Shelter Dog Recommendation System

C964 – Computer Science Capstone

Andrew Thaxton

Table of Contents

A1. Letter of Transmittal

Mr. John Smith

Humane Society

1234 56th St

Indianapolis, IN 46236

Dear Mr. Smith.

The current state of animal shelters across the country is disheartening to say the least. They are severely overcrowded leading to animals in need being turned away or euthanized, depending on the shelter. People are choosing to get their dogs from a breeder to get a specific type of dog that has specific traits without considering how many other breeds have those same traits or similar traits and are easily found in shelters.

I have come up with a solution to this issue. I propose a program that learns the traits of various dog breeds from the data that we provide. The program will then use its knowledge and the user's inputted preferred dog breed to recommend one of the top five dog breeds found in shelters. Finally, the program will display some visualization tools so the user can see the accuracy of the results and a comparison in each of the traits used to define the dog breeds. This will help educate future dog owners that there are similar dogs to the ones they desire and encourage them to consider adopting from the shelter. This will lead to more future pet owners to adopt reducing the overcrowding in the shelter and opening availability to those dogs in need.

The funding required to implement this shelter dog recommendation system will be approximately $11,000. Most of the funding will be used on development costs and man-power related costs. It will include any development time determined to be in the scope of the project. My knowledge and experience in python and working with cross-functional teams will make me an ideal leader for a team to develop this system as a solution to the overcrowding of animal shelters across the country.

I look forward to hearing from you and hope to have the opportunity to work with you on this project. Please let me know if you have any questions or concerns.

Sincerely,

Andrew Thaxton

Lead Developer

A2. Project Proposal

Problem Summary

Animal shelters across the United States are struggling with the overcrowding of animals. This leads to cages lining the hallways, multiple animals to a single cage, animals in need being turned away, and in many cases, animals being deemed unlikely for adoption and being euthanized. This is the sad truth as people look to breeders for specific dog breeds that they are familiar with as opposed to looking at their local dog shelter for a dog in need that has the same or very similar traits.

Data Product Benefit

This system will take the knowledge it has built using the data that we provide it and the user's inputted preferred dog breed and provide one of the top five shelter dog breeds that has very similar traits. It will even provide visualization tools that the user can review and be able to see how similar the dog breeds are and in what ways they are most similar. The shelter dog recommendation system looks to educate potential adopters and encourage them to consider a dog from their local shelter and cut down on the overcrowding situation.

Data Product Outline

The data product will be developed using Python along with their data analysis libraries. The program will use Jupyter Notebook, an interactive data science environment. Jupyter Notebook can separate code into segments of code that can run separately, allowing for incremental development. The program can be run from inside Jupyter Notebook. After initiating some of the code the user will be prompted by the command line interface to provide their desired dog breed. Once provided, the program can continue to be run and will provide the recommended shelter dog breed, a bar chart comparing the two dog breeds' traits, a pie graph displaying the percent similarity between the two, and a pie graph comparing the top fifteen shelter dog breeds to the total number of dog breeds. All these tools will serve their primary purpose to educate the user about shelter dog breeds and encourage them to consider looking at their local dog shelter.

Data Description

The dataset used to train the shelter dog recommendation system was acquired from Kaggle.com and will be prepped manually by myself to be used as a training dataset. The dataset contains a ranking of thirty different traits on a scale from one to five for each of the dog breeds along with ranking for cat breeds which will be discarded as unneeded data. I will use this information to provide which of the five shelter dog breeds should be recommended based on these traits to be used as training data for the system. The traits to be included are:

- Adaptability
- All Around Friendliness

- Exercise Needs
- Health Grooming
- Trainability
- Adapts Well to Apartment Living
- Affectionate with Family
- Amount Of Shedding
- Dog Friendly
- Drooling Potential
- Easy To Groom
- Easy To Train
- Energy Level
- Friendly Toward Strangers
- General Health
- Good For Novice Owners
- Incredibly Kid Friendly Dogs
- Intelligence
- Intensity
- Potential For Mouthiness
- Potential For Playfulness
- Potential For Weight Gain
- Prey Drive
- Sensitivity Level
- Size
- Tendency To Bark Or Howl
- Tolerates Being Alone
- Tolerates Cold Weather
- Tolerates Hot Weather
- Wanderlust Potential

The data is in a JSON format with three levels. The first being dog breed vs cat breed, the second being the different dog breeds, and the third being the list of traits and their ranking. As I mentioned, the data will need to be prepped to discard unnecessary, duplicate, or bad data. The data will also need to be provided with the shelter breed that should be assigned to each dog breed to be used as training data.

Objective and Hypothesis

The primary objective of the shelter dog breed recommendation system is to educate potential adopters about shelter breeds and their similarity to breeds that the potential adopter is familiar with. It will serve to encourage potential adopters to consider looking at their local shelter before going to a breeder to find their new dog. The shelter dog recommendation system hypothesizes that if a user prefers a particular dog breed that they will also like a similar dog breed that is found in shelters. By recommending a similar dog breed that can easily be found in shelters more potential adopters will look

to their local shelter before going to a breeder. This will reduce the overcrowding in shelters opening space for more dogs in needs and reducing the need for euthanasia.

Project Methodology

The development methodology that will be used to develop the shelter dog recommendation system is the agile methodology. This methodology was selected because it can be integrated with Jupyter Notebook's features for continuous testing through running individual segments of code. The agile methodology is based on iterative development that is broken into several phases. It required constant collaboration between stakeholders and cross-functional teams that will constantly test and improve the project. The aspect of the agile methodology that is most important is the collaboration between the development team and the customer. This allows for flexibility and quality of the project as there is a constant stream of feedback at each phase. The phases of agile are:

1. Requirements: Gather requirements of the project from the customer.
2. Development: Develop the software based on the requirements.
3. Testing: Test using quality assurance to check for errors and the ability to meet the user's acceptance criteria.
4. Delivery: Integration of the project.
5. Feedback: Receive feedback from stakeholders to generate requirements for the next iteration.

Funding Requirements

This project will require funding of $11,000. This amount will cover the development cost of the product, including the time used by the development team, quality assurance team, and integration team. If any out-of-scope requirements are added during the agile development cycle, additional funding may be required. This will be determined by the complexity of the requirement and any extra time needed.

Solution Impact

The shelter dog recommendation system with have a tremendous impact on shelters and potential adopters. Potential adopters will be able to make educated decisions when looking for the right dog for them. They can use the knowledge that they already possess and expand it by comparing the breeds they are familiar with to those found in shelters. The shelters will benefit from more adoptions coming through and reducing the overcrowding problem that was leading to drastic measures.

Data Communication Precautions

The data that will be used in the development of the shelter dog recommendation system is publicly available data that does not violate HIPPA, FERPA, or PCI DSS regulations. As this system will

also be publicly available and does not contain any information on users no precautions will be necessary.

Developer's Expertise

The developers that will be working on the development team are all college graduates with varying experience. All developers have some experience with Python though not all are familiar with Python's extensive machine learning library. They all have experience in agile development which will be essential to timely development of this project. The development team will be led by myself as I have extensive knowledge and experience with Python, machine learning, and agile along with my leadership skills and experience with previous project. We will have the final product completed in an efficient and timely manner with minimal to no setbacks.

B. Project Proposal

Problem Statement

The shelter dog recommendation system is a much-needed addition to the Humane Society that will improve the adoption rate of shelter dogs and reduce the overcrowding in the shelters. This system will utilize the K nearest neighbor machine learning algorithm to recommend a shelter dog based upon the user preferred dog breed. The system will serve to educate potential adopters that there are dogs commonly found in shelters that have similar traits to dogs they may be familiar with. There will be visualization aids that will show comparisons between the breed they select and the recommended shelter breed along to further educate the potential adopters.

Customer Summary

The shelter dog recommendation system will be used by the potential adopters that come into the Humane Society to look at some dogs. It will be an additional feature that the potential adopters can utilize when considering different dogs. This system will take what the customer knows already and use that information to provide them with alternative options that they may not know to consider without any need for special skills.

Existing System Analysis

The shelter dog recommendation system will be a new addition to the shelters. Currently there is no such system in place. Potential adopters must rely on their current knowledge and trust the knowledge of the volunteers which can vary based on biases. This system will be able to provide an unbiased recommendation based on machine learning algorithms.

Data

The data used to develop the shelter dog recommendation system is a dataset available on Kaggle.com and can be found at the following link: https://www.kaggle.com/hocop1/cat-and-dog-breeds-parameters

The data includes is a JSON file containing information on cats and dogs breeds, however only the dog breeds portion will be used. There are ratings on a one to five scale on over 270 different dog breeds in 30 different traits including:

- Adaptability
- All Around Friendliness
- Exercise Needs
- Health Grooming
- Trainability
- Adapts Well to Apartment Living
- Affectionate with Family
- Amount Of Shedding
- Dog Friendly
- Drooling Potential
- Easy To Groom
- Easy To Train
- Energy Level
- Friendly Toward Strangers
- General Health
- Good For Novice Owners
- Incredibly Kid Friendly Dogs
- Intelligence
- Intensity
- Potential For Mouthiness
- Potential For Playfulness
- Potential For Weight Gain
- Prey Drive
- Sensitivity Level
- Size
- Tendency To Bark Or Howl
- Tolerates Being Alone
- Tolerates Cold Weather
- Tolerates Hot Weather
- Wanderlust Potential

As I mentioned previously, there is some data that is unnecessary and will be removed. I will also use the Euclidean distance algorithm to determine which of the top five shelter dog breeds is closest to each of the dog breeds. This is necessary to include as they will be used as labels in the training data for the k nearest neighbor machine learning algorithm.

Project Methodology

The project will be developed using the agile methodology. This is to ensure consistent communication between the development team and the stakeholders to ensure the requirements are met and any issues that arise are dealt with efficiently and communicated accordingly. The phases of the agile methodology are:

1. Requirements: The development team will work with the stakeholders to determine and define requirements. These requirements are subject to change throughout the process based upon stakeholder feedback.
2. Development: The program will be developed based upon the designated requirements
3. Testing: There will be continuous testing performed by the quality assurance team using black box and white box testing accordingly to test logic and usability.
4. Delivery: The program will be delivered to the customers for acceptance testing. Should the program fail this test it will immediately be sent back to be adjusted. If it passes, then it will be implemented in all the shelters.
5. Feedback: The feedback will be provided by the stakeholders and used to determine if any additional requirements are necessary, which may also lead to additional funding being necessary.

Project Outcomes

The project deliverables will include any documents/programs that are created during development. This will include a schedule of project development, the budget distribution, documentation of problems that arise, storyboards, and flowcharts.

The product deliverables will include the shelter dog recommendation system, the source data, and the dataset used to train the machine learning system.

Implementation Plan

We will be utilizing the agile methodology for our implementation plan. The requirements will be determined by the stakeholders and the development team to begin. The rollout will consist of three separate phases: prototype, testing phase, and final distribution.

As soon as development begins the development team will work closely with the stakeholders to consistently test and improve the prototype. Testing will then be conducted by the quality assurance team to ensure the functionality and logic have minimal errors. It will then be rolled out to a single shelter for acceptance testing. Once the program has been determined to be accepted it will be rolled out to the rest of the shelters across the country.

The most critical deliverable will include a complete and fully functional shelter dog recommendation system along with the documentation of issues and implementation. Other deliverables will include documents from the planning and designing phases of the product development.

Evaluation Plan

Using the agile methodology, the quality assurance team will test the program consistently throughout all the stages of development. Any errors or incomplete requirements will be identified through communication between the development team and stakeholders to be fixed before the final product is deployed.

The shelter dog recommendation system will be evaluated based on the accuracy of the shelter dog recommendations. This can be determined by taking the traits of the shelter dog and the user's selected dog and comparing them. This can be done for each individual trait the combined percent difference between the two breeds' traits. As we observe the results, we can adjust the k value for the k nearest neighbor algorithm for more accurate results.

Resources and Costs

Programming Environment

The data product will utilize Python 3 and its data analysis and machine learning libraries and Jupyter Notebook to develop the shelter dog recommendation system. These resources are available to Mac OS and Windows OS computers, are open source and free of cost. This means the programming environment will cost a total of $0.

Environment Cost

The shelter dog recommendation system will be hosted on local computers at each of the shelters. The shelters should be able utilize the computers that they currently use to fill out adoption applications and other paperwork to host this new program as well. Leaving the environment cost at $0.

Human Resource Requirements

The cost of this project will primarily come from staff development hours. This will include the 90 hours put in by the development team coming out to $9,000 and the 20 hours put in by the testing team coming out to $2,000 for a grand total of $11,000.

Timeline and Milestones

The timeline and milestones are shown in the table below. The approximate time for this project is about three weeks.

Event	Start Date	End Date	Duration (hours)	Dependencies	Assigned Resources
1. Project requirement meeting	2/15/2022	2/15/2022	5	NA	Development team, Quality Assurance team, Stakeholders

2.Project kickoff meeting	2/16/2022	2/16/2022	5	Task 1	Development team, Quality Assurance team, Stakeholders
3.Planning (Start phase 1)	2/17/2022	2/18/2022	10	Task 2	Development Team
4.Design	2/19/2022	2/20/2022	10	Task 3	Development Team
5.Data Analysis	2/20/2022	2/22/2022	10	Task 4	Development Team
6.Shelter Dog Recommendation System development	2/23/2022	2/25/2022	15	Task 5	Development team
7.Deliver prototype (End phase 1)	2/26/2022	2/27/2022	5	Task 1-6	Development team, Stakeholders
8.QA testing (Start phase 2)	2/28/2022	3/1/2022	10	Task 1-7	Quality Assurance Team
9.Review and adjust for feedback	3/2/2022	3/3/2022	10	Task 1-8	Development team, Stakeholders
10.Deliver to first shelter (End phase 2)	3/4/2022	3/5/2022	5	Task 1-9	Development team, Stakeholders
11.Acceptance testing and debugging (Start phase 3)	3/6/2022	3/7/2022	10	Task 1-10	Development team, Stakeholders
12.Final delivery and implementation (End phase 3)	3/8/2022	3/9/2022	5	Task 1-11	Development team, Stakeholders
Total			100		

D. Post-Implementation Report

Project Purpose

The purpose of this project was to create a shelter dog recommendation system that utilized machine learning algorithm to recommend one of the top five shelter dog breeds based on a dog breed

that the user entered and to provide some visualization tools that the user could review to compare their desired dog with the dog recommended by the system.

This shelter dog recommendation system was intended to be used as an educational tool for potential adopters to use. Originally, potential adopters would have to rely on their own knowledge and potentially the knowledge of the Humane Society's volunteers to decide on a dog to adopt. With the new system, the potential adopter can use what they know and discover alternative breeds that can easily be found in shelters. This would lead to an increase in adoption rate, therefore, reduce the overcrowding in the shelters.

Dataset

The dataset used to develop this shelter dog recommendation system was from Kaggle.com at https://www.kaggle.com/hocop1/cat-and-dog-breeds-parameters. The data was provided in a JavaScript Object Notation (JSON) file format. Figure 1 is a sample of the raw data.

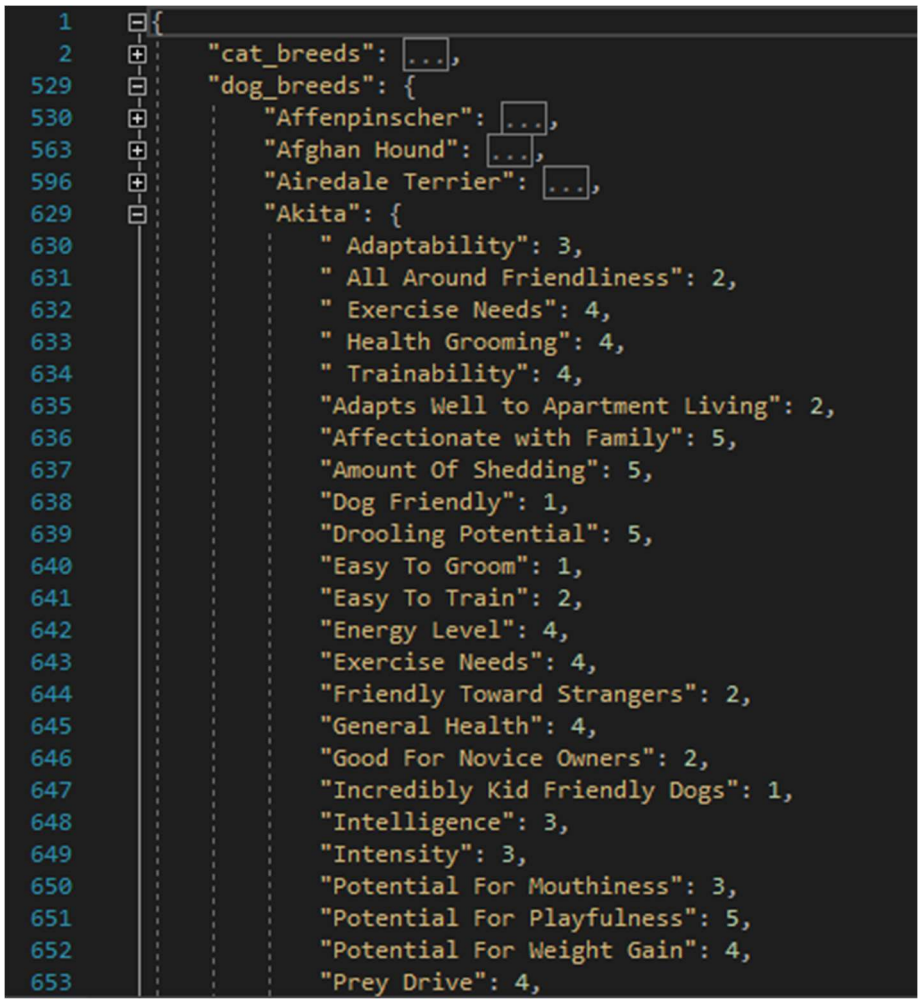

Figure 1: Raw Data Sample

The data contained an entire cat breed section along with the dog breed section, but as we chose to focus on dog breeds that section was manually removed. There were also a few dog breeds listed without traits that were manually removed. Finally, an additional category was manually added to each dog's trait called "Shelter Breed". This contained the recommended shelter breed for each of the dogs to be used as the label for the training data for our k nearest neighbor algorithm. The recommended breed was based on its Euclidean distance (found manually) from each dog breed. The top five shelter breeds were determined from an article called "15 Most Common Dog Breeds Found In Shelters" at https://iheartdogs.com/15-most-common-dog-breeds-found-in-shelters/. We chose to use only the top five as fifteen was too taxing on our machine learning algorithm and negatively impacted the computation time.

This data was then processed into a data frame using the code in figure 2. Which left us with the processed data in Figure 3.

```
with open("rating_processed.json") as f:
    data = json.load(f)
```

```
import pandas as pd
df = pd.DataFrame(columns=["Breed", "Adaptability", "All Around Friendliness", "Exercise Needs", "Health Grooming", "Trainability
```

```
row = 0
for i in data['dog_breeds']:
    df.loc[row, "Breed"] = i
    for j in data['dog_breeds'][i]:
        df.loc[row, j] = data['dog_breeds'][i][j]
    row = row + 1
```

Figure 2: Data processed into the data frame

	Breed	Adaptability	All Around Friendliness	Exercise Needs	Health Grooming	Trainability	Adapts Well to Apartment Living	Affectionate with Family	Amount Of Shedding	Dog Friendly	...	Potential For Weight Gain	Prey Drive	Sensitivity Level	Siz
0	Affenpinscher	3	3	4	2	3	5	5	1	4	...	3	3	3	
1	Afghan Hound	4	4	4	2	3	5	5	4	4	...	1	5	5	
2	Airedale Terrier	2	4	5	3	5	1	4	2	4	...	4	5	3	
3	Akita	3	2	4	4	4	2	5	5	1	...	4	4	5	
4	Alaskan Klee Kai	3	3	4	3	4	3	4	4	2	...	2	5	4	
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
209	West Highland White Terrier	4	4	4	3	4	4	5	4	5	...	4	4	4	
210	Whippet	3	5	4	3	4	5	5	2	4	...	1	5	5	
211	Wirehaired Pointing Griffon	3	5	4	3	4	1	5	1	4	...	3	4	4	
212	Yorkipoo	4	4	4	2	4	5	5	1	3	...	2	3	4	

Figure 3: Processed Data Sample

Data Product Code

The predictive function of the data product was used to create the shelter dog recommendation system. The machine learning algorithm we chose to use was the k nearest neighbor (KNN) algorithm. The code we utilized was from the python libraries panda and sklearn. We first selected the data that was applicable to train the algorithm, split it into training and test data and ran it into the KNN algorithm using the code in figure 4.

```
X = df.iloc[:, 1:-1].values
y = df.iloc[:, 31].values
```

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.20)
```

```
from sklearn.neighbors import KNeighborsClassifier
classifier = KNeighborsClassifier(n_neighbors=13)
classifier.fit(X_train, y_train)
```

```
KNeighborsClassifier(n_neighbors=13)
```

```
y_pred = classifier.predict(X_test)
```

Figure 4: Machine Learning Code

The value of 5 was originally chose arbitrarily as the k value. This was then adjusted through trial and error using a confusion matrix and classification report to determine precision and accuracy of our machine learning code. An example of this can be found in figure 5.

```
from sklearn.metrics import classification_report, confusion_matrix
print(confusion_matrix(y_test, y_pred))
print(classification_report(y_test, y_pred, zero_division=1))
```

```
[[ 7  5  0  0  0]
 [ 1 13  1  0  0]
 [ 0  2  3  0  0]
 [ 0  1  0  2  0]
 [ 6  0  0  0  2]]
                          precision    recall  f1-score   support

American Pit Bull Terrier       0.50      0.58      0.54        12
                   Boxer       0.62      0.87      0.72        15
               Chihuahua       0.75      0.60      0.67         5
      German Shepherd Dog       1.00      0.67      0.80         3
       Labrador Retriever       1.00      0.25      0.40         8

                 accuracy                           0.63        43
                macro avg       0.77      0.59      0.63        43
             weighted avg       0.70      0.63      0.61        43
```

Figure 5: Confusion Matrix and Classification Report Example

We were then able to use the machine learning program that we created. First, we took an input from the user which could be any breed of their choosing. We then used the string that was entered to search for the chose breed from the data, grabbed the trait data and fed this data into our machine learning program. Finally, we just had to display the recommendation made by our program. An example of this can be found in figure 6.

```
user_dog = input("Enter your preffered dog breed: ")
```

```
Enter your preffered dog breed: Afghan Hound
```

```
#Shelter dog prediction
user_dog_search = df.loc[df['Breed'] == user_dog]
X_user_dog = user_dog_search.iloc[:, 1:-1].values
user_dog_pred = classifier.predict(X_user_dog)
print(user_dog_pred)
```

```
['Boxer']
```

Figure 6: Recommendation Example

We also included three visualization tools including a double bar graph and two pie charts all using the python library matplotlib. The bar graph simply takes the trait data from the user's inputted breed and puts it up against the trait data from the recommended breed for comparison. An example of the bar graph can be seen in figure 7.

```
#Selected dog vs. Shelter dog bar chart comparison
pred_dog_search = df.loc[df['Breed'] == user_dog_pred[0]]
X_bar = list(df.columns.values)
Y_user_dog = X_user_dog[0]
pred_dog_values = pred_dog_search.iloc[:, 1:-1].values
Y_pred_dog = pred_dog_values[0]
X_axis = np.arange(len(X_bar[1:-1]))
f = plt.figure()
f.set_figwidth(80)
f.set_figheight(10)
plt.bar(X_axis - 0.2, Y_user_dog, 0.4, label = user_dog)
plt.bar(X_axis + 0.2, Y_pred_dog, 0.4, label = user_dog_pred[0])
plt.xticks(X_axis, X_bar[1:-1])
plt.xlabel("Traits")
plt.ylabel("Rating")
plt.title(user_dog + " vs. " + user_dog_pred[0])
plt.legend()
plt.show()
```

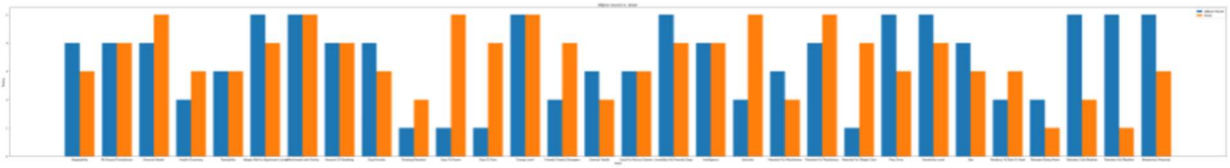

Figure 7: Double Bar Graph Example (the bar graph can be enlarged in the program)

The first of the pie charts shows those top fifteen breeds found in shelters from the article compared to the massive number of different breeds of dogs. This can be seen in figure 8.

```
#Pie chart comparing the 15 top shelter breeds to the number of other dog breeds
shelter_vs_all = ['Top Shelter Breeds', 'All Other Dog Breeds']
other_breeds = row - 14
breed_num = [15, other_breeds]
fig = plt.figure(figsize =(10, 7))
plt.pie(breed_num, labels = shelter_vs_all)
plt.show()
```

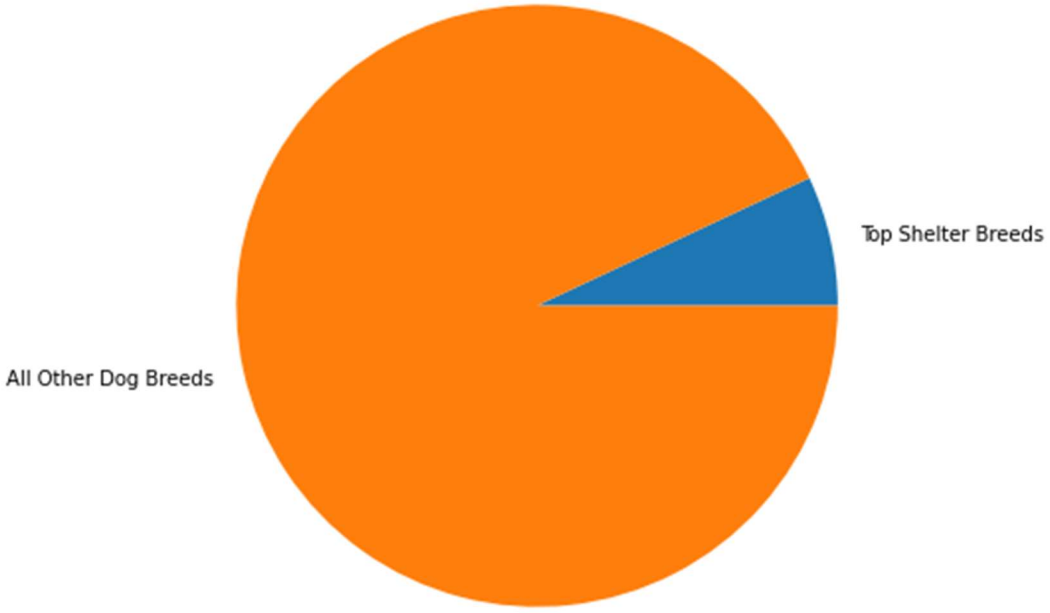

Figure 8: Number of Breeds Pie Chart and Code

The second pie chart shows the percent similarity between the chosen dog breed and the recommended dog breed. This was found by finding the absolute value of the user dog trait minus the recommended dog trait divided by the user dog trait divided by the number of traits for each of the traits, then adding those together. A sample of this is shown below in figure 9.

```
# Pie chart displaying the percent difference between the traits of the user's selected dog and the predicted shelter dog
sum_perc_diff = 0
j = 0
for i in Y_user_dog:
    float_user = float(i)
    float_pred = float(Y_pred_dog[j])
    trait_diff = abs(((float_user - float_pred)/float_user) / 30.0)
    sum_perc_diff = sum_perc_diff + trait_diff
    j = j + 1
rounded_perc = round(sum_perc_diff, 2) * 100
perc_similar = 100 - rounded_perc
percentage_label = ['', str(perc_similar) + '% similar']
percentage = [rounded_perc, perc_similar]
fig2 = plt.figure(figsize =(10, 7))
plt.pie(percentage, labels = percentage_label)
plt.show()
```

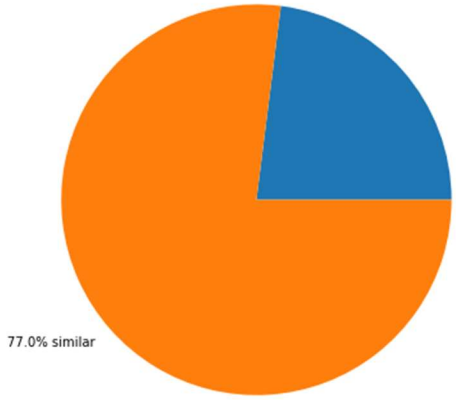

Figure 9: Percent Similar Code and Pie Chart

Hypothesis Verification

The initial hypothesis assumed that if the user likes a particular breed of dog, then they will more than likely like a similar dog. The hypothesis for the shelter dog recommendation system stated that if a user prefers a particular dog breed that they will also like a similar dog breed that is found in shelters. By recommending a similar dog breed that can easily be found in shelters more potential adopters will look to their local shelter before going to a breeder. This will reduce the overcrowding in shelters opening space for more dogs in needs and reducing the need for euthanasia. This hypothesis will need to be validated once it has been implemented in the shelters. It can be verified by monitoring the adoption rate since the system has been implemented.

Effective Visualizations and Reporting

This shelter dog recommendation system includes three data visualizations demonstrated previously. This included the double bar graph and the two pie charts. The double bar graph changes depending on the breed selected and the breed recommended as it compares the two breeds traits and allows the users to get a visual representation of how the two breeds are similar and how they differ. The first pie chart will remain the same as it just displays the top fifteen dog breeds found in shelters

and compares them to the total number of dog breeds. Finally, the second pie chart will also change depending on the selected dog breed and the recommended dog breed as it displays the percent difference or percent similar of the two dog breeds.

Accuracy Analysis

The current accuracy of the system is about 63% based on the classification report. This, however, can improve with adjustments to the k value and as the system is used over time. While working on this, I had the realization that this program could be modified in the future to allow the user to just enter a rating for how important each trait is to them on a scale of 1 to 5 and just enter that data into the algorithm. This would provide more differentiated data than the set data it is currently provided and will further improve the program.

Application Testing

Application testing was done throughout the development of the shelter dog recommendation system. Constant testing was performed as the agile methodology would suggest. The functionality test was used on the prototype of the shelter dog recommendation system. Once the shelter dog recommendation system was completed, it was implemented in a single shelter where it was usability tested and acceptability tested.

Application Files

The application files are all contained within the Final Docs folder. The program file itself must be run within Jupyter Notebook. The folder contains the following files:

- Shelter Dog Recommendation Program.ipynb: the Jupyter Notebook file
- rating_processed: the processed data file in JSON format
- rating_raw: the raw data file in JSON format
- Shelter Dog Recommendation System: The file containing all the information on the project including how to run it

User's Guide

The hosting computer must be capable of using Jupyter Notebook. The following are the steps to access and use the Shelter Dog Recommendation System:

1. Download Final Docs folder.
2. Open Jupyter Notebook.
3. Navigate to Final Docs folder and select Shelter Dog Recommendation Program.ipynb

The next series of steps is to run the program

4. Click the Run button for the first nine coding cells. (This will prepare the machine learning algorithm)
5. Click Run for the tenth coding cell to be prompted to enter a dog breed.
6. Enter your dog breed in the command line interface with each word starting with a capital and press enter.
7. Click Run for the eleventh coding cell to receive you recommended shelter breed. **(IF NOTHING COMES UP BE SURE YOU TYPED IN YOUR DOG BREED CORRECTLY.)**
8. Click Run for the twelfth coding cell for the comparison double bar graph.
9. Click Run for the thirteenth coding cell for the Shelter Breeds vs. All Other Breeds pie chart.
10. Click Run for the fourteenth coding cell for the Percent Similar pie chart.
11. Repeat steps 5-10 if you make an error or just wish to select a different breed.

Summation of Learning Experience

This capstone project pushed me outside of my comfort zone to learn new coding techniques and libraries. I enjoyed being able to work on a project that had meaning to me because I came up with it. I did realize that I rendered the machine learning portion of this pointless as I had to manually find the correct shelter breed for each dog manually and could have just used a lookup function to find that dog. However, if I were to improve upon this project, I would change it so that the user would input a rating for each of the traits and used that data instead to run through the machine learning program. I just felt that strayed a bit far from my initial proposal, so I stuck with my original idea despite being redundant. Overall, I felt this was a great learning experience and one that I can look back on when pursuing my career and working on future projects.

E. Sources

Baynazarov, Ruslan. “Cat and Dog Breeds Parameters.” Kaggle, 26 Jan. 2019, https://www.kaggle.com/hocop1/cat-and-dog-breeds-parameters.

H, Scott. “15 Most Common Dog Breeds Found in Shelters.” IHeartDogs.com, 25 Nov. 2020, https://iheartdogs.com/15-most-common-dog-breeds-found-in-shelters/.