PetRisk: A Data-Driven Breed-Based Disease Risk Analysis System Using FastAPI

Aarushi
School of Information Technology, Vivekananda Institute of Professional Studies - Technical Campus, AU- Block (Outer Ring Road), Pitampura, Delhi - 110034, India
aarushihari05@gmail.com

Malya Kapoor
School of Information Technology, Vivekananda Institute of Professional Studies - Technical Campus, AU- Block (Outer Ring Road), Pitampura, Delhi - 110034, India
malyakapoor4@gmail.com

Meenu Chopra
School of Information Technology, Vivekananda Institute of Professional Studies - Technical Campus, AU- Block (Outer Ring Road), Pitampura, Delhi - 110034, India
drmeenuchopravips@gmail.com

Cosmena Mahapatra
School of Information Technology, Vivekananda Institute of Professional Studies - Technical Campus, AU- Block (Outer Ring Road), Pitampura, Delhi - 110034, India
cosmenavips@gmail.com

Abstract- ***With the changing face of veterinary medicine, proactive care is becoming more and more necessary. This paper presents a dedicated backend system that will be used in predictive pet healthcare with a focus on minimising breed-specific diseases. Using FastAPI and SQLAlchemy, the system supports the assessment of risks in detail and the identification of similarities between breeds based on the past trends of diseases. We combine a modular similarity model that will produce personalized preventive care plans. The findings indicate that an API-based, structured architecture can go a long way in enhancing the provision of health insights to offer a scalable foundation of modern veterinary tools.***

Keywords— Pet HealthCare, Disease Prediction, FastAPI, Similarity Analysis ,SQLAlchemy, Risk Analysis

I. INTRODUCTION

The closer a human being and their pets get, the more advanced health monitoring would be required. Predictable health issues in cats and dogs are often due to genetic predisposition, but many current veterinary systems do not have the structured data analysis that can deliver pet owners with actionable insights to predictable health issues.

To fill this knowledge gap, we created PetRisk, a back-end application, which is developed on a solid API architecture. The main aim of PetRisk is to compute disease risks using an information-based approach, which will enable early intervention and custom care plans that had been hard to orchestrate without expert clinical knowledge until now.

II. METHODALOGY

The method uses a pipelined approach that involves storing, validating, feature extraction, calculating similarity and results production. This methodology uses a modular design, whereby each process, such as CRUD functions, service logic, and database operations, is isolated to improve maintenance.

A. Dataset Description

The database that will be used in this particular system is stored in a relational database management system (RDBMS). It contains information on diseases of different breeds. It includes characteristics such as breed name, species, disease, and its prevalence. The data is categorized under three main tables which include Breed, Diseases, and BreedDiseasesLinks. The table of BreedDiseasesLinks uses a many-to-many relationship that allows for many breeds to be connected with multiple diseases.

B. Data Cleaning

Data cleaning is done at the backend stage. In order to avoid multiple breeds, uniqueness constraint is imposed on breed, and for disease entries in a request, checking of duplicates is done based on sets at the CRUD layer stage. Prevalence values are made sure to fall in the permissible range of 0 and 1 with database constraints. Additionally, there are references between foreign keys in different tables. All these make up the data validation strategies employed by the system.

C. Feature Selection

The selected features include the breed name, species, disease name, and disease prevalence. These features are obtained through SQLAlchemy ORM queries and are key elements in carrying out similarity and risk analysis. The purposeful selection of these features ensures that only relevant data is used in calculations, which reduces complexity in the process.

The database that will be used in this particular system is stored in a relational database management system (RDBMS). It contains information on diseases of different breeds. It includes characteristics such as breed name, species, disease, and its prevalence. The data is categorized under three main tables which include Breed, Diseases, and BreedDiseasesLinks. The table of BreedDiseasesLinks uses a many-to-many relationship that allows for many breeds to be connected with multiple diseases.

D. Model Development

The system implements a rule-based analytical model instead of a traditional machine learning model. The core logic is implemented in the service layer, where similarity between breeds is calculated based on shared diseases. The similarity measure follows a Jaccard-based approach, where the ratio of common diseases to the total unique diseases between two breeds is computed.

The formula used is:

Similarity

$$(A,B)=|DA\cap DB| \,/\, |DA\cup DB|$$

(Where D_A and D_B represent the sets of diseases associated with two different breeds.)

The workflow is seamless: an API call triggers the service layer, which queries the database, calculates the similarity scores, identifies common ailments, and ultimately generates a preventive healthcare plan in a structured JSON format.

E. Model Assessment

We have evaluated its performance considering three important indicators such as accuracy, speed of response, and user-friendliness. The system was very efficient in identifying similar breed risks and Fast API guaranteed low-latency response. Although the modular nature of the system predisposes it to being easily updatable, its overall effectiveness is still tied to the amount and quality of the underlying database.

III. LITREATURE REVIEW

Research carried out in veterinary medicine underlines the fact that the need to implement strategies that will guarantee early diagnosis and prevention of breed-specific diseases. The traditional method is much reliant on diagnostic processes and human abilities, and the latter cannot always result in the derivation of data-driven inferences. With the invention of data mining and web technologies, researchers can now utilize clever tools of data processing to process voluminous data and to discover patterns.

Several measures based on similarity such as the Jaccard Index have been demonstrated to be very useful measures in the analysis of data to assist in clustering and comparison of data. They allow realizing the connection between the items based on the similarities. But very little is known as to how these ways have been put into practice in the pet medicine business. This paper discusses a system that fills in the gap in research.

A. Abbreviations and Acronyms

API-Application Programming Interface.
ORM - Object Relational Mapping.
DBMS-Database Management System.

B. Units

In this model prevalence of a disease is assumed to be a number without units which falls between 0 and 1. The more the value is drawn to 0, the less likely the disease will affect a given breed; the more the value is drawn to 1 the more likely the disease will affect a given breed.

C. Formulas

The breeds similarity is calculated as indicated by the Jaccard similarity index:
Similarity:

$$(A,B) = |D_A \cup D_B| / |D_A \cap D_B|$$

By so doing, a normalized similarity value will be given in the range [0, 1].
Studies conducted in veterinary medicine emphasize that it is essential to adopt strategies that ensure early diagnosis and prevention of breed-specific illnesses. The traditional approach relies significantly on diagnostic methods and human expertise, and this cannot always result in the extraction of data-driven conclusions. With advances in data mining and web technologies, researchers can now utilize intelligent data processing tools for large-scale data processing and pattern recognition.

Various similarity-based approaches, such as the Jaccard Index, have proven to be highly efficient tools in data analysis for purposes of clustering and comparison. They allow recognizing connections between objects based on shared properties. Yet, there is little evidence of how these methods have been used in the field of pet medicine. This paper discusses a system that fills in the gap in research.

IV. DISCUSSION AND RESULT

Besides, the system will develop a preventive care program depending on the risks identified, which will make them more applicable. The fact that SQLAlchemy ORM is used also means that efficient queries are used and FastAPI is also characterized by being fast and the ability to properly handle multiple requests simultaneously.

The system will be characterized by low latency and output consistency as far as efficiency is concerned. Modular design approach, which divides the CRUD functionality, service functions and database procedures increases flexibility and expandability. The quality of the final product will however be dependent on the quantity and quality of input data. The similarity will not be true reflectors of reality when a small sample of data are used. Nevertheless, the system proves the possibility of the use of the structured data and the analysis based on rules in the pet health care system.

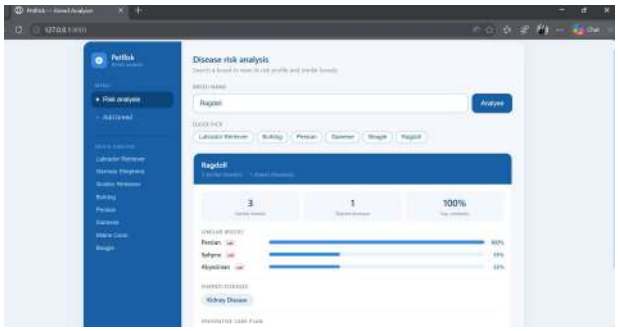


Figure 1 User Interface

V. FUTURE WORK

The system however can be optimized to be more efficient by adding some improvements to the system. First, machine learning algorithms, including Naive Bayes, to predict diseases, should be added. It will help the program to work with the similarity measure far better, and implement some data analysis techniques in order to develop even more sophisticated predictions. Additionally, the new version of the program might have additional details on various diseases.

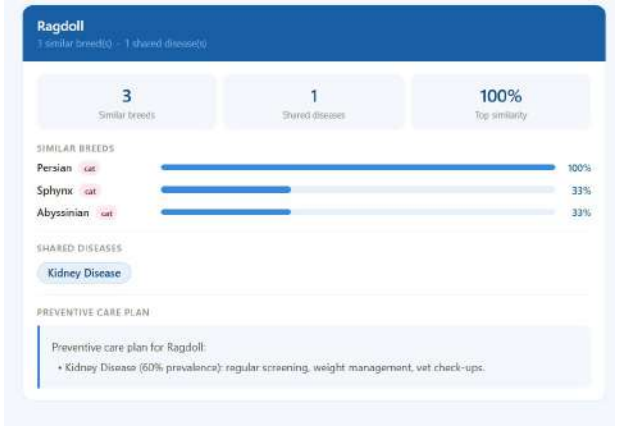


Figure 2 Result

The program could get some other databases to examine with a veterinarian or make use of some APIs to enrich the database. Another improvement can be adding data to the dataset to accommodate data of other animals such as cats, rabbits etcetera.

Nevertheless, discussing the technical advancements, it should be stated that a cloud version of the application should be developed in order to make it more available to the users. In addition, it is possible to think of creating a mobile version of the software as an application. And, lastly, the program might be connected to the Internet of Things (IoT) and a few sensors and health trackers that pets can have to monitor their health.

VI. CONCLUSION

On the conclusion, PetRisk seems to be one of the most efficient tools to evaluate the diseases related to particular pet breeds with the help of the appropriate backend design. When using FastAPI, SQLAlchemy, and similarity analysis algorithm, users will have a chance to obtain valuable

information about the diseases and how to prevent them. Following, modern technologies can be used to construct healthcare services.

The proposed solution, besides the possibility of storing the relations between breeds and diseases in a well-organized manner, also conducts an analysis at the moment. Furthermore, owing to the architecture modularity, it will be rather easy to expand the application in the future. Despite the fact that the current model has a few limitations associated with its rule-based framework, there are numerous opportunities to enhance it and implement the methods of machine learning.

In general, the given solution has its contribution to the sphere of animal healthcare, as it helps avoid diseases, but not to treat them.

REFERENCES

[1] FastAPI Documentation, https://fastapi.tiangolo.com
[2] SQLAlchemy Documentation, https://www.sqlalchemy.org
[3] J. Han and M. Kamber, *Data Mining Concepts and Techniques*
[4] Research papers on veterinary disease prediction
[5] IEEE Paper Writing Guidelines