DATA SCIENCE CANVAS

Customer: Data Detectives | Project name: Dog Aging Project | Iteration | Date: 07.Feb. 2024

Problem Statement

With the development of society, dogs play an increasingly important role in people‘s lives, which makes people pay more attention to the health of dogs. The primary problem is the challenge in effectively predicting and managing health issues in dogs, which includes identifying more relevant factors and more accurate risk of disease.

Solution

We design a user friendly application based on machine learning to predict dog’s individual health condition types, the health issues throughout on dogs lifespan, and inform user which variables determine health risks.

Modeling

Naive Bayes; Random Forest; Gradient tree boosting; Logistic regression; MLP; Extreme gradient boosting

Users & Use

-Veterinarians, canine researchers, and dog owners can use this model to deepen their understanding of dogs in order to better study/care for dogs.
-People who want to raise a dog can use this as a reference to choose the breed of dog they want to raise.
-Insurance companies can develop more appropriate insurance plans for dogs based on the disease risks of different types of dogs.

Data Acquisition

Data available:
-Basic information about dogs according to the dog id: pure or mixed, sex, age, physical activity, disease, diet.
-Basic information about dog owners: income.
-Environment data: surface of physical activity, months weather hot per year.
Data to be acquired:
-Breed information: breed type, sub breed, breed group, breed.
-The census division to US states mapping

Analytics Formulation

Input:
-Information about breed: breed group, individual breed, breed type, pure or mix.
-Dog’s basic information: age, sex.
-Owner’s information: demography, income.
-Others: climate, physical activity.
Output:
-Risk of getting different diseases.
Methodology:
-Multilabel classification to get probabilities

Data Preparation

-Data Cleaning: Handle missing values, outliers, and convert data types as necessary.
-Feature Engineering: Extract, construct, and select the most useful features from the raw data.
-Data Splitting: Divide the data into training, validation, and test sets.

Evaluation

-K-fold stratified cross validation
-Evaluation Metrics: AUC-ROC for classification models

Success Criteria

Success is based on both objective and subjective criteria:
Objective:
-high and transparent predictive accuracy
-Numerous opportunities for further advancement and diverse application possibilities
Subjective:
-User satisfaction, improved healthcare outcomes for dogs.
-Data privacy & protection

MVP

A basic web application allowing users to input dog and their owner information and receive a simple condition types risk prediction.

Constraints

-Dataset Constraints: Missing datapoints for certain survey-questions, data privacy & protection
-Algorithmic Constraints: computing complexity, computation time is relatively long

Business Value

-For veterinarians: The ability to predict diseases in dogs more accurately leads to earlier interventions, potentially reducing the severity and cost of treatment.
-For dog owners: Preventive health measures and early diagnosis can significantly reduce long-term healthcare costs for their pets. By providing valuable, personalized health insights, the project increases engagement from dog owners.
-For insurance companies: The predictive insights from the project enable more accurate risk assessment for individual pets. Insurance companies can use this data to tailor policy pricing, reflecting the actual health risk of each pet.
-New products/services: Insights gleaned from the data can guide the development of new products or services tailored to specific canine health needs, such as specialized diets, supplements, or health monitoring devices.

Key Actors

-Internal: Data scientists, software developers, UI/UX designers.
-Customer Stakeholders: Veterinarians, dog owners.
-External: Technology Vendors and Cloud Service Providers, Animal health experts

Technology stack

-Communication and project management: Zoom, WhatsApp, miro, discord
-Software Development: React, HTML, CSS, JS, Pycharm, Jupyter notebook, Visual Studio Code, Flask