Enhancing Pet Adoption Rate Predictions through Multi-Modal Machine Learning

Dongming Shen (dshen999@mit.edu)
Tessie Xie (cc906@mit.edu)

15.095 Machine Learning under a Modern Optimization Lense

Prof. Dimitris Bertsimas

Fall 2023

1 Introduction

This project delved deep into the PetFinder.my Adoption Prediction Dataset from Kaggle to tackle a poignant challenge: predicting the pet adoption speed as an ordinal classification label (main goal) and potentially accelerating the adoption process for future pets adoptions (stretch goal). The core of our analysis pivots on a key objective—to explore how multi-modality in data inputs, encompassing tabular, textual, and image features, impacts the performance of adoption speed predictions. This multifaceted approach is predicated on the hypothesis that a richer, more holistic dataset can unveil nuanced patterns and correlations that single-modality data might overlook.

Beyond predictive performance, we extends the objectives to interpretation. Through this lens, we aim to extract meaningful insights that could guide rescuers and pet owners in optimizing pet profiles. A nuanced understanding of the data's interpretative landscape will inform recommendations for profile enhancements, focusing on aspects that resonate most with potential adopters.

Embracing this dual focus on multi-modal prediction and interpretative insights, we aspire to forge a pathway that not only elevates the predictiveness of our model but also translates into practical strategies for elevating the visibility and appeal of pets in need of a home. Ultimately, our vision is to leverage machine learning not just as a predictive tool but as a catalyst for compassionate action that enriches animal welfare and adoption efficacy.

2 Data

2.1 Data Overview

The dataset contains various types of data:

- Tabular data on pet features (around 15,000 rows in `train.csv`). Key data fields include: pet characteristics (age, breed, color, etc.), health status, number of pets in the profile, adoption fee.
- Image data and the associated Google Vision API analysis metadata.
- Text data on pet descriptions with sentiment metadata from Google Natural Language API.
- Target variable, *AdoptionSpeed*, a 5-level categorical variable ranging from 0 to 4.

2.2 Exploratory Data Analysis

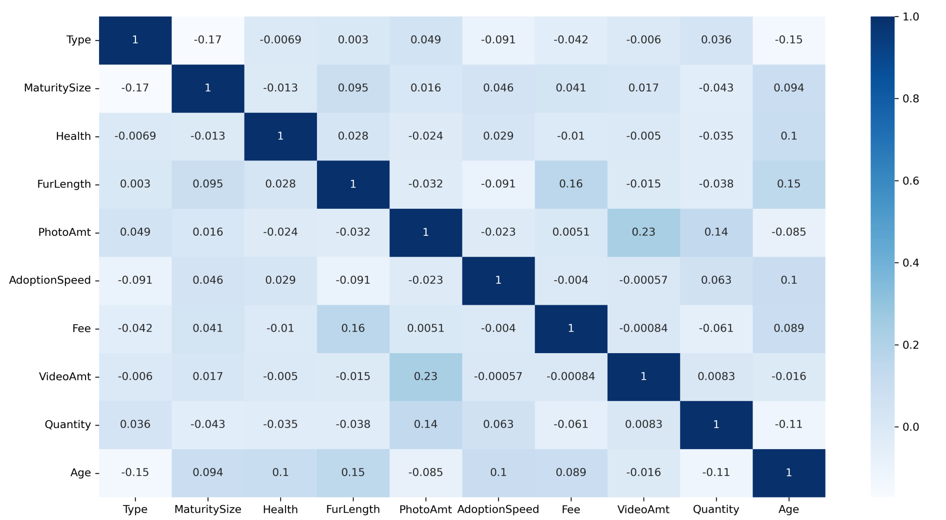

Figure 1: Correlation Matrix of Numerical Variables in Dataset

As part of our data analysis process, we conducted an Exploratory Data Analysis (EDA) on our tabular dataset. We first created a correlation matrix (Figure 1) for the numerical features to understand how these variables relate to one another. We found no strong correlations between any numerical feature pairs, saving our efforts in eliminating strong pairwise collinearity. This initial analysis provided us a starting point for making informed decisions about data handling and modeling in the next stages.

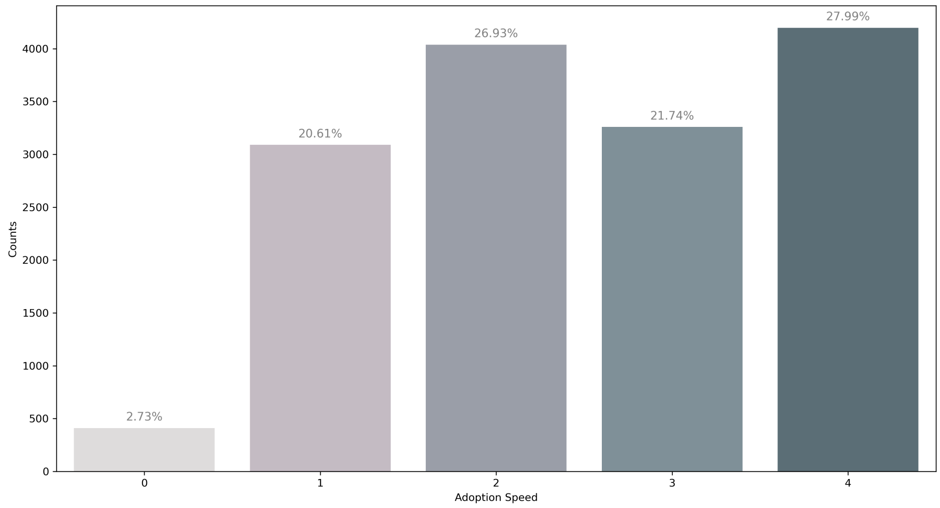

Figure 2: *AdoptionSpeed* classes rates

In addition, Figure 2 presents a bar chart that illustrates the distribution of our categorical label *AdoptionSpeed*. The chart reveals an imbalance across the different classes. To address this and ensure our training and testing datasets are representative of the overall dataset, we implement stratified sampling as our method for dividing the data. This technique helps maintain the proportion of each class in both the train and test sets.

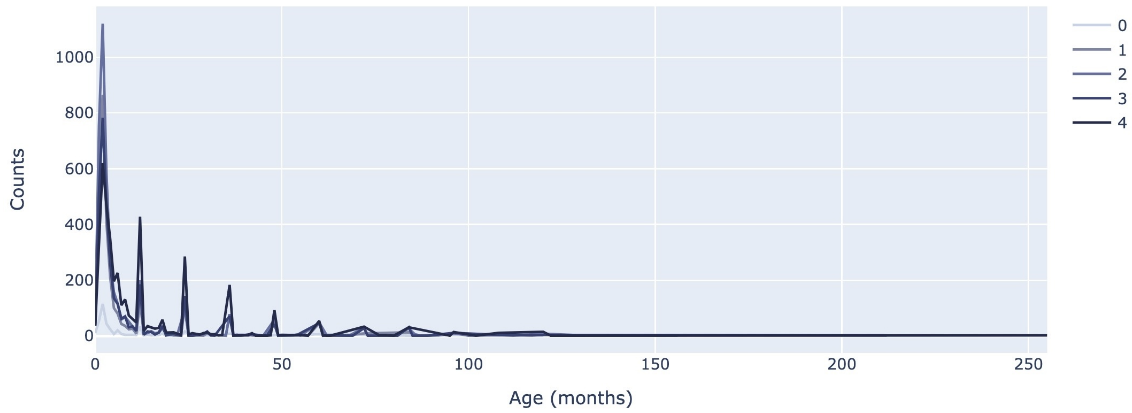

Figure 3: *AdoptionSpeed* trends by *Age*

We also analyzed the relationship between the age of the pets, measured in months, and their *AdoptionSpeed*. The corresponding relationships are showed in Figure 3. Observations from the data indicate that pets with the fastest adoption rate (*AdoptionSpeed* $= 0$) comprise the smallest group. Therefore, for a more robust analysis, we focused on pets within the *AdoptionSpeed* categories of 1 to 4.

The trend in pet *AdoptionSpeed* relative to age presents a clear pattern: pets in the younger bracket, especially those between 0 to 10 months, tend to be adopted more rapidly. During this age window, a significant proportion of pets are classified within the *AdoptionSpeed* $= 1$ category, followed by categories

2, 3, and 4 in descending order of frequency. This is visually represented by the gradient color intensity of the initial peak in our analysis. As age advances past the 10-month mark, there is a discernible transition to slower adoption categories, revealing an obvious preference for younger pets. This highlights the importance of age as a factor in adoptive decisions.

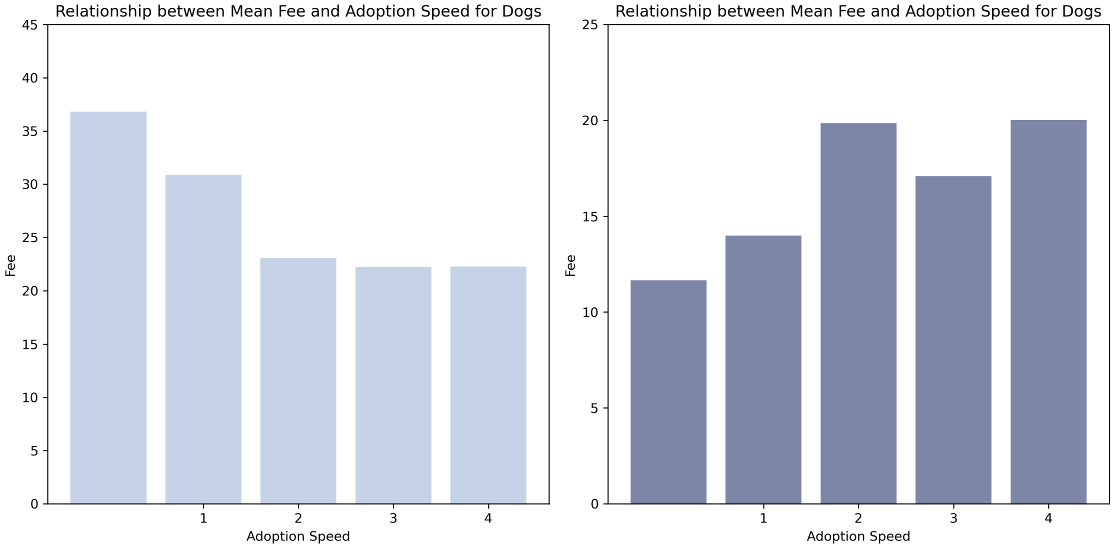

Figure 4: Relationship between Mean *Fee* and *AdoptionSpeed* for Dogs and Cat

Following that, we analyzed the potential effect of adoption fees on how quickly pets are adopted. In Figure 4, we examined a chart that lays out the average adoption fee for each category of *AdoptionSpeed*. The chart reveals variations in the average fees among the different categories, indicating that pets with different rates of adoption are associated with different fee levels. These variations lead us to believe that the amount of the adoption fee could play a role in influencing how quickly a pet is adopted, with certain fee levels possibly acting as a deterrent or an incentive for potential adopters.

3 Methodology

3.1 Feature Engineering

3.1.1 Tabular Data Processing

The dataset consists of various attributes related to pets, including their type, age, breed, color, and health status. Our primary goal is to predict the *AdoptionSpeed* of each pet, which serves as our label. In the process of preparing our tabular data for analysis, we made several key modifications:

- **Dropped Columns:** We excluded less relevant columns such as *Name*, *Description*, *PetID*, and *RescuerID*. These columns were deemed to have a limited impact on the *AdoptionSpeed* prediction.
- **Creation of New Columns:**
  - *RescuerID_cnt*: To capture the number of pets rescued by each rescuer, we transformed the *RescuerID* column into a count metric. This was achieved by grouping the data by *RescuerID* and counting the occurrences of each ID.
  - *if_name_missing*: A binary column indicating whether the pet's name is missing, which could influence adoption chances.
  - *Description_len* and *Description_word_cnt*: To quantify the extent of each pet's description, we utilized the Natural Language Toolkit (NLTK) library to calculate the length and word count of the *Description* field.

3.1.2 Sentiment and Readability

1. **The Provided Sentiment Data:** The Kaggle dataset included extra sentiment features processed through Google's Natural Language API. This supplementary information, providing insights into the sentiment and key entities of the descriptions, was incorporated.

2. **Additional Sentiment and Readability Features:** We expanded our analysis by generating additional sentiment and readability metrics.

   - Employing a BERT-based model (`nlptown/bert-base-multilingual-uncased-sentiment`) for sentiment analysis. Specifically, for each pet description, the model provided sentiment probabilities, which we stored across five designated columns (*sentiment_1* to *sentiment_5*).

   - We calculated three readability indices: *Flesch Reading Ease*, *Gunning Fog Index*, and *Automated Readability Index* using the `textstat` library. These indices provide a quantitative measure of the complexity and readability of the text.

3.1.3 Image Metadata

The Kaggle dataset also includes image metadata derived from Google's Vision API. This data offers various annotations for each pet's images, such as Face Annotation, Label Annotation, Text Annotation, and Image Properties. The file format for this data is `PetID-ImageNumber.json`. Notably, some properties like Face Annotation are absent if not applicable to the image. We extracted numerical features from this file as the image metadata.

3.1.4 Text Embedding - BERT

BERT (Bidirectional Encoder Representations from Transformers) is a groundbreaking model in natural language processing (NLP). Developed by Google, BERT revolutionizes the understanding of context in language by processing words in relation to all the other words in a sentence, rather than in isolation, using attention mechanism. This makes BERT exceptionally good at grasping meanings and complexities in language, a feature crucial for analyzing pet descriptions and breed types in our dataset. Specifically:

- Each pet's *Description* field is first tokenized using BERT's tokenizer. This process converts text into tokens that can be understood by the model.
- We used pre-trained BERT to processes these tokens to produce text embeddings. The embeddings from the last layer of BERT, which capture the highest level of understanding, are used.
- These embeddings are transformed into a 1x768 vector for each description through average pooling, effectively condensing the information while retaining crucial features.
- We applied dimensionality reduction through Principal Component Analysis (PCA), where 200 principal components were found to retain over 97% of the variance in the data.
- The same embedding process is applied to the *Breed* fields, enabling the model to capture real-world semantic meanings associated with different breeds.

3.1.5 Image Embedding - BEiT

BEiT (Bidirectional Encoder Representations from Image Transformers) represents a significant leap in image processing, analogous to BERT in NLP. Developed by Microsoft, BEiT is specially designed to understand and analyze images, identifying patterns, and features that are not immediately apparent. Its capability to process images in a contextual manner, much like BERT does with text, makes it highly effective for tasks like image classification and object detection, which are essential for analyzing the pet images in our dataset. Specifically, we apply the following steps to generate image embeddings when applicable:

- Up to four images per pet are processed through BEiT's feature extraction system. This involves analyzing the images and identifying relevant features.
- We used pre-trained BEiT to generate embeddings for these images, capturing complex visual information in a structured format. Specifically, we extracted the last layer.

- These embeddings are then averaged to create a single 1x768 vector per pet, ensuring a concise yet comprehensive representation of each pet.
- Unlike text embeddings, PCA was not utilized for image embeddings, as the principal components did not effectively encapsulate the image data. This decision was informed by the unique nature of image data, which often requires a more nuanced approach for dimensionality reduction.

3.2 Evaluation Metric

The Quadratic Weighted Kappa (QWK) metric was chosen as the evaluation metric for this project. QWK is particularly suitable for cases where the classification categories are ordinal, as is the case with our *AdoptionSpeed* labels. In particular, *AdoptionSpeed* is an ordinal variable with five classes (0 to 4), where each class signifies the time taken for a pet to be adopted. These classes inherently have an order - a pet adopted in 1 day (class 0) is distinctly different from one adopted in more than 90 days (class 4). In this case, traditional classification measures like accuracy or multi-class AUC would treat these classes as nominal, ignoring the ordered nature of the outcome. QWK, on the other hand, effectively accounts for the order and the degree of disagreement between the predictions and the actual labels, making it more suitable for this ordinal classification task.

Calculation of QWK: QWK compares the agreement between two ratings - here, the predicted and actual *AdoptionSpeed*. It ranges from 0 (random agreement) to 1 (complete agreement). The steps for calculating QWK are as follows:

1. Construct an $N \times N$ histogram matrix O, where $O_{i,j}$ is the count of instances with actual rating i and predicted rating j.
2. Calculate the $N \times N$ matrix of weights w, defined as $w_{i,j} = \frac{(i-j)^2}{(N-1)^2}$. This matrix penalizes disagreements based on the distance between ratings.
3. Compute the $N \times N$ histogram matrix of expected rating E under the assumption of independence between actual and predicted ratings. This is calculated as the outer product between the actual rating's histogram vector of ratings and the predicted rating's histogram vector of ratings, normalized such that E and O have the same sum.
4. Finally, from these three matrices, QWK is calculated as $\kappa = 1 - \frac{\sum_{i,j} w_{i,j} O_{i,j}}{\sum_{i,j} w_{i,j} E_{i,j}}$.

By accounting for the ordered nature of the labels and penalizing predictions based on how far they are from the actual class, QWK provides a robust and meaningful measure of performance for our models.

3.3 Modeling

3.3.1 General Model Pipeline

The modeling approach began with a data split of 0.8 train to 0.2 test, utilizing stratified sampling based on the label *AdoptionSpeed* to ensure a balanced distribution across classes. The fundamental structure of the model pipeline involved a regression model paired with a threshold optimizer. The optimizer's role was to convert numerical predictions into categorical values representing *AdoptionSpeed*, categorized as follows: (fastest) 0 for same-day adoption, 1 for adoption within the first week, and so on, up to (slowest) 4 for no adoption after 100 days.

A rounding threshold optimizer, termed 'OptimizedRounder', is a custom-built tool designed to maximize the QWK score. It operates by adjusting the thresholds that separate the classes in the regression model's continuous output. These thresholds are optimized such that the predicted categorical outcomes align as closely as possible with the actual categories, in terms of the QWK score. This optimization is crucial as it ensures the predictions respect the ordinal nature of the *AdoptionSpeed* classes.

3.3.2 Models Used and Multi-Modality

Regression Model Used:

A variety of regression models were employed within this pipeline, they are included in Table 1. The latter two models utilized categorical encoding for categorical variables due to their built-in support, whereas the others used one-hot encoding.

Table 1: Regression Models Used in the Pipeline

Model Type	Description
Linear Regression	Standard linear approach to regression.
Ridge Regression	Linear regression with L2 regularization.
Optimal Regression Tree	Optimal Regression Tree (ORT) in IAI.
Random Forest Regressor	Ensemble method using decision trees.
Gradient Boosting Regressor	Boosting method for regression.
LGBM Regressor	Light Gradient Boosting Machine.
CatBoost Regressor	Categorical Boosting Regressor.
LGBM Regressor (categorical)	LGBM with categorical encoding.
CatBoost Regressor (categorical)	CatBoost with categorical encoding.

Multi-Modality Analysis:

We explored the impact of multi-modality on model performance. Each single modality components are included in Table 2. Various combinations of these modalities were tested, each representing a different set of features derived from the data.

Table 2: Description of Individual Modality Components

Modality Code	Description
_	Pure tabular features derived from the dataset.
s	Sentiment features provided in the dataset.
m	Metadata features derived from image analysis.
t	Text data features created from sentiment and readability analysis.
T200	Text embeddings from descriptions using BERT, with 200 PCA components.
b200	Text embeddings from breed type using BERT, with 200 PCA components.
G	Image embeddings obtained using the BEiT model.

Finally, the best model and modality combination was chosen for training on the complete dataset, followed by submission to Kaggle for evaluation against the actual hidden test set.

For better understanding and for a clearer visualization of our complete multi-modal framework from feature engineering to final prediction, we summarized the whole pipeline in Figure 5.

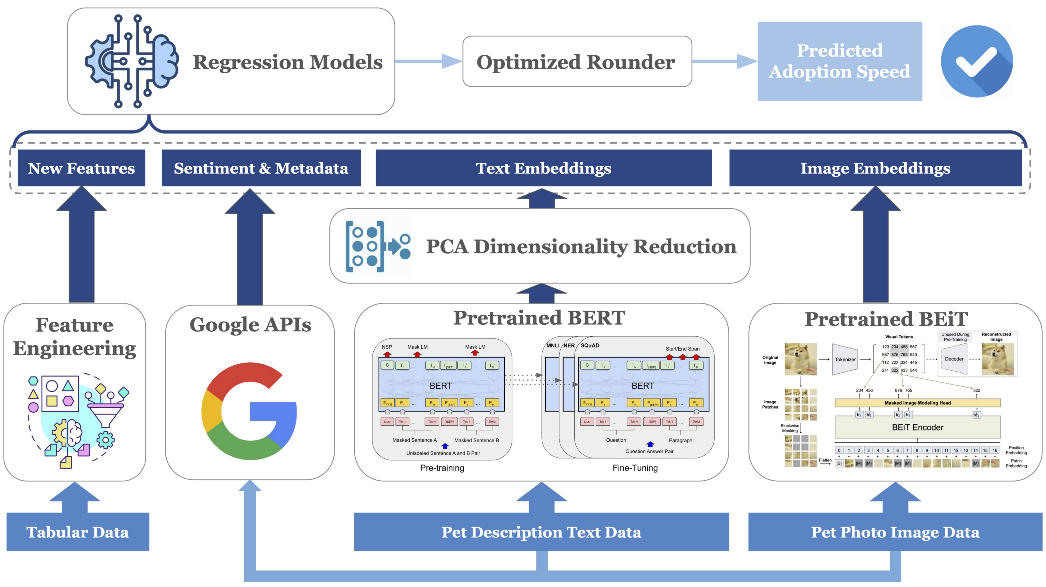

Figure 5: The Complete Multi-Modal Pipeline

4 Results and Insights

4.1 Comparing Modalities and Models

In our evaluation of models and modalities, we represented the QWK performance across different (model used, modality group) combination in a heatmap format, as depicted in Figure 6. This visualization is critical in discerning the performance of various combinations of modalities and models used in the study.

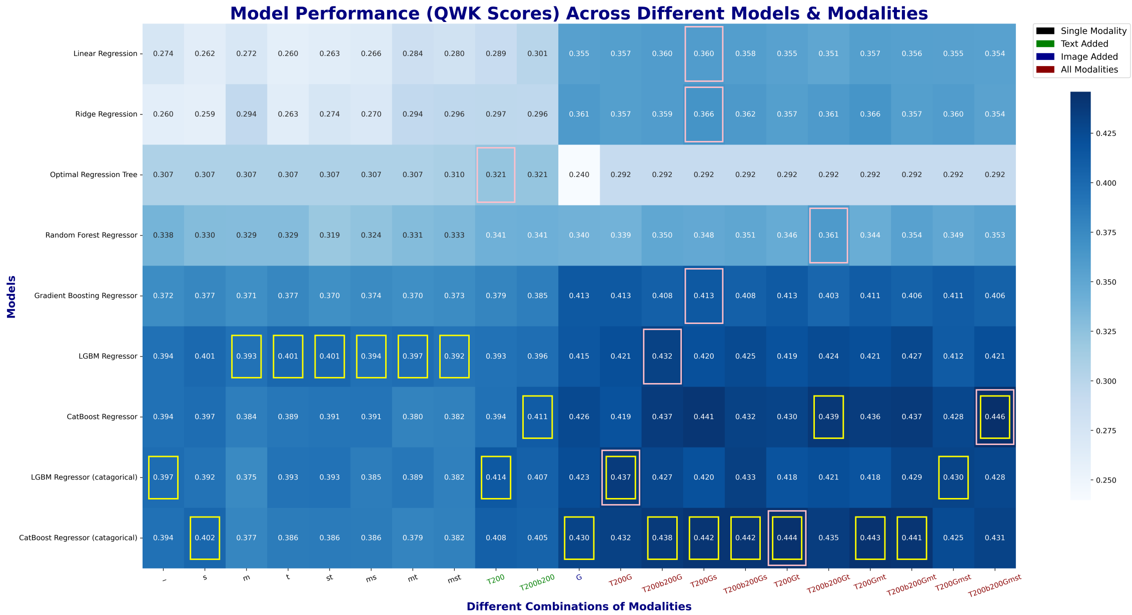

Figure 6: Comparison of Performance

Interpreting the Heatmap:

- Each cell within the grid represents the QWK score, with deeper shades of blue indicating higher scores, thus better model performance.
- The rows of the heatmap correspond to the regression models that were evaluated.
- The columns represent the different combinations of modalities tested, with color coding to distinguish between the types of modalities incorporated.
  - Black column headers indicate single modalities, primarily tabular features and metadata.
  - Green column headers signal the inclusion of text embeddings derived from BERT.
  - Blue column headers denote the inclusion of image embeddings from BEiT.
  - Red column headers reflect the incorporation of all modalities: tabular, text, and image.
- Smaller yellow boxes highlight the best-performing model for each modality combination.
- Larger pink boxes indicates the best-performing modality combination for each model trained.

Key Insights:

1. The best models, as showed by the last four rows highlighted with yellow boxes, include the advanced ensemble and boosting methods, specifically CatBoost and LGBM, both in their standard and categorical forms. The categorical versions are slightly better and faster to train and tune.
2. Notably, all of the best-performing modalities for each model, encased in pink boxes, correspond to the columns with red names. This suggests that including all modalities—tabular data, text embeddings, and image embeddings—consistently yields the most robust results.
3. The Optimal Regression Tree (ORT) model emerged as an outlier of the above point, as it was the only model negatively impacted by the addition of image modality data. This deviation is likely

due to the high dimensional feature space, coupled with constraints in executing comprehensive hyperparameter tuning necessary to harness the full potential of this complicated model.

4. The gradation from lighter to darker blue across the heatmap suggests a general trend: models benefit from the richness of information provided by multi-modal data. The more comprehensive the data modality, the better the models perform, particularly when all modalities are combined, as reflected in the red columns.

This analysis underscores the significance of multi-modal machine learning in enhancing predictive performance. By leveraging diverse data representations, models can capture a more holistic view of the dataset, which is paramount in tasks that involve varied data types, such as those in pet adoption prediction.

4.2 Best Models

Hyperparameter Optimization: Following a rigorous evaluation of models and modalities, we narrowed our focus to the most promising candidates: the LGBM categorical and the CatBoost categorical models. These were paired with the richest combination of modalities (red columns). To refine these top-performing model-modality pairings, we employed Optuna for hyperparameter tuning, executing 500 iterations within a 3-fold Cross Validation framework. This process aimed at fine-tuning the models to achieve the best possible performance. We get the following result in Figure 7.

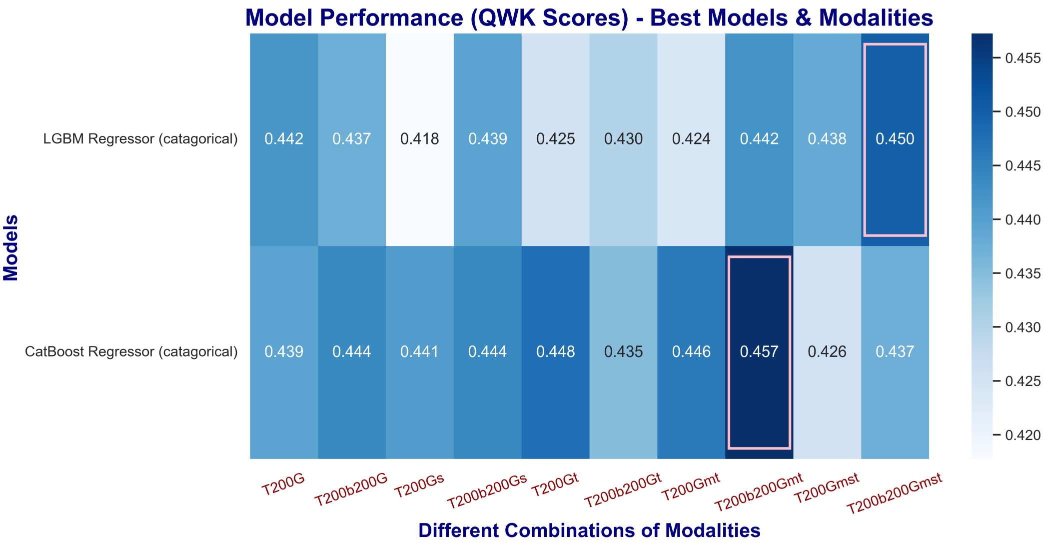

Figure 7: Comparison of Performance - Best Models and Best Modalities

Final Evaluation on Kaggle: The hyperparameter optimization process led us to identify the best hyperparameters for both the LGBM categorical and CatBoost categorical models, for different modalities. Upon training these models with the entire dataset using their respective best modality combinations—namely, 'T200b200Gmst' for LGBM categorical and 'T200b200Gmt' for CatBoost categorical—we proceeded to evaluate them on Kaggle's unseen test set. The results were highly competitive: the LGBM model achieved a QWK score of 0.4399, while the CatBoost model scored 0.4414. These performances would place both models at the 31st position among approximately 1800 participating teams. These outcomes highlight the robustness of the chosen models and modalities and demonstrate the effectiveness of our multi-modal framework.

4.3 Interpretations

4.3.1 SHAP Feature Importance

Analyzing feature importance is crucial for understanding the factors that influence the predictive models. We employ SHAP (SHapley Additive exPlanations) values to interpret the impact of the features on the our best CatBoost model's output. The following are interpretations based on the SHAP values plot for the top 10 features (in ranked order from most important features at the top), as shwon in Figure 8:

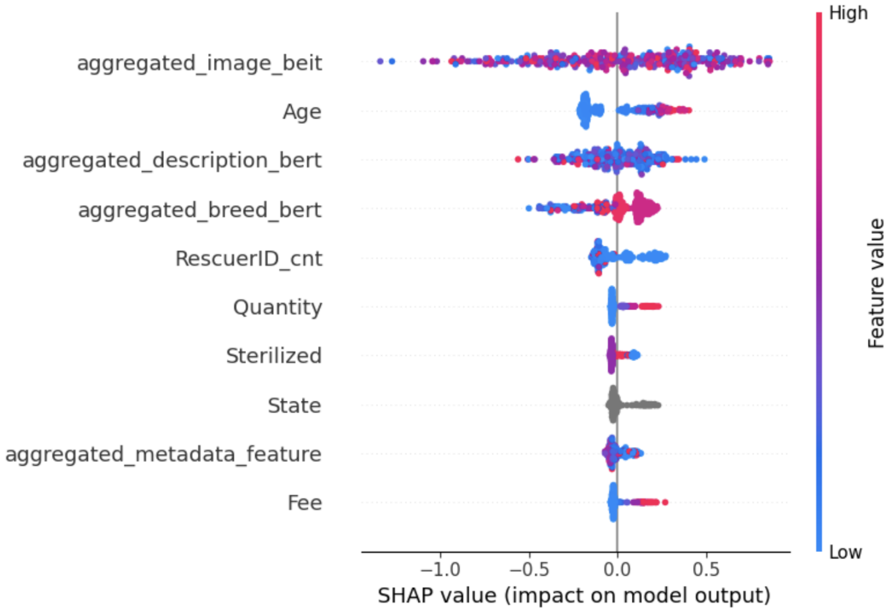

Figure 8: SHAP Importance of The Best CatBoost Model

1. **Aggregated Image BEiT Embedding:** Due to aggregation over 768 features, direct interpretation is challenging. However, the spread of SHAP values and the top position suggested a diverse impact on adoption speeds, indicating the important role of visual features.
2. **Age:** The SHAP values demonstrate a clear trend where higher age values, indicative of older pets, corresponds to longer adoption speeds. This aligns with intuitive expectations that younger pets are often adopted more quickly.
3. **BERT Embeddings:** The aggregated description and breed embeddings from BERT (aggregated_description_bert and aggregated_breed_bert) also present interpretation challenges due to the deep learning based embedding in addition to dimensionality reduction and aggregation. Nonetheless, their position indicates a significant impact on the model output.
4. **RescuerID_cnt:** The plot reveals some concentration of pink dots for higher counts have negative SHAP values, which may imply that pets rescued by more experienced rescuers, are likely to be adopted faster, which logically makes sense.
5. **Quantity:** Intuitively, a higher quantity of pets tends to lead to a slower adoption process, as all pets in a group must be adopted for a successful outcome. The SHAP values proves this, with higher quantities generally leading to negative impacts on adoption speeds.
6. **Sterilized, State, and Aggregated Metadata Feature:** These features show a mixed impact on adoption speeds, with no clear trend discernible from the SHAP plot. They appear to have a varied influence depending on individual instances.
7. **Fee:** The SHAP values for the fee are skewed towards the negative impact on the model's output, suggesting that higher fees are associated with slower adoption speeds. This is in line with the expectation that cost can be a barrier to adoption.

These insights from the SHAP plot provide a good understanding of how different features influence the likelihood of pet adoption. They highlight the complexity of the adoption process and the importance of considering a multifaceted approach when designing interventions to improve adoption rates.

4.3.2 Interpretation of Images and Descriptions (Tables in Appendix)

From the analysis of variable importance and SHAP values, we observed that image and text data significantly impact the predictions of our model in determining pet adoption speeds. Thus, a more comparative study of visual characteristics was conducted, analyzing four images from each category of

AdoptionSpeed 0 (fastest) and 4 (slowest), with negative and positive SHAP values for image embeddings, respectively, as shown in Table 3.

In the group with an *AdoptionSpeed* of 0 with negative SHAP, the images are very attractive. Pets in these photographs are depicted in a manner that emphasizes their cuteness and charm, with the high quality of the imagery, favorable lighting, and vivid coloration. Such favorable visual presentations enhance the pets' appeal to prospective adopters, which possibly leads to their expedited adoption.

In contrast, some images corresponding to an *AdoptionSpeed* of 4 with positive SHAP (indicating that the images negatively affected the adoption speed, making it slower) showcase noticeable deficiencies. In particular, this category often features images where the pets' faces or figures are obscured or poorly visible, thus providing very limited information about their physical attributes. Moreover, some images are subject to suboptimal backgrounds with bad lighting, potentially hinting less-than-ideal living and health conditions. These aspects might foster adverse impressions regarding the pets' well-being or hygiene, thereby negatively affected their rate of adoption. Recognizing these visual factors is indispensable for a comprehensive understanding of the variances in adoption speeds as influenced by photographic representations.

In conjunction with the analysis of image data, textual descriptions of the pets were also compared. We created Table 4, contrasting descriptions of four pets each from the fastest (*AdoptionSpeed* 0) and slowest (*AdoptionSpeed* 4) adoption groups, again with negative and positive SHAP values for descriptions, respectively. The content and quality of the descriptions between these groups exhibited marked disparities. Descriptions associated with rapidly adopted pets are characteristically thorough and elaborate. They often encompass good amount of details, including the pet's physical attributes, health status, behavioral tendencies, and even narratives of their interactions with past owners or rescuers. This depth of detail tends to engage potential adopters more effectively, thereby lead to a higher interest in adoption.

Conversely, descriptions pertaining to pets with languid adoption trajectories are generally terse and lack substance. Even when lengthy, as seen in the second row for the slower *AdoptionSpeed* group, the content conveyed is oftentimes not compelling, possibly with overall negative attitudes, and fails to advocate for adoption. Another notable issue is the language barrier. Many descriptions in the slower adoption group are in languages other than English or sometimes unreadable, like the instance in the fourth row, which may significantly narrow the scope of audience engagement, especially on a platform predominantly operated in English-speaking regions. These elements, in concert, are influential in the diminished adoption velocities observed for these pets.

4.4 Insights and Potential Impact

Our pet adoption speed prediction model is designed to estimate the time range it would potentially take for a pet to be adopted, using the pet profile data provided by the owner or rescuer. For example, a predicted *AdoptionSpeed* $= 0$ indicates an estimated adoption time of 1 day while a predicted *AdoptionSpeed* $= 4$ indicates an estimated adoption time of more than 90 days. With such information, the online pet adoption platform can offer an estimated adoption timeline, a crucial piece of information for pet owners and rescuers. Specifically, such insights are helpful in aiding their preparation and planning for the pet's care, particularly in managing food supplies and other essentials. Furthermore, equipped with an understanding of the anticipated adoption speed, owners and rescuers are empowered to strategically refine their pet profiles to enhance the appeal of their pets. The platform can also help advertise the pet based on the predicted adoption speed to accelerate the adoption process when necessary.

In addition to a simple prediction, we went one step ahead taking advantage of the aforementioned analysis on feature importance as well as the quality of images and descriptions. Specifically, we realized domestic features of pets (such as age, breed, color, health conditions, etc.) are indeed very influential to adoption speeds. Nevertheless, images and textual descriptions with suboptimal quality are likely to negatively affect this speed. As a result, we propose here an innovative pet-adoption online posting framework, aimed at increasing the likelihood of adoption and accelerating the adoption process when possible. A flowchart of this framework is shown in Figure 9. The framework is based on the important insight that good quality pictures and descriptions in pet profiles can greatly affect how likely and how fast it is for the pets to be adopted.

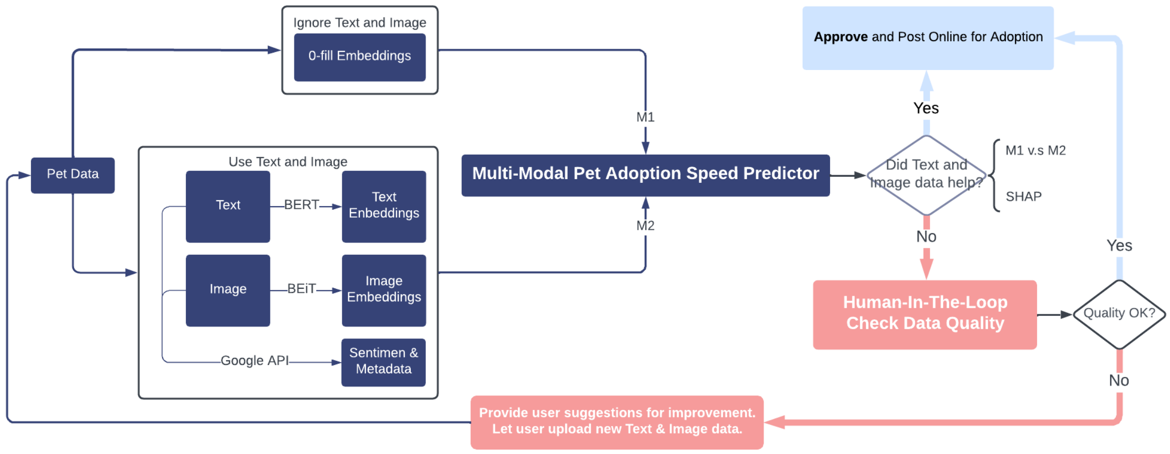

Figure 9: The Proposed Pet-Adoption Online Posting Framework

The framework operates as follows:

1. Upon receiving pet data, two feature sets are constructed: the first set ignores text and image data, utilizing zero-filling for embeddings (assuming a baseline where these data are absent), while the second set fully incorporates textual descriptions and images.

2. These feature sets are then fed into our best multi-modal pet adoption speed prediction model to generate two separate adoption speed predictions.

3. We compare the two predicted speeds and their corresponding SHAP charts, aiming to detect the effectiveness of the text and image data. If the inclusion of this data is found to enhance adoption speed, the listing proceeds to posting.

4. Conversely, if the data do not contribute to an increased adoption speed or, worse, detract from it, a Human-In-The-Loop (HITL) approach is invoked. Specifically, a manual review of the text and images assesses quality concerns such as clarity, composition, lighting in images, and positivity, readability, and detail in descriptions.

5. Should quality issues be identified, users are provided with constructive feedback and the opportunity to refine and resubmit their pet profile with the updated descriptions or images, thus optimizing the pet advertising before it is posted.

This framework leverages our insights into the significance of text and image quality in adoption, which, unlike other domestic features, texts and images can be easily improved if are under non-ideal conditions. By ensuring the pats are advertised in the best possible light through clear, engaging descriptions and high-quality images, the framework holds the potential to markedly improve adoption speeds and likelihoods. Furthermore, it empowers rescuers and pet owners to present their pets more effectively, fostering a proactive environment that prioritizes the well-being of the pets.

Potential Impact: The implementation of this framework can lead to a transformative change in the pet adoption landscape. By systematically improving the quality of pet listings, we can expect not only faster adoptions but also a more engaging and trustworthy platform for potential adopters. This could result in a positive feedback loop, where higher-quality listings attract more visitors, leading to increased adoption rates and, ultimately, enhancing the overall efficacy of the adoption process.

5 Conclusion

We performed a comprehensive exploration of the PetFinder.my Adoption Prediction dataset, with the primary intent to predict and potentially expedite the adoption process of pets. Central to our approach was the integration of multi-modality in data inputs, encompassing tabular, textual, and image features, to enhance predictive performance. Our analysis confirmed that such an holistic approach significantly boosts the model's ability to predict adoption speed as an ordinal classification label.

On top of multi-modality, our model pipeline effectively combined regression algorithms with a threshold optimizer to map continuous predictions to discrete adoption speed categories. This strategy proved successful, as evidenced by our model's performance, which ranked within the top 31st position among approximately 1800 teams in a Kaggle competition.

Further, we delved deep into the interpretation analysis of the most influential features using SHAP values. This allowed us to understand the impact of individual predictors and identify that high-quality images and detailed, positive textual descriptions contributed markedly to faster adoption speeds.

The insights from interpreting text and image data were instrumental in crafting a suggested pet-adoption online posting framework aimed at improving pet adoption speeds and likelihoods. This user-centric model employs the HITL approach and encourages the submission of high-quality text and image data, flagging and providing feedback on profiles that might hinder adoption due to poor quality, and thus, stands to significantly improve adoption rates.

In conclusion, our project highlights the transformative potential of multi-modal machine learning in real-world applications. By leveraging diverse data types and interpretative models, we not only enhance predictive accuracy but also contribute actionable recommendations that can have a tangible, positive impact on animal welfare and online adoption platforms.

Appendix

Table 3: Comparison of Images from *AdoptionSpeed* $= 0$ and *AdoptionSpeed* $= 4$

AdoptionSpeed $= 0$ with SHAP < 0	*AdoptionSpeed* $= 4$ with SHAP > 0
	
	
	
	

Table 4: Comparison of Description from $AdoptionSpeed = 0$ and $AdoptionSpeed = 4$

$AdoptionSpeed = 0$ with SHAP < 0	$AdoptionSpeed = 4$ with SHAP > 0
I raised her since she was a puppy. Recently, I shifted to a bigger house. One of my neighbors also owns a few dogs, and Shy is getting infected by them. Those dogs are having serious ticks. No matter what treatment I send Shy to, she keeps having ticks again and again. It's getting worse! I couldn't help her. I couldn't see her suffering this way. She was very chubby a few months ago, and now she has reduced so much weight. If anybody is interested, please contact me as soon as possible. I'm not expecting money. I just want a very responsible owner who could handle her ticks, send her to proper treatment, and also keep in touch with us. She is very loyal and lovable. She is toilet trained. Contact: Neshalyn	Please feel free to contact us: Stuart
Pls read before you call for adoption This cat is very special to us and she will only be given to the right family (after interview) who will commit to take a good care of her and agreed to give updates / photo of the cat during the first year after rehoming. REASON FOR LETTING GO: MY 2 YO BABY DAUGHTER HAVING A SEVERE ALLERGIES TO CAT DANDER Chook Chook (porridge in Cantonese) was adopted from a porridge restaurant in Pudu KL as stray cat. She is now clean and healthy. She is extremely affectionate when require attention and always there for me. She bath once every month and eat Royal Canin biscuit only. Yearly check up (vaccination and deworming) is done every year. She is very discipline, she would not scratch your beloved furniture.	this few kitten need home. the mama cat get pregnant again.. pls contact me asap if can.. from hulu langat
Neyo rescued by me n wife three weeks ago. Found near my office in a drain. We urgely sent him for full vaccinated and deworm and cleaned. He is very calm and gentle but a little bit afraid of new surroundings but improved after time. We already have two kittens in possession, both males and unfortunately one out of two feel envious with Neyo and always try to scratch Neyo. Luckily Neyo is not type of picking a fight. So we need a terrace house adopter as we believed Neyo loves large spaces and it is better if new adopter have no cat/cool and calm cat at home. Please call. Kindly no SMS. We are willing to send him to new adopter if nearby. RM70 fee is actually not for vaccinated and deworm fee but for cat foods.	LOOKING FOR A FOREVER HOME FOR HER
Cougar is a very healthy and lovable dog. He is pure white in colour and has beautiful brown eyes. He understands simple English and is well trained. Very playful and likes attention. He was born in March. I would like to put up Cougar for adoption as I will be migrating soon.	请给小猫一个温暖的家。。可以 sms 给我，谢谢

Contribution

As A Group:

- Came up with the topic and researched the dataset.
- Designed the multi-modal pipeline and proposed the pet-adoption online posting framework.
- Determined the models to be used.

Tessie Xie:

- Conducted EDA to gain insights into the dataset and assist with feature engineering.
- Selected photos and descriptions to interpret the results.
- Contributed to the report sections on Introduction, Data, Interpretation of Images and Descriptions, Insights, and Potential Impacts.
- Prepared the PowerPoint presentation.
- Enhanced the clarity and engagement of figures and graphs.

Dongming Shen:

- Conducted feature engineering.
- Utilized pre-trained BERT and BEiT to generate text and image embeddings.
- Trained the pet adoption speed prediction models; visualized the results for different model and modality combinations.
- Proposed the improved Pet-Adoption Online Posting Framework.
- Contributed to the report sections on Methodology, Results, and Impact.