PAWFECT MATCH

GROUP 7:
BIJOYETA KAYAL, CHANELLE GONZALEZ, CRIMSON AMARO, NOUREDDINE BELHAJ, SANJANA ISLAM, SHIH-HSIEN MA, TRISTAN MARCUM

FIND YOUR PAWFECT MATCH

Are you looking for your soul mate? Someone to be your best friend, always be by your side or greet you at the door everyday with hugs and kisses, and never nag or fight with you... then, get a dog!

Lucky for you, we can help you find a furry friend. Using various machine learning models and data sets, we have developed different methods to find the perfect dog for you.

DIGGING FOR DATA

Data Resources:

- Kaggle for image detection files with corresponding labels
- AKC web scrape for breed data and additional images

Data Cleaning:

- Basic cleaning for nulls, creating field to join AKC with Kaggle data, attaching scraped images to main AKC data frame

Data Issues:

- Fields missing values, height and weight fields were inconsistent and using difference measurement units (metric and standard)
- Sampling bias between groups in image detection files. (Herding group under-represented/ Toy group over represented)

Group	Kaggle Breeds	AKC Breeds	Percent Represented
Herding	12	30	**40%**
Hound	19	32	59%
NonSport	11	21	52%
Sport	18	32	56%
Terrier	19	31	61%
Toy	16	21	**76%**
Working	19	31	61%

PAWFECT TOOLS

Data:

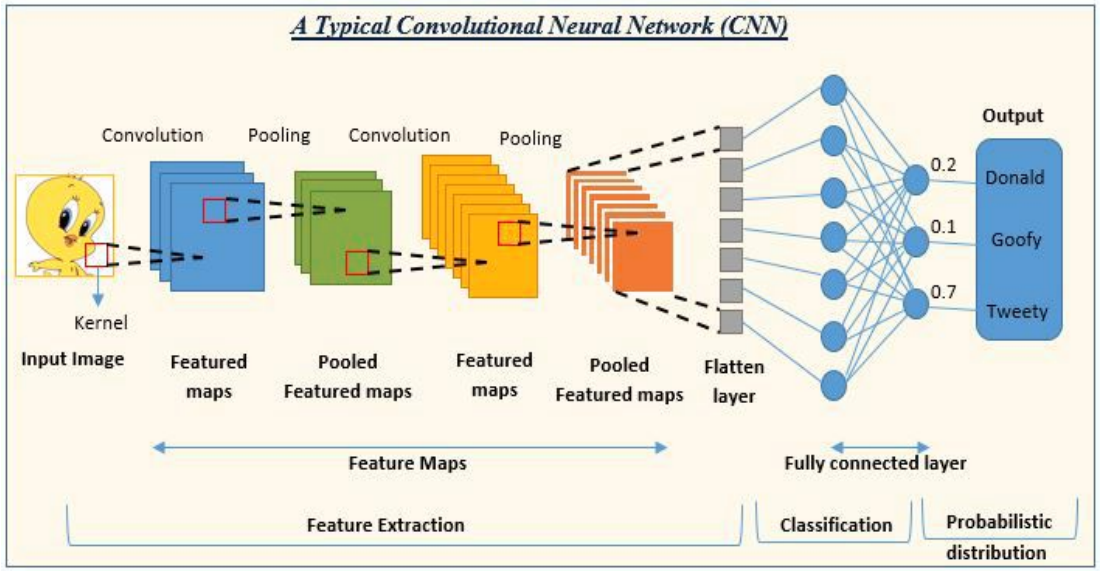
- Python/ Pandas
- SQL Lite

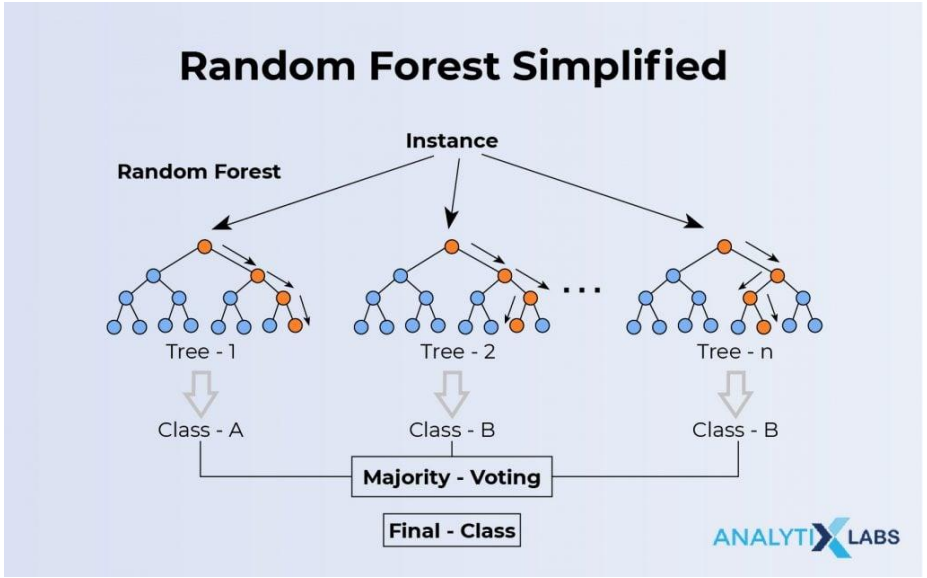
Modeling:

- Convolutional Neural Network(CNN)
- Random Forest Modeling
- Decision Tree Modeling
- Synthetic Data Generation (SMOTE, Manual)
- Standard, MinMax, Robust Scaler
- Power and Quantile Transformer
- Pickle / BoJo for Model Saving

Application:

- Streamlit
- Cloud-Based Deployment





DEMO

GO
FETCH!

PAWFECT DATA

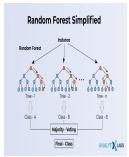
Dataset Analysis:

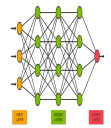
- Source is AKC data cleaned: 277 rows and 27 columns
- Selecting the right candidates for Prediction :
  - Breed Name: 277 values
  - Group: 8 values

Deciding on Feature Set:

- 13 Columns mix of numerical and categorical popularity, height, weight, expectancy, grooming frequency, shedding, energy level, trainability and demeanor

PAWFECT MODEL - BY USER



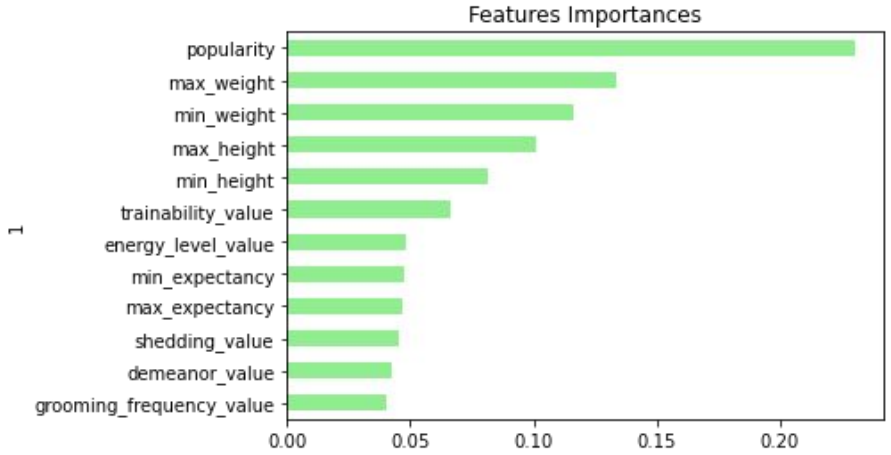



Candidates: Supervised ML for Multi Class prediction

- Random Forest Modeling
- Decision Tree Modeling
- Neural Network

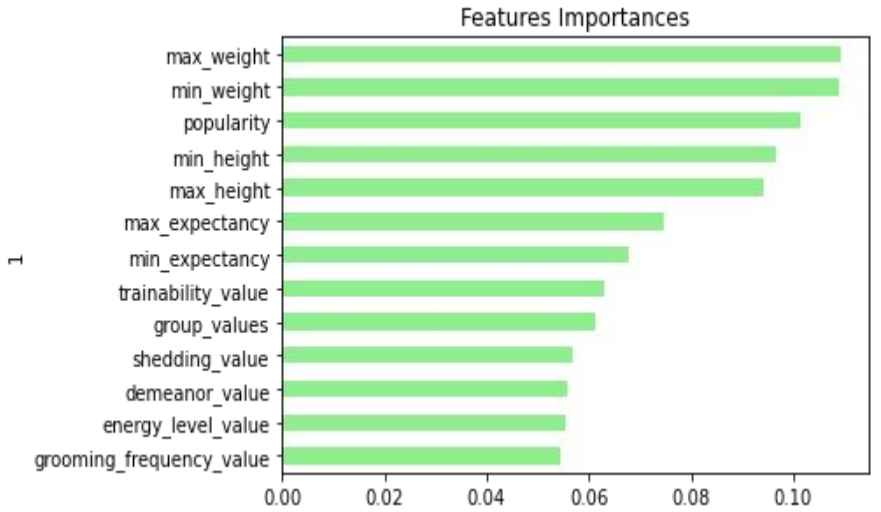
Group Prediction :

- Random Forest: accuracy= 58%
- Decision Tree: accuracy = 48%



Breed Prediction: required additional data processing: data set had one breed per line, we used SAMPLE method to generate synthetic data

- Random Forest: accuracy=100%
- Decision Tree: accuracy=99.9%



RUFF PATCHES

Data Issues:

- Dirty data
- Lack of breed images in Kaggle data set including a sampling bias between groups in image detection files.
- Having to find datasets for latitude & longitude by zip codes that coincide with animal shelter dataset

Modeling Issues:

- Transfer learning requires heavy GPU resources. Even though Google Colab had free GPU runtime options however it was only limited to training one model a day, while also losing intermittent connection since everything runs in cloud.
- 277 different breeds/ outputs created problem in Random Forest, so we need to split the data by group, then breed for modeling.

MORE PAWSIBILITIES

Data/ Modeling:

- Find additional images for all breeds to add to the image detection model to promote accuracy and remove any sampling bias.
- Add genetic and health data by breeds
- Add cost of dog and projected average ownership costs
- Feeding the data set back with user inputs

Application:

- Add additional functionality to allow for finding additional breed rescue centers and breeders by breed selected.
- Add additional functionality for creating a more detailed user profile to store your search parameters, store your favorite breeds, alert you to new adoption events near you and/ or new breeder whelps.
- Add additional functionality to allow for dog-friendly social media interaction.
- Adopt a v-PAW

BARKING UP THE RIGHT TREE

Given the time allotted for data collection and calibration, our models performed well resulting in an accurate and appropriate match. With more time and data, we can make it even better.

This application has pawesome growth potential with many ideas to expand its functionality.

We can officially say... Finding the right dog for you, just got a lot easier!

QUESTIONS?

BEGGING FOR MORE?