

Animal Shelter data mining

...

Lucas May Petry
Maíke de Paula Santos

Introdução

- 7,6 milhões de cães e gatos vão para abrigos anualmente nos EUA
 - Destes, 2,7 milhões são submetidos à eutanásia
- Imensa quantidade de entradas em abrigos
 - Torna-se possível descobrir padrões
- Abrigo de Animais Austin
 - Maior abrigo de animais municipal da região central do Texas
 - Abriga aproximadamente 18 mil animais todo ano

Pré-Processamento

...

O conjunto de dados

- Provido pelo Abrigo de Animais de Austin
 - Dados de 2013 a 2016
- Conjunto de dados possui os seguintes atributos:
 - **ID:** Identificador único de cada animal;
 - **Nome:** Nome do animal;
 - **DateTime:** Data em que foi realizado o registro;
 - **OutcomeType e OutcomeSubType:** Identifica o que aconteceu com o animal no abrigo;
 - **AnymalType:** Tipo do animal(gato ou cachorro);
 - **SexUponOutcome:** Informação sobre sexo e castração/esterilização do animal;
 - **AgeUponOutcome:** Idade do animal ao deixar o abrigo;
 - **Breed:** Raça do animal;
 - **Color:** Cor do animal.

O conjunto de dados

- Provido pelo Abrigo de Animais de Austin
 - Dados de 2013 a 2016
- Conjunto de dados possui os seguintes atributos:
 - **ID:** Identificador único de cada animal;
 - **Nome:** Nome do animal;
 - **DateTime:** Data em que foi realizado o registro;
 - **OutcomeType e OutcomeSubType:** Identifica o que aconteceu com o animal no abrigo;
 - **AnimalType:** Tipo do animal(gato ou cachorro);
 - **SexUponOutcome:** Informação sobre sexo e castração/esterilização do animal;
 - **AgeUponOutcome:** Idade do animal ao deixar o abrigo;
 - **Breed:** Raça do animal;
 - **Color:** Cor do animal.

O problema

- Qual será o provável destino do animal ao entrar no abrigo?
 - Destino pode ser: adoção, morte, eutanásia, retorno ao dono e transferência para outro abrigo.

Solução

- Gerar um modelo de dados a partir das informações disponíveis a fim de prever o destino do animal.

Conjunto de dados adicionais

- Adquiridos dos websites *Dogs Breeds List* e *Cats Breeds List*
 - Dados adicionais não foram encontrados para cerca de 5% dos registros no conjunto original
- Conjunto de dados possui os seguintes atributos:
 - **type:** Tipo do animal(gato ou cachorro);
 - **breed:** Raça do animal;
 - **other_names:** outros nomes de raça para o animal;
 - **size:** tamanho do animal;
 - **life_span_low/high:** tempo de vida mínimo e máximo estimado;
 - **price_low/high:** preço de compra mínimo e máximo estimado;
 - **intelligence:** grau de inteligência do animal;
 - **adaptability:** grau de adaptabilidade do animal;
 - **child_friendly:** valor indicando o quão amigável com crianças o animal é;
 - **cat_dog_friendly:** valor indicando o quão amigável com gatos ou cachorros o animal é;
 - **grooming:** grau de necessidade de cuidados para com o animal;
 - **health_issues:** valor indicando o quão propenso o animal é a ter problemas de saúde.
 - **shedding_level:** valor indicando o quanto o animal solta pelos;

Conjunto de dados adicionais

- Adquiridos dos websites *Dogs Breeds List* e *Cats Breeds List*
 - Dados adicionais não foram encontrados para cerca de 5% dos registros no conjunto original
- Conjunto de dados possui os seguintes atributos:
 - = **type:** Tipo do animal (gato ou cachorro);
 - = **breed:** Raça do animal;
 - = **other_names:** outros nomes de raça para o animal;
 - **size:** tamanho do animal;
 - **life_span_low/high:** tempo de vida mínimo e máximo estimado;
 - **price_low/high:** preço de compra mínimo e máximo estimado;
 - **intelligence:** grau de inteligência do animal;
 - **adaptability:** grau de adaptabilidade do animal;
 - **child_friendly:** valor indicando o quão amigável com crianças o animal é;
 - **cat_dog_friendly:** valor indicando o quão amigável com gatos ou cachorros o animal é;
 - **grooming:** grau de necessidade de cuidados para com o animal;
 - **health_issues:** valor indicando o quão propenso o animal é a ter problemas de saúde.
 - **shedding_level:** valor indicando o quanto o animal solta pelos;

Modificações no conjunto de dados

Atributo original	Atributo(s) derivado(s)
name	hasName
sexUponOutcome	sex e isIntact
ageUponOutcome	monthsOld
breed	breed1, breed2 e isMix
color	color1 e color2

Escolha do algoritmo

...

Algoritmos vistos em aula

- KNN
- SVM
- Naïve Bayes
- ID3
- C4.5

Algoritmos vistos em aula

- ~~KNN~~
- SVM
- Naïve Bayes
- ID3
- C4.5

Algoritmos vistos em aula

- ~~KNN~~
- ~~SVM~~
- Naïve Bayes
- ID3
- C4.5

Algoritmos vistos em aula

- KNN
- SVM
- Naïve Bayes
- ID3
- C4.5

Algoritmos vistos em aula

- KNN
- SVM
- Naïve Bayes
- ID3
- C4.5

Análise dos resultados

...

Algoritmos e Resultados			
Algoritmo Utilizado	Parâmetros	Score Kaggle	Acurácia no Conjunto de Treino
C4.5 (J48)	-C 0.25 -M 3	0.93655	64.2037%
C4.5 (J48)	-C 0.25 -M 2	0.96796	64.3122%
REPTree	-M 2 -V 0.001 -N 3 -S 1 -L 1 -I 0.0	1.80906	66.7627%
C4.5 (J48)	-C 0.7 -M 2	1.86157	71.5328%
RandomTree	-K 0 -M 1.0 -V 0.001 -S 1	6.14302	83.9687%