PetFinder.my
Animal Adoption Prediction

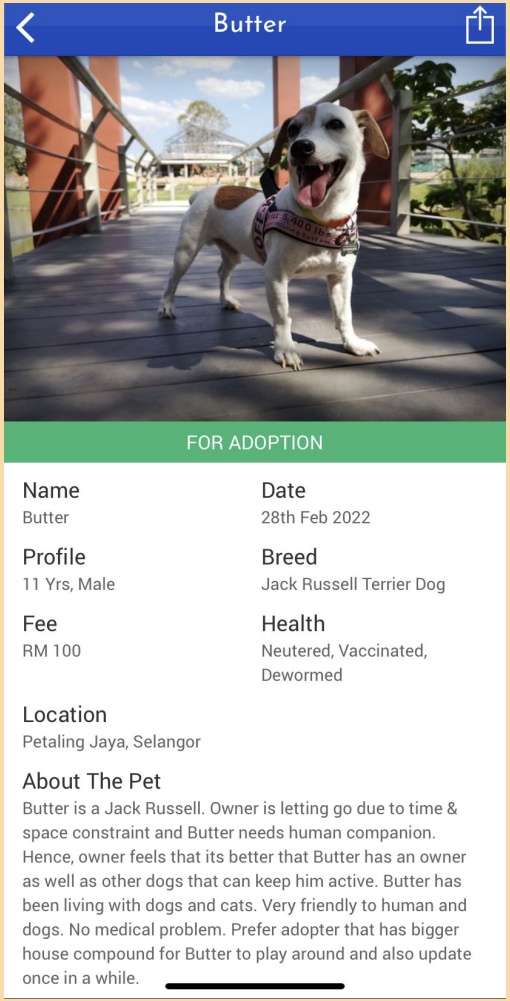
01

INTRODUCTION

Background, problem statement & rationale

Background

- Leading animal welfare platform since 2008
- Provide online profiles for animals waiting for adoption
- Aims to boost the animal adoption rate
- Database of more than ten thousand of animals, includes metadata & adoption information



Task introduction

Task Goal

- Predict the adaptability of pets
- Improving pet's profiles appeal

Problem Statement

When new animals come in, given it's metadata, what would be the estimated time for new pets to be adopted?

Rationale

- Find a home for the pets
- Improving adoption performance
- More efficient resource allocation by reducing sheltering cost

02

Data description

Three types of dataset

Data description

Target variable: AdoptionSpeed (Categorical output from 0 to 4)

1. Tabular Data

Characteristic of pets, for example, PetID, Name, Age, etc

2. Sentiment Data

Pet profile's (text) description went through Google's Natural Language API, with analysis on sentiment and key entities

3. Images

Pets that have photos went through Google's Vision API, for providing image properties

Source: https://www.kaggle.com/c/petfinder-adoption-prediction

Data description

1. Tabular Data

- **PetID** - Unique hash ID of pet profile
- **AdoptionSpeed** - Categorical speed of adoption. Lower is faster. This is the value to predict. See below section for more info.
- **Type** - Type of animal *(1 = Dog, 2 = Cat)*
- **Name** - Name of pet *(Empty if not named)*
- **Age** - Age of pet when listed, in months
- **Breed1** - Primary breed of pet *(Refer to BreedLabels dictionary)*
- **Breed2** - Secondary breed of pet, if pet is of mixed breed *(Refer to BreedLabels dictionary)*
- **Gender** - Gender of pet *(1 = Male, 2 = Female, 3 = Mixed, if profile represents group of pets)*
- **Color1** - Color 1 of pet *(Refer to ColorLabels dictionary)*
- **Color2** - Color 2 of pet *(Refer to ColorLabels dictionary)*
- **Color3** - Color 3 of pet *(Refer to ColorLabels dictionary)*
- **MaturitySize** - Size at maturity *(1 = Small, 2 = Medium, 3 = Large, 4 = Extra Large, 0 = Not Specified)*
- **FurLength** - Fur length *(1 = Short, 2 = Medium, 3 = Long, 0 = Not Specified)*
- **Vaccinated** - Pet has been vaccinated *(1 = Yes, 2 = No, 3 = Not Sure)*
- **Dewormed** - Pet has been dewormed *(1 = Yes, 2 = No, 3 = Not Sure)*
- **Sterilized** - Pet has been spayed / neutered *(1 = Yes, 2 = No, 3 = Not Sure)*
- **Health** - Health Condition *(1 = Healthy, 2 = Minor Injury, 3 = Serious Injury, 0 = Not Specified)*
- **Quantity** - Number of pets represented in profile
- **Fee** - Adoption fee *(0 = Free)*
- **State** - State location in Malaysia *(Refer to StateLabels dictionary)*
- **RescuerID** - Unique hash ID of rescuer
- **VideoAmt** - Total uploaded videos for this pet
- **PhotoAmt** - Total uploaded photos for this pet
- **Description** - Profile write-up for this pet. The primary language used is English, with some in Malay or Chinese.

Training dataset

- Some 15 000 dogs and cats
- A mix of different types of variables (categorical, ordinal, quantitative)

Testing dataset

- Some 4 000 dogs and cats

Data description

2. Sentiment Data

Variables (derived from text data)

- Score: the overall emotion of a document
- Magnitude: how much emotional content is present within the document
- Languages

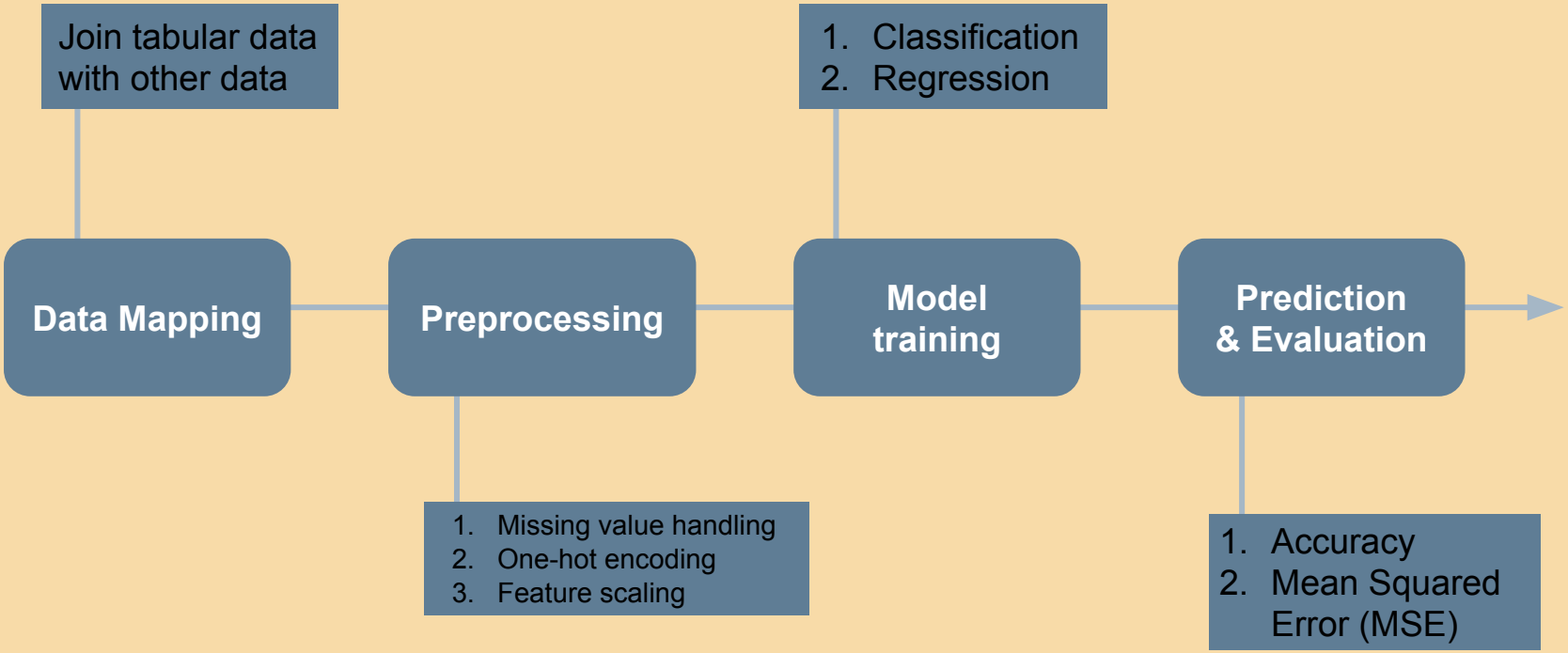
3. Images

03

Methodology

Data preparation and Machine Learning Model Training

Methodology – Overview



Data Preparation - Mapping (1)

Images
(Profile image of pets)

Text data
(Description of pets)

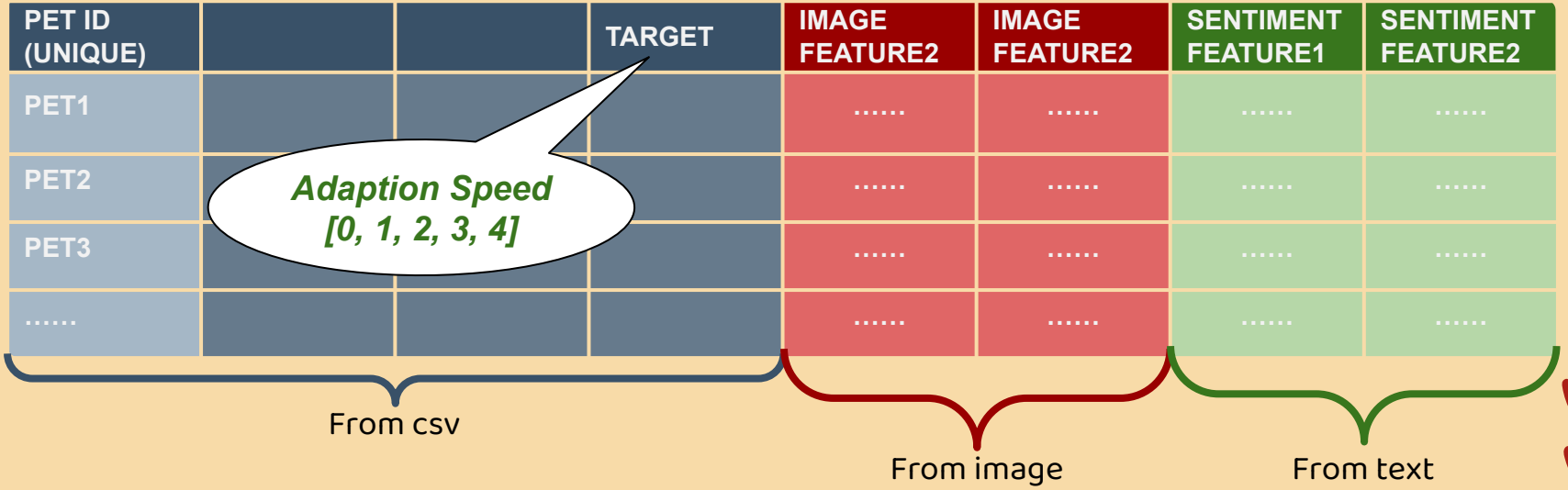
Tabular data
(Characteristic of pets)

	FEATURE1	FEATURE2	Adaption Speed
#ROW1			0
#ROW2			3
#ROW3			1
.....			4

Data Preparation - Mapping (2)

Gaol: (1) Combine different types of data into one single table

(2) Feed data to train ML models + predict the the pet adoption speed



Data Preparation - Data Mapping (3)

analysed

JSON

```
[ "pet_id"   : "PET1",
  {"label 1" : "score"},
  {"label 2": "score"},
   ……
]
```

analysed

JSON

```
[ "pet_id"   : "PET1",
  {"sentiment 1":"score"},
  {"sentiment 2":"score"},
          ……
]
```

Target table

PET ID (UNIQUE)				IMAGE LABEL1	IMAGE LABEL2	SENTIMENT 1	SENTIMENT 2
PET1				{ SCORE 0~1 }	{ SCORE 0~1 }	{ SCORE 0~1 }	{ SCORE 0~1 }
PET2				……	……	……	……
PET3				……	……	……	……
……				……	……	……	……

From csv

From image

From text

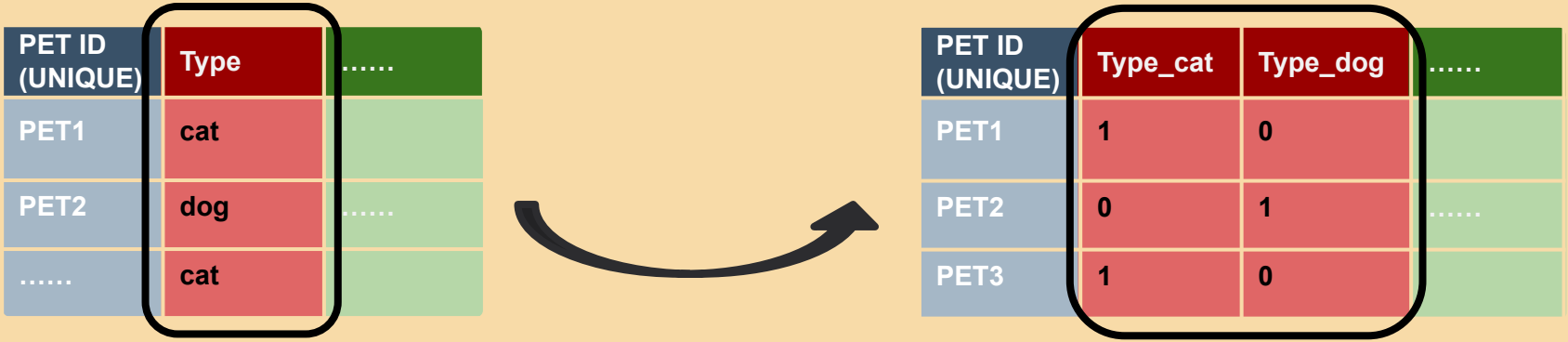
Data Preparation - Preprocessing

Goal:

(1) Missing value handling

(2) Convert categorical data (one hot encoding)

(3) Feature scaling (normalization)

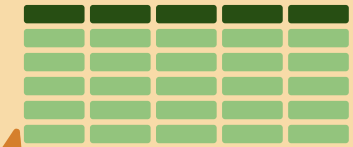


ML Model Training

Prepared Data

Multi-dimensional table

1



Train Model

Classification / Regression

(1) Random Forest
(2) XGboost
(3) Neural Network

2

Predict

Predict the adoption speed [0, 1, 2, 3, 4] by testing data

3

Evaluation

See the performance

(1) Accuracy or
(2) MSE

4

04

Q&A

Feel free to ask any question