Predicting Common Dog Diseases using Supervised Machine Learning Classification

Barbarona, Lucky | Candel, Sheena | Senepete, Jayve | Sy, Jascha

Abstract

In this study, the researchers developed a machine learning model that predicts common dog diseases in Cagayan de Oro City's Veterinarian Office, providing pet owners with a guide to identify and mitigate these conditions early on. The model uses a dataset from Github that includes various dog diseases and their corresponding symptoms, which underwent data cleaning and preprocessing for optimal modelling. The study utilized K-Nearest Neighbor, Support Vector Machine, and Naive Bayes Classification, and used an ensemble as a voting classifier for better predictions, achieving a 95% accuracy. The model is deployed through a website application, and the researchers suggest that collaboration with the academic veterinary field can further improve its purpose. Moreover, expanding the dataset can avoid overfitting/underfitting, and incorporating additional common dog diseases can widen the scope of predictions.

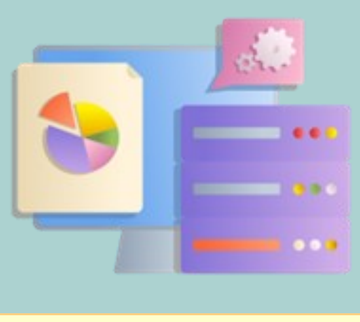
Introduction

Animal health is a critical aspect that affects food production, environmental welfare, and pet owner's safety. A variety of animal diseases have been identified, making it difficult to categorize symptoms, and diagnostic tests are recommended. Predicting the emergence and spread of animal diseases on local and global scales is essential to prevent the spread of animal diseases and minimize the risk of exposure. Common animal health issues were identified through databases of insured pets, and Supervised Machine Learning Classification was established to predict common dog diseases. This study aims to help pet owners take better care of their pets and prevent unwanted events due to animal health problems. The lack of care from some pet owners and the absence of animal health records hinder the development of local animal health status. With the use of the internet and search engines, pet owners evaluate their pet's medical symptoms before seeking an official medical diagnosis.

Methodology

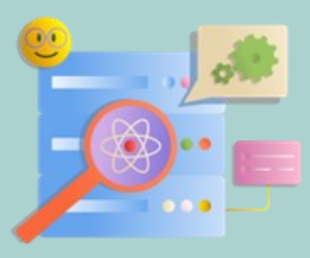
Data Acquisition

The dataset used is from Github and was confirmed and checked in Cagayan de Oro City Veterinarian Office



Data Preprocessing

The data is reduced and cleaned according to the present common dog diseases in local areas which removes one disease and some symptoms that are hard to determine.



Data Modeling

The machine learning models used in this study are, KNN, Naïve Bayes Classification, Support Vector Machine and Ensemble. The test size used for train test split of the data is 0.2 with random state of zero (0).

Website Development

The researchers used Flask to display the website application

List of Diseases

1	Canine Distemper
2	Canine Parvovirus
3	Heartworm
4	Kennel Cough
5	Rabies
6	Leptospirosis
7	Kidney Disease

Prognosis

Chances of Recovery

Excellent—Having a high chance of surviving for a certain period of time or that treatment is likely to be successful.

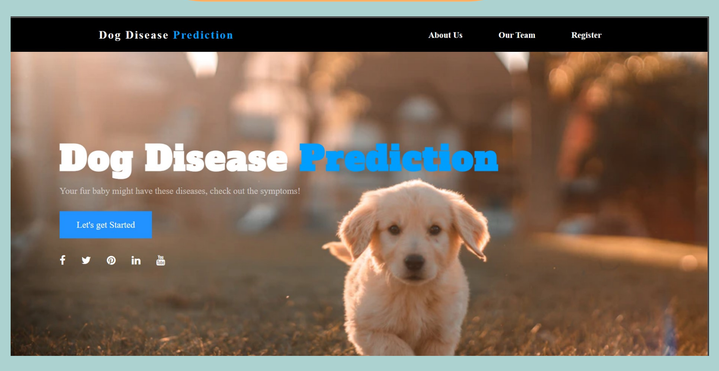
Good—it typically means that the individual has a good chance of recovery or a positive outcome and has a high chance of surviving for a certain period of time

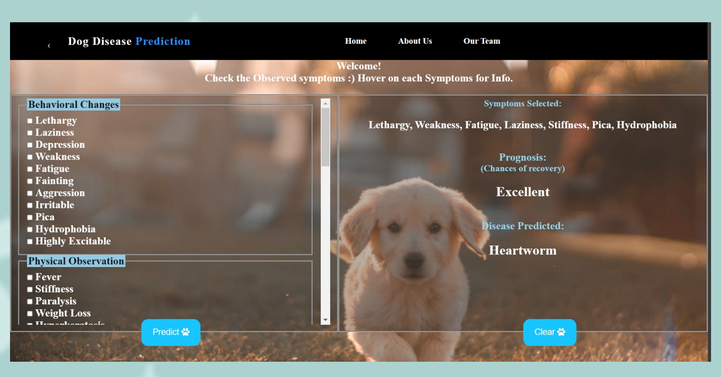
Fair- it typically means that the individual has a moderate chance of recovery or a positive outcome.

Poor- A poor prognosis is a negative prediction of the outcome, but it does not mean that it will not recover or that there are no options left and may not heal completely, or will take a long time to heal and the individual may have permanent limitations.

Results and Discussion

Website





The model obtained 86% accuracy in KNN, 95% in SVM, 95% in Naïve Bayes, which results to 95% accuracy in Ensemble model. These results means that all the model used performed well

The researchers recommend that to improve the website's purpose, collaboration with the academic veterinary field would be a great help to improve this study. Moreover the data needs to expand to avoid overfitting/underfitting. Additional common dog diseases are also better for wider scope in prediction. Hence, this research is a stepping stone for giving the right attention and care not just for dogs but the dog owners as well, and could possibly help the fur community if the study is improved and continuously developed. To better serve its purpose, deploying the web application online is one of the goals of this study.