



PETMATCH UPDATE

Denise, Matt & Theresa

PROBLEM

- Lot of pets in shelters wait a long to get adopted
- Owners often don't know what pet is best for them, driving up return rates
- Shelters can only hold so many animals and the alternative is the streets
- Pet abandonment poses concerns in road safety, public safety, unwanted public disturbances, and declined human mental health => which can cost a lot of money



SOLUTION

- Provide users with personalized, curated rescue animal recommendations
- PetMatch will learn from users and provide on-demand user access to adoptable dogs that best suit them, all from the comfort of their favorite mobile device => faster adoptions
- Shelters will provide data about the animals themselves and contact methods for anyone interested in adoption
- Success == Increased adoption rates and lower return rates



DATA + MODEL

- Data
 - Cats
 - 46805 cats
 - 15 features (age, gender, breed info, animal attributes, animal behavior with environments)
 - Dogs
 - 97694 dogs
 - 20 features (age, gender, breed info, animal attributes, animal behavior with environments)
- Model
 - [Content-Based Filtering] Item-based similarity models (Unsupervised)
 - Linear Kernel
 - Cosine Similarity
 - Laplacian Kernel



CAT MODEL RESULTS AND FINDINGS VI

- Metric: Average Similarity score of all examples for top 5 results
 - Linear Kernel has a maximum value of 0-15, where 15 is best
 - Cosine Similarity and Laplacian Kernel have range 0-1, where 1 is best
- Findings
 - Cosine Similarity is better at differentiating one cat from another

Model Name	Score
Linear Kernel	13.7663
Cosine Similarity	0.917753
Laplacian Kernel	0.999386



CAT MODEL RESULTS AND FINDINGS V2

<todo> only run with cosine similarity, scales WAY better than linear kernel which breaks system RAM like it was tissue paper... but run with all data!

DOG MODEL RESULTS AND FINDINGS VI

- Metric: Average Similarity score of all examples for top 5 results
 - Linear Kernel has a maximum value of 0-15, where 20 is best
 - Cosine Similarity and Laplacian Kernel have range 0-1, where 1 is best
- Findings
 - Cosine Similarity is better at differentiating one dog from another

Model Name	Score
Linear Kernel	18.2681
Cosine Similarity	0.913404
Laplacian Kernel	0.999454

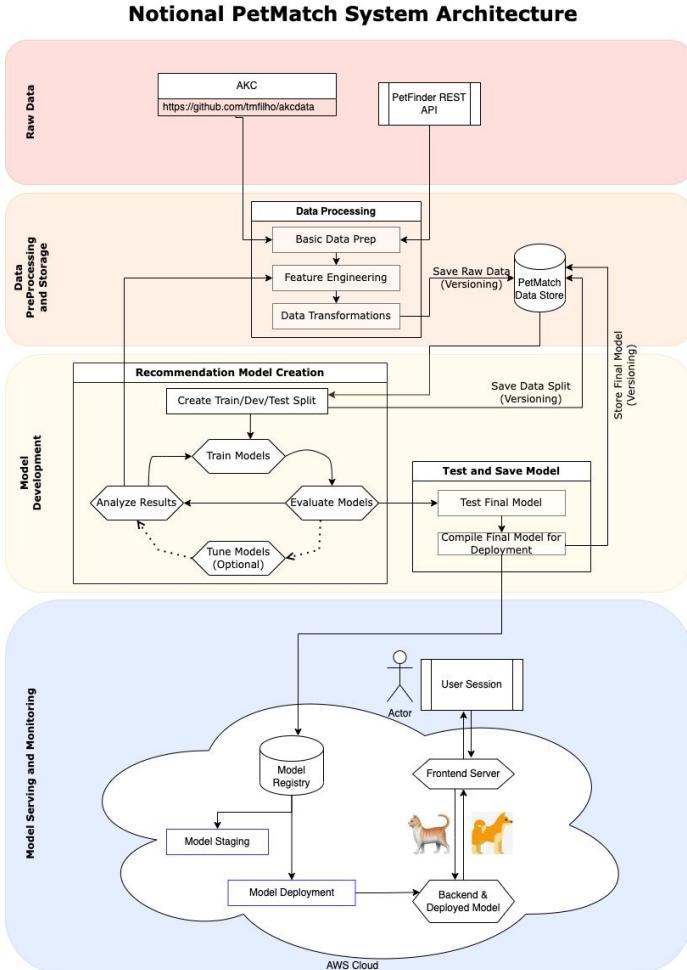


DOG MODEL RESULTS AND FINDINGS V2

<todo> only run with cosine similarity, scales WAY better than linear kernel which breaks system RAM like it was tissue paper... but run with all data! Laplacian doesn't take up much ram either (most efficient) but not discriminative enough

MLE STACK - UPDATES

- Added the AKC dataset
- Switch from a serverless backend to Flask
- Using a Streamlit app to gather user preferences



DEMO- UI

>

PetMatch Playground



Jade

[Meet Jade! Jade is a 4 month old Hound mix. Jade has oodles of puppy energy, loves to tussle with...]

See more details about this animal

Not For Me

Like



Habanaro

[These kitties were found in a pipe in a converted in rural Oklahoma and were very spicy. Her wonderful foster...]

See more details about this animal

Not For Me

Like

CAT DEMO VI - COSINE SIMILARITY EXAMPLE RESULTS



DOG DEMO VI - COSINE SIMILARITY EXAMPLE RESULTS



LESSONS LEARNED

- Modeling serverless compute locally is not the most feasible at this time
- If you have two identical recommendation sets, Laplacian Kernel model assigns a much higher score than Cosine Similarity model => Latter is more discriminative
- User rankings data requires more pre-processing than initially expected
- Content-based Filtering doesn't seem to be included in mainstream modeling libraries, required code from scratch
- Collaborative Filtering' support in modeling libraries does not seem as easy to use as traditional supervised learning algorithms



FUTURE WORK

- Execute all Item-based similarity models for dogs
 - Data Preprocessing and Feature Engineering already complete!
 - Next up is running models
- Implement Collaborative Filtering models
 - Continue collecting user rankings
 - Add timestamp to user rankings for time-sensitive recommendations
- Incorporate best model in MLE Stack
- Time permitting, incorporate distance more effectively

