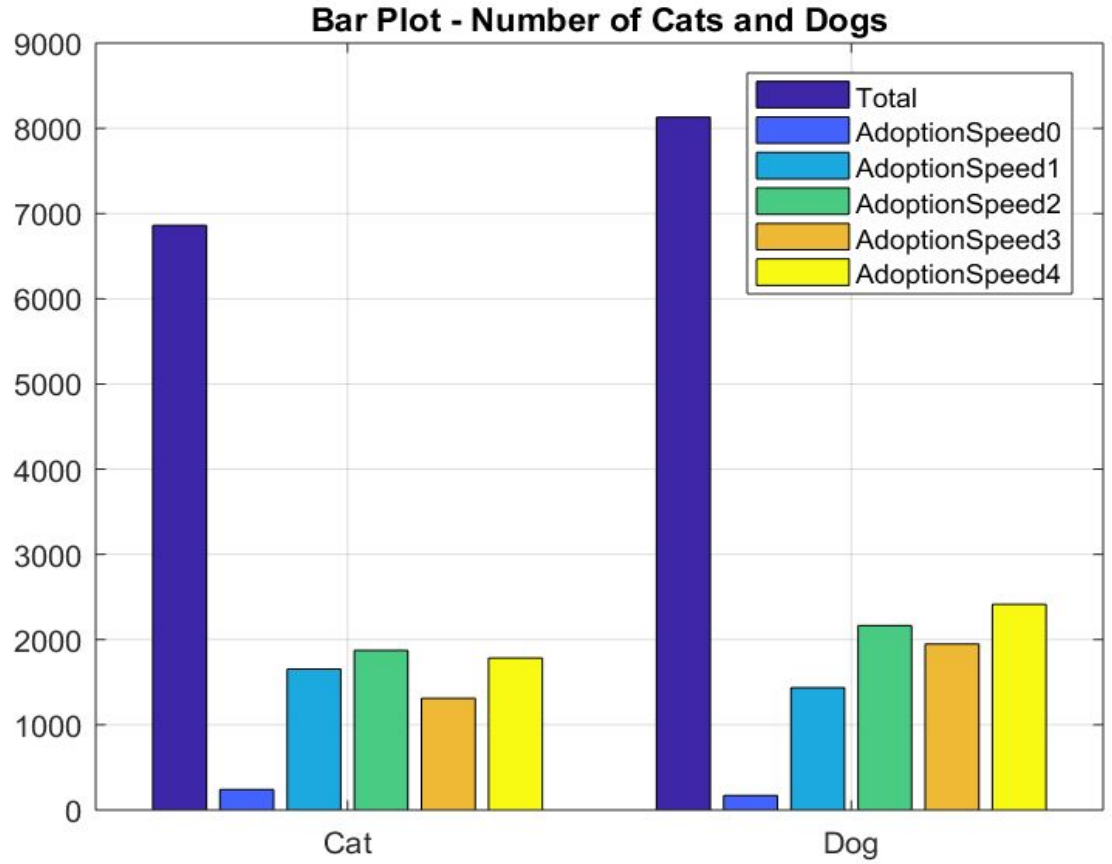


Plot Tips:

- Adoption speed can be treated as a categorical variable with the categories being 0, 1, 2, 3 and 4 with 0 being the highest adoption speed (quick adoption) and 4 being not adopted..

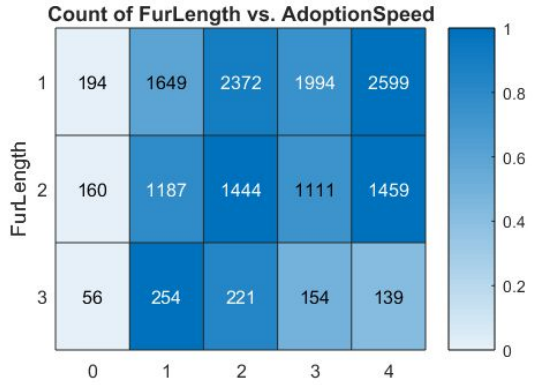
- More pets fall within the adoption rate 4 category than any other - ie. ~**4200 pets out the total 14000 pets** are not adopted even after **100 days** of being listed**.**
- The second highest number of pets fall in the adoption speed 2 (adopted within 1 month of being listed)
- Adoption speed 1 (adopted within 7 days) and 3 (adopted within 90 days) have almost ~3100 and ~3250 pets in their respective categories
- **Lesser than 500 pets** are adopted at **the best adoption speed** 0 (adopted on the same day as listed).

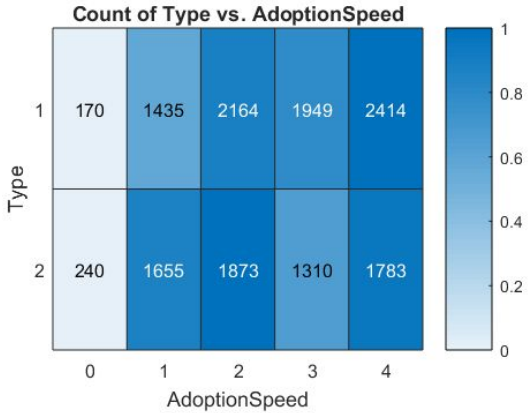
Stacked Bar Plot: Pet Type and Adoption Speed

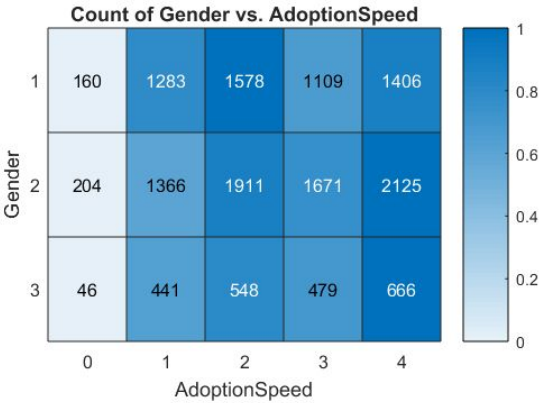


- Visualize relationship between adoption rate and pet type
- More dogs than cats in dataset
- Cat Population: Adoption Speed 2>4>1>3>0
- Dog Population: Adoption Speed 4>2>1>3>0
- **Adoption speed 2 most prevalent in cats** and **4 most prevalent in dogs.**

2-D Heat Map: Adoption Speed vs Categorical Features







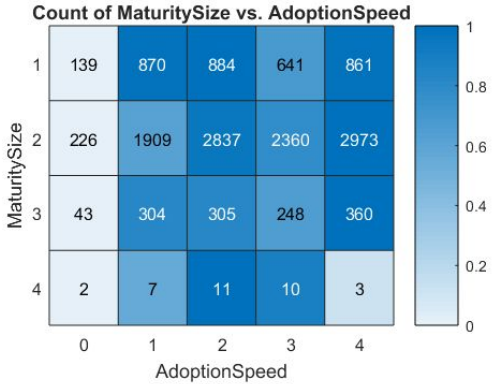
Plot Tips:

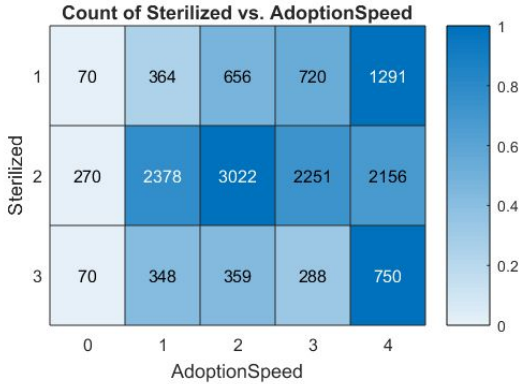
- X-axis: Adoption Speed
- Y-axis: Categorical feature
- Numbers in boxes can be used to interpret size of population in a particular category.
- Heat maps can be used to visualize population distribution and capture high-level trends.

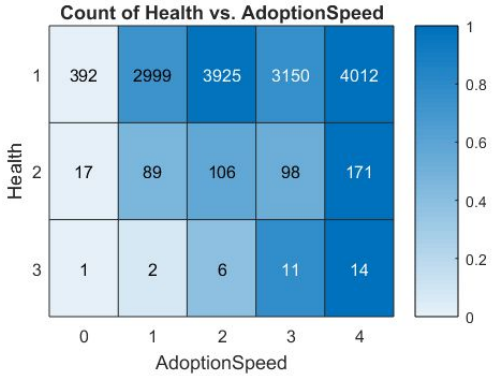
Inference:

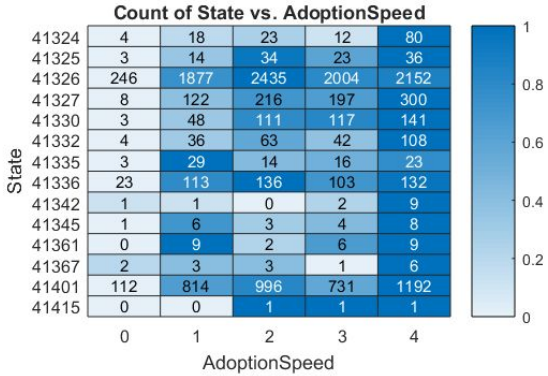
- A higher proportion of pets with fur length 3 are adopted at speed 1 - ie. Pets with fur length 3 are adopted quickly but the trend does not hold true for pets with fur lengths 1 and 2.
- A higher proportion of cats are adopted at quicker speeds than dogs
- For genders 2 (female) and 3 (others), most pets have adoption speed 4. For pets with gender 1 (male), more pets are adopted at higher speeds of 1 and 2

2-D Heat Map: Adoption Speed vs Categorical Features





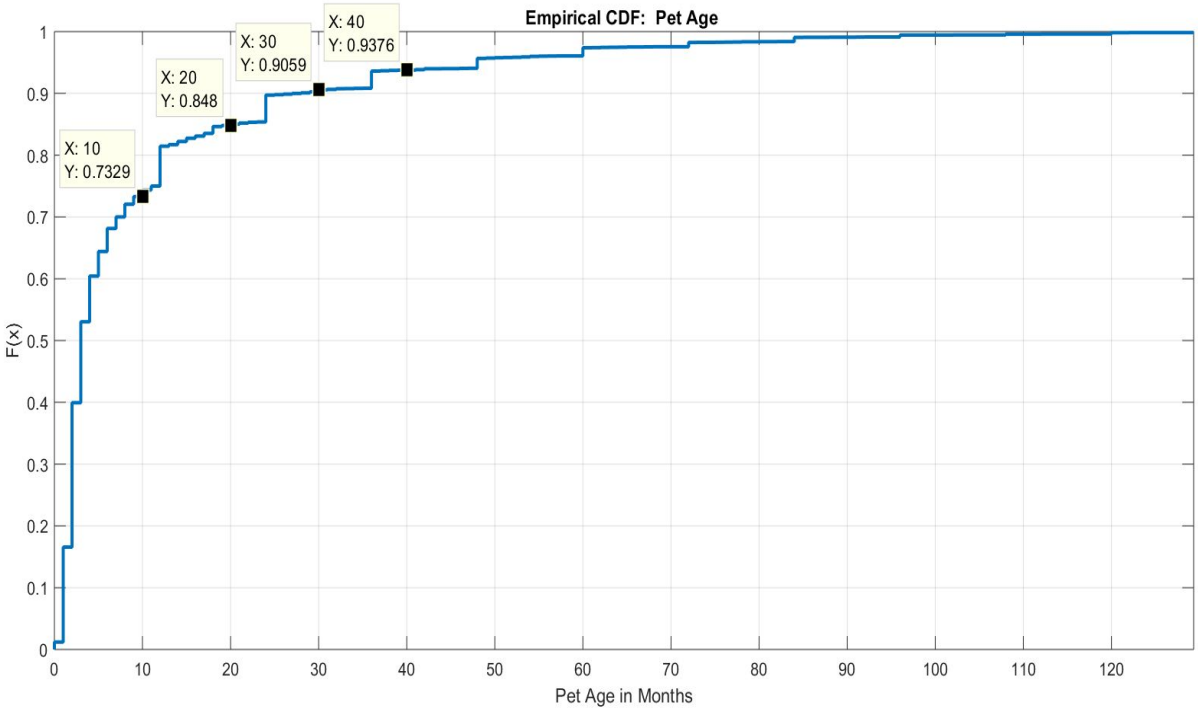




Inference:

- Most pets are of Maturity Size 2 (medium sized) followed by 1 (small), 3 (large) and 4 (extra large).
- More pets are not sterilized and are adopted at higher speeds when compared to sterilized pets
- Most pets are left at the shelter in healthy condition (Health = 1), health condition does not show a trend with respect to adoption speed
- 58.12% of all pets are found in state location 41326 and 25.6% of all pets are found in state location 41401.

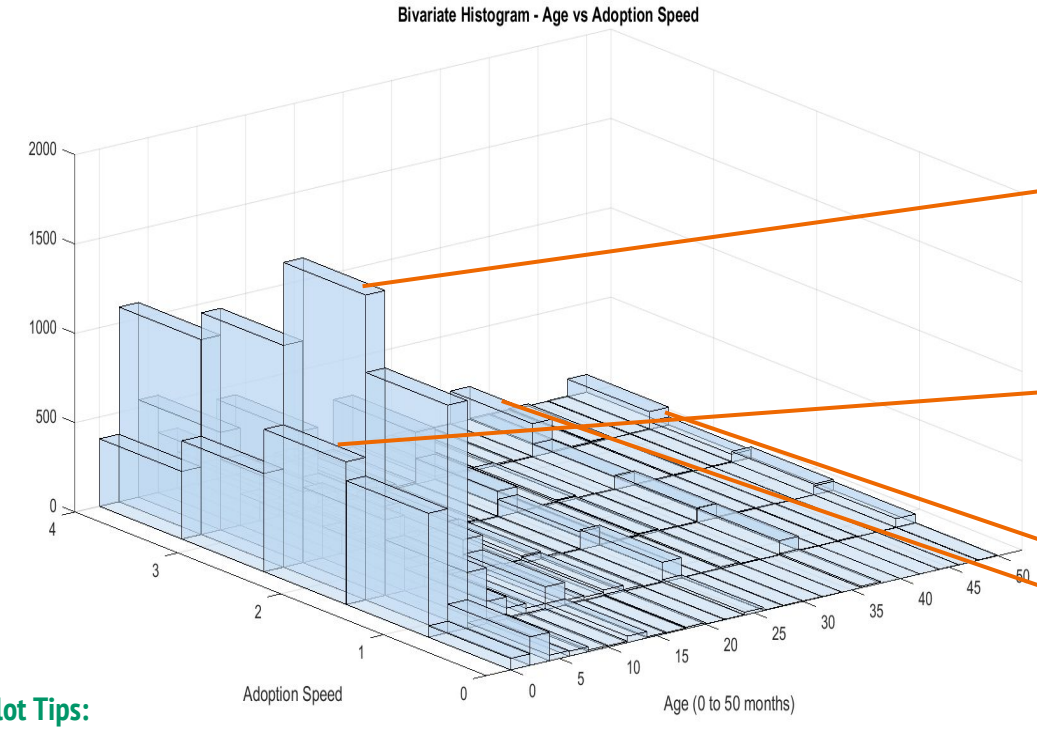
Age distribution: Cumulative Distribution Function



Inference:

- **73.29%** of the total number of pets are **lesser than 10 months old.**
- 11.51% of the pets are between 10 and 20 months old
- 9.41% of the pets are greater than 30 months old and 6.24% of the pets are less greater than 40 months old
- Since majority of the pets are lesser than 50 months old, pets lesser than 50 months old are visualized in a bivariate histogram shown in the next slide

Bivariate Histogram: Pet Age vs Adoption Speed

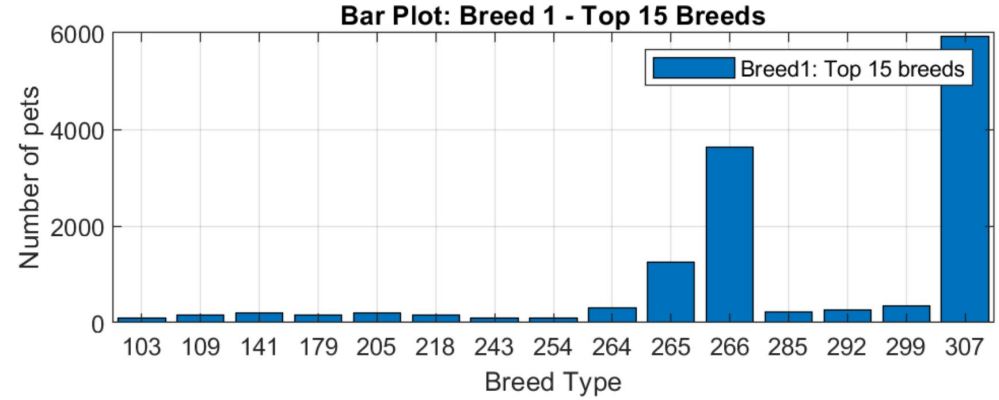


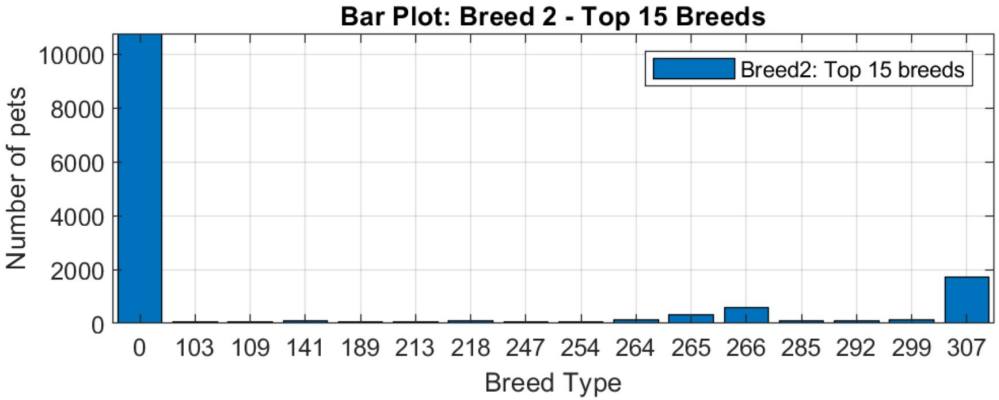
- Bin size in X-axis is two months
- For pets that are between 0 and 4 months old, most frequent adoption speed is 2 followed by 1, 4 and then 0.
- For pets that are between 4 and 6 months old, more pets are adopted at adoption speed 3 followed by 4, 2,1 and 0
- For pets that are older than 6 months, the trend suggests that they are adopted at slower speeds as more pets fall in the adoption speed 4 category than any other speed category

Plot Tips:

- X-axis: Age (0-50 months)
- Y-axis: Adoption Speed.
- Z-axis: Number of pets.

Breed Population





- For 14993 pets, there are 188 different breed types.
- Breed 1 has 176 different types of breeds and Breed 2 has 135 different types of breeds.
- **13114 pets** fall within the **top 15 types of Breed 1** breed types.
- **14296 pets** fall within the **top 15 types of Breed 2** breed types.
- Among Breed 1 breed types, 307 is the most common followed by 266 and 267.
- Among Breed 2 breed types, 0 is the most common breed type followed by 307 and 266.

Agenda

- ➔ Introduction
- ➔ Data Exploration
- ➔ Data Preprocessing
- ➔ Results and Observations

Agenda

- ➔ Introduction
- ➔ Data Exploration
- ➔ **Data Preprocessing**
- ➔ Results and Observations

Data Preprocessing

One hot Encoding:

- **Type** (cat or dog)
- **Age** (months)
- **Gender**
- **Breed** (Breed ID1 & Breed ID2)
- **Color, Fur Length**
- **Maturity Size**
- **Medical Attributes** (vaccinated, sterilized, health condition, etc.)
- **Adoption Fee**

TF-IDF & Word2Vec Embedding:

- **Advertisement Description.**

Ablation Study on Name Feature:

- **Removed Name feature** since it was not useful after performing an ablation study on the same.

Agenda

- ➔ Introduction
- ➔ Data Exploration
- ➔ Data Preprocessing
- ➔ Results and Observations

Agenda

- ➔ Introduction
- ➔ Data Exploration
- ➔ Data Preprocessing
- ➔ **Results and Observations**

Results and Observations

- **Algorithms Implemented:**
  Support Vector Machine
  Decision Tree
  Neural Network

- **Evaluation Criteria:**
  Micro F1 Score

- **Other Evaluations:**
  TF-IDF vs Word2Vec Embeddings
  Binary Classification Results

Models - Performance Comparison

Support Vector Machine

- **Model Parameters**
  Kernel: RBF (Gaussian).

- **Results**
  F1 Score = 0.32

Decision Tree

- **Model Parameters**
  Criterion: Gini Gain.

- **Results**
  F1 Score = 0.37

Artificial Neural Network

- **Model Parameters**
  Two Hidden Layers: 2000 & 500 neurons.
  Learning Rate: 1e-5.

- **Results**
  F1 Score = 0.42

Note:
These results are based on Word2Vec preprocessing applied to the advertisement description attribute.

Observations - Word2Vec Embedding vs TF-IDF

Word2Vec Embedding

- We convert the “Description” attribute to dense 300 dimensional vectors using the skip-gram model trained on Google News corpus**.**

Results:

SVM F1 Score:	**0.32**
Decision F1 Score:	**0.37**
Neural Network F1 Score:	**0.42**

TF-IDF

- We use Bag of Words representation for the “Description” attribute weighted using Term Frequency Inverse Document Frequency (TF-IDF).

Results:

SVM F1 Score:	**0.33**
Decision Tree F1 Score:	**0.35**
Neural Network F1 Score:	**0.44**

Observations - Binary classification

Binary Classification - Classifying whether the pet will be adopted within 100 days.

- **Class Labels:**
  1 → Pet adopted within 100 days.
  0 → Pet not adopted after 100 days.
- **Results with Word2Vec Embedding on the description attribute:**

SVM F1 Score:	0.71
Decision Tree F1 Score:	0.69
Neural Network F1 Score:	**0.77**

- **Results with TF-IDF preprocessing on the description attribute:**

SVM F1 Score:	0.71
Decision Tree F1 Score:	0.70
Neural Network F1 Score:	**0.78**

Questions?

Thank you!

References

1. Dataset: https://www.kaggle.com/c/petfinder-adoption-prediction/data
2. Introduction statistics: https://www.aspca.org/animal-homelessness/shelter-intake-and-surrender/pet-statistics