**AN ANIMAL DISEASE PREDICTION SYSTEM USING MACHINE LEARNING**

**BY**

**JOHN KIPNGENO**

**IN16/00024/20**

**MERCY CHEPKEMOI KILEL**

**IN16/00032/20**

**SCHOOL OF INFORMATION SCIENCE AND TECHNOLOGY**

**A Project proposal submitted to the school of information science and Technology for the study leading to a Project Fulfillment of the Requirements for the award of the Degree**

**DECLARATION**

This work has not been copied nor taken from any other done work.

Signature: ...................................... Date: ....................................................

John Kipngeno

Signature: ...................................... Date: ....................................................

Mercy Chepkemoi

This thesis has been submitted for examination with the approval from the University by the supervisor

Signature: ………………………………………………. Date: …………………………………………………………………

**DEDICATION**

This thesis is Dedicated to all the farmers who have find challenged looking for lasting solutions to control and make prediction of diseases affecting their animals.

**ACKNOWLEDGEMENT**

We would like to acknowledge the University for providing the adequate resources which has contributed to the success of our research. Also acknowledgement to the founders of the powerful network browsers that instill much of the education.

**TABLE OF CONTENTS**

DECLARATIONi

DEDICATIONii

ACKNOWLEDGEMENTiii

ABSTRACTiv

CHAPTER ONE1

1.0 Introduction1

1.1 Background2

1.2 Problem Statement3

1.3 Objectives of the Study3

1.4 Scope and Boundary3

1.5 Justification4

CHAPTER TWO6

2.0 LITERATURE REVIEW6

2.1 Overview and Analysis7

2.2 Body10

2.2.1 Prevalence of livestock disease 10

2.2.2 **Livestock disease identification using Prediction Techniques** 11

CHAPTER THREE12

3.0 METHODOLOGY12

3.1 Introduction12

3.2 General Requirements12

3.3 Tools and Specification13

3.4 Implementation of the Animal Disease Prediction System15

3.4.1 Procedure15

3.5 Used Machine Learning Algorithm15

3.6 Data Collection and Analysis15

CHAPTER FOUR16

4.0 SYSTEM ANALYSIS AND DESIGN, TESTING AND VALIDATION16

4.1 System Analysis16

4.2 Requirement Specification16

4.2 Functional Requirements16

4.2 Non-Functional Requirements16

4.2 Hardware and Software Requirements16

4.3 System Design17

4.4 Testing and Validation21

CHAPTER FIVE

5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS26

5.1 SUMMARY26

5.2 CONCLUSIONS27

5.2 RECOMMENDATIONS27

REFERENCES28

**ABSTRACT**

The main purpose of this research is to build a system and a model with machine Learning that is used to predict animal diseases from symptoms. The main issue is that Late detection of animal diseases has a major effect on both the economy and health in both animals and humans.

The research objectives are: To curb the challenge of late treatment of animals from late disease detection, reduce human-animal contact diseases, to design a model with random forest classifier and Decision tree classifier that predict animal disease from symptoms using machine learning. The system I developed using the python jupyter notebook and the GUI written with the Flask framework and Html. The Data used in training the model used is from an internet source retrieved as a csv file.

This system will help in early detection of animal diseases and early treatment and isolation, which will reduce both human-animal contact diseases and prevent the spread of the disease to other healthy animals.

**CHAPTER ONE**

1. **INTRODUCTION**
   1. **Background**

According to the research on Animal Health, ecological health is a fundamental asset for the sustainability of both human and animal existence. Animal health is an integral component of ecological health. For instance ‘a small country such as Kuwait which is 17,818km2 surrounded by much larger countries such as Saudi Arabia, Iraq and Iran, makes animal health a national asset and ranks high in importance in the same class of local natural resources such as oil‘. The most recent foot and mouth disease (FMD) cases were identified in Saudi Arabia and Iraq in 2016. The world must be alert to new threats from animal diseases, thus a need to greater concern to the animal health, “We need to prioritize on the animal health sector” FAO director General Qu Dongyu said at a Launch Event of a Dairy Monitoring System. Strong animal disease prediction systems are the key to prevent diseases, ensure safe and nutritious food and protect farmers interests. FAO regards animal health as critical to food and nutrition security and achieving many of the sustainable development goals, especially those related to improving production, eradication of poverty and ending hunger to those depended to the animal farming and curbing the human diseases from contact with animals, et al. 2014)

The recent appearance of bovine Spongiform encephalopathy (BSE) in the UK have increased awareness and concern about the potentially devastating impacts of these and other animal diseases. This thus requires a system that predicts the occurrence and the infections of the disease by an input of the known symptoms and the observations of the unusual behavior of the animals, (Smith et al 2013)

It has become clear that there is an urgent need to understand and monitor the status of animal health from historical perspective and to build an early disease detection system for animal diseases from the unusual observations of the animal behavior and the determined symptoms, with the collective goal to track the status of animal health. Livestock is a source of subsidiary income for many families in Africa especially the resource-poor who maintain few heads of animals. Therefore with a animal’s disease prediction system, (Signh et al 2015)

The concepts of animal 'health' and 'welfare' is increasingly becoming linked. But are they really indissociable, or even synonymous? Animal health is considered as simply the absence of disease. However, recent advances have been made to take into consideration the health of animals for a sustainable improvement of animal productivity,(Zhang et al 2014).

Supervised machine learning has been one of the branches of machine learning that has eased the development of powerful systems that can be trained with data. Different machine Learning algorithms have been used (such as Logical regression, Random Forest, Decision Tree) to make the disease identification which contributes to early treatment easier. One recent study has demonstrated the real-time prediction of Zika virus outbreaks in Brazil using the k-nearest neighbors (KNN) algorithm by adapting dynamic variables as predictors, such as the number of daily international travelers, as a predictor. K-nearest neighbor is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It’s easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.  The datasets were related to different disease contexts, (Morsy et al. 2019). Neural Networks was implemented and their performance was accessed in different tasks including the construction of missing data in a time series. Neural Networks implemented presented a considerable prediction error which may be due to the use of few training data. The strength of the algorithms was easier to implement, interpret, and very efficient to train. Made no assumptions about distributions of classes in feature space. The main challenge was that the algorithms accepted large volumes of data, but the procedure led to inaccurate estimates and required more data, (Kemp et al. 2010)

With the various testing and implementation of the various algorithms, a combination of the Random Forest and the decision tree algorithms could yield an effective disease prediction system due to the ease of interpretation, high performance on large datasets, extremely fast and takes less time to train the model.

* 1. **Problem Statement**

Late identification of animal diseases has posed a threat to both economy animal health and human health. Some diseases such as the Aflatoxicosis, African horse sickness, Akabane disease, African swine fever, Anthrax, Australian bat lyssavirus, Avian influenza, Avian paramyxovirus are not easily identified from a single symptom but with the presents of more than one symptom most diseases can be easily predicted. Animals are prone to many diseases, some of which can decrease productivity and lower production, if not identified at an early stage, can also contribute to the death of cattle, which is greatly impeded by the sustainable development of the national economy, (Rosa et al. 2021). There are many unusual behaviors that are observed which may be a sign of a developing disease in an animal therefore with little knowledge of the diseases identification of the disease seems a challenge.

Therefore, with the presence of the disease prediction system the farmer or any animal attendant can key in the unusual observations or the symptoms and the system predict the disease that the animal is suffering from and thus can make consultation about the disease and finally the animal can be isolated and treated at the early stage.

* 1. **Objectives of the Study**
     1. **Main Objective**

The main objective of the study is to study the previously developed animal disease prediction systems and make the improvements by developing an animal disease prediction system that curb the main challenges of the developed systems such as the prediction accuracy.

* + 1. **Objectives**

1. To analyze current developed systems for animal disease prediction
2. To find out the challenges of the developed animal disease prediction systems
3. To develop an animal prediction system that accepts symptoms as input to predict animal’s disease.
4. To test and implement the developed animal disease prediction system.
   1. **Scope and Boundary**

The livestock diseases have been well researched and documented but a disease cannot be identified from a single symptom but with more than one symptom the disease can be easily identified, the system makes use of the symptoms to predict a disease, it accepts symptoms as input from the user then predict the disease. This study will focus more on early identification of the symptoms through frequent monitoring of the animal, the user then chooses the animal’s name from a dropdown of various animal names the chooses the five symptoms identified during the close monitoring and the system predicts the disease which triggers isolation and treatment of the ill animal.

* 1. **Justification**

This project will reduce the spread of diseases between the animals and reduce the occurrence of the human-animal contact related diseases. This will also improve the disease control activity by early detection of the disease, isolation of the affected animals and creation of awareness to curb another occurrence of the disease by either vaccination or any other preventive activity. With the reduction of animal diseases and the promotion of animal health the human lifestyle and the living standard of the dependent population is improved. This will thus spear the economy.

**CHAPTER TWO**

**2.0 LITERATURE REVIEW**

**2.1 Introduction**

This literature review focuses on livestock disease prediction within different geographical region depending on the symptoms input by the owner. Resources used are from 2015 to 2022 time period. This review will shed light on measures that have been coined with the aim of improving economic growth, only achieved through quick and early animal diagnosis which may rather reduce mortality rate thus improved animal farm productivity.

**2.2 Animal Diseases**

(Gilbert et al. 2014), He used structured questionnaires used to compare reporting behaviors between farmers according to their exposure to disease. He used a survey of 69 livestock farmers to assess factors associated with farmers’ decisions to submit biological samples as a means of identifying early stages of disease infection. The survey posed two different kinds of questions to farmers: first, ‘prospective assessment’ in which farmers were asked how often they performed specific actions, and second, ‘retrospective assessment’ in which farmers were asked to recall specific events over the last 6 months. Analysis revealed that farmers prospective assessments generally overestimate actual disease reporting. In addition, farmers were asked questions about their attitudes and motivations towards veterinarians and disease surveillance. Results also indicated that farmers who were members of herd health schemes would be more likely to submit samples.

(Magwood et al. 2018), used a questionnaire survey of 201 pig farmers in Madagascar about African Swine Fever (ASF) in order to analyze their responses to controlling diseases. Their results show that as the disease spread increases, more farmers are willing to report ASF cases but this also depends on farm-related characteristics and to adapt new technologies, farmers’ knowledge about ASF, administration of the classical swine fever vaccine and previous experiences with ASF.

(Scotch et al. 2012), adopted a retrospective case-control design in which 89 farmers were surveyed by telephone using this approach. Analysis of farmers’ reasons for arising challenge of early prediction of disease, such as the little knowledge, is illustrated both quantitatively and qualitatively using quotations from open-ended survey questions. Separate quantitative modelling is conducted to show how different factors such as knowledgeable farmers are associated easy disease identification and those who have not.

Livestock Disease effects, found out that Livestock is a source of subsidiary income for many families in Africa especially the resource-poor who maintain few heads of animals. One of the major obstacles to achieving the targeted growth rates in the sector is the prevalence and outbreaks of diseases. Disease outbreak among the animals is predicted based on certain conditions and it is also concerned with a specific animal and disease. Animal owners are often unaware of whether the disease is mild or might prove fatal and precautions are to be taken at the appropriate time, (Mak et al. 2010)

This livestock disease is a great threat to animal health as well as to human who are in direct contact with animals and who consumes the product of the animal that has been infected by a certain disease. Livestock animals usually distribute in remote areas with relatively poor conditions for disease diagnosis rapidly and accurately. It is necessary to detect the disease outcome in the livestock to take precautionary measures to avoid spread amongst them. The system will predict the livestock disease based on the symptoms. It will also alert the livestock owner if the predicted disease may cause sudden death. The concepts of animal 'health' and 'welfare' is increasingly becoming linked. But are they really indissociable, or even synonymous? Animal health is considered as simply the absence of disease. However, recent advances have been made to take into consideration the health of animals for a sustainable improvement of animal productivity, (Chang et al 2021).

Animal disease identification using Prediction Techniques” Cattle animals are prone to many diseases, some of which can decrease productivity and lower the quality of dairy products and, if not identified at an early stage, can also contribute to the death of cattle, which is greatly impeded by the sustainable development of the national economy. This paper presents a technique that explains how the use of dataset can diagnose cattle diseases that are rare in farm animal medical facilities that can have cost-effective medical solutions, (Rosa et al 2021)

Effective and viable Animal infection reduction, this deals with an approach to providing a system that contains a dataset of symptoms and records of various animal health conditions. To generate results from the obtained data, the method utilizes the intelligent analysis mechanism functionality of data set. This program is a first aid mechanism that analyzes the signs and symptoms to send you results with algorithms for the prediction of livestock diseases based on data computation. The objective of this systematic review was to answer the research  
the question “what are risk factors associated with lameness in animals that are housed in  
free stall barns or tie stall facilities”. Furthermore, we performed a synthesis of the current, (Kelly et al 2012)

Lameness in animals has been an ongoing concern of great relevance to animal welfare and productivity in modern livestock production. The concepts of animal 'health' and 'welfare' is increasingly becoming linked. But are they really indissociable, or even synonymous? Animal health is considered as simply the absence of disease. However, recent advances have been made to take into consideration the health of animals for a sustainable improvement of animal productivity. Many studies have examined associations between various factors related to housing, management, and the individual animal and the occurrence of lameness, (Zhang et al 2019)

Impact of Livestock disease, the impact of livestock disease has been cited in many publications, but the ability to monitor change is limited as the available data is contained in disparate publications and reports, usually from individual countries, and there are few longitudinal studies of disease prevalence and impact. Donors to international development projects are increasingly interested in being able to monitor change in a country’s performance particularly in response to investment, (Elhady et al. 2018)

Our system intends to use machine learning algorithm technique to predict animal diseases which will help in early disease prediction using the symptoms and signs obtained from observation of the animal behavior and the physical signs. The users of the system which are mostly the livestock farmers would be able to make early prediction and early treatment of the various diseases. The disease control activity would also be of lightweight since early isolation of the infected animals will be of great importance by reducing the spread of the predicted disease.

**2.2 Used Machine Learning Algorithms**

Machine learning algorithms (MLAs) can build a very flexible model to predict the risk of disease in real-time and employ unlimited scopes of data structures, such as texts and images.

Prediction of Zika Virus; One recent study has demonstrated the real-time prediction of Zika virus outbreaks in Brazil using the k-nearest neighbors (KNN) algorithm by adapting dynamic variables as predictors, such as the number of daily international travelers, as a predictor. KNN is a simple, supervised machine learning algorithm that can be used to solve both classification and regression problems. It’s easy to implement and understand, but has a major drawback of becoming significantly slows as the size of that data in use grows.  The datasets were related to different disease contexts. We considered the performance measures of accuracy, precision and recall for comparative analysis. The average accuracy values of these variants ranged from 64.22% to 83.62%. KNN showed the highest average accuracy (83.62%), followed by the ensemble approach KNN (82.34%). A relative performance index is also proposed based on each performance measure to assess each variant and compare the results. This study identified KNN as the best performing variant based on the accuracy-based version of this index, followed by the ensemble approach KNN. This study also provided a relative comparison among KNN variants based on precision and recall measures, (Morsy et al. 2019).

Animal Disease identification by symptom analysis, With Neural Networks and Logical regression demonstrated tremendous abilities for the identification of animal diseases performed well with 91% classification accuracy. The number hidden layers in the model must be selected independently for each particular case. Neural Networks are efficient at handling data that have no cancellation or linearity among them, but they are ineffective for modelling time series, Neural Networks was implemented and their performance was accessed in different tasks including the construction of missing data in a time series. Neural Networks implemented presented a considerable prediction error which may be due to the use of few training data. The strength of the algorithms was easier to implement, interpret, and very efficient to train. Made no assumptions about distributions of classes in feature space. The main challenge was that the algorithms accepted large volumes of data, but the procedure led to inaccurate estimates and required more data, (Kemp et al. 2021)

Animal Health Analysis and Disease Prediction, Random Forest behaves efficiently in terms of time complexity when analyzing the computational complexity. When analyzing the computational complexity of the algorithm. Doshi et al. (2018); implemented RF in animal disease monitoring due to the support incorporated for multilabel classification(MLC) and emphasized that this technique is effective when managing missing values and it is resistant to overfitting of the model, maximum depth of each tree and the number of trees in the forest are the most sensitive classifier parameters. Increasing the number of trees improves the accuracy of the prediction, however it makes the program slower. Easy to be executed in a distributed manner, (Signh et al 2015)

Livestock Disease Monitoring and prediction, this technique requires low computer time to train the model and the ease of its subsequent interpretation. It presented the best accuracy values and area under the curve. DT does not work well with large scattered data sets, since its performance decreases as data volume increases and that it does not ideally detect anomalies and manage missing data. When the DT algorithm was applied to each split, they produced varying results with the third split giving the best accuracy of 73.1%. Then Bagging Ensemble was applied to further increase the accuracy and the third split gave accuracy of 75.1%. When compared with the standard DT algorithm it was observed that after applying split feature reduction and bagging ensemble the accuracy was increased by 4.6%. The Algorithm was highly intuitive and easy to understand, required a smaller number of data preparation steps and can perform multiple rows. The identified challenges of this algorithm where it was time consuming in its training phase and limited performance in solving regression, (Balducci et al. 2012)

With the various test in the used algorithms, Neural Networks is an excellent choice when working with big datasets because they are easy to adapt, which reduces errors based on the data used for training. Random Forest is ideal for working with massive data sets since it needs less time to preprocess the data, is competent in global time complexity and works well with scattered data sets. SVM is suitable for managing small data sets that do not contain to many outliers and its performance is increased when the dimensional space of the data is simple and attributes are slower. Decision Tree requires little computational time and generates a high level of accuracy; it is efficient in terms of computation and scalability.

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Author and Year | Paper Title | Data Set | Machine Learning Type | Strength | Challenge | Performance |
| Saggi et al. (2010) | Animal Disease identification by symptom analysis | Animal disease dataset | Logical Regression, NN | It was easier to implement, interpret, and very efficient to train.  Made ni assumptions about distributions of classes in feature space. | The algorithms accepted large volumes of data, but the logistic regression procedure leads to inaccurate estimates and required more data. | The average accuracy values of these variants ranged from 64.22% to 83.62%. KNN showed the highest average accuracy (83.62%), followed by the SVM approach  (82.34%). |
| Akhtar et al.2019 | Prediction of Zika Virus | Disease dataset | SVM, KNN | We can implement the algorithm with ease.  It is very effective against large data  The decision boundaries that are formed can be of arbitrary shapes. | Its training is slow therefore not efficient for large data  Curse of dimensionality: Domination of distances by irrelevant attributes.  Finding the correct value of k may be time expensive sometimes.  Very high computation cost due to its distance measure. | The accuracy values of these variants ranged from 70% to 90%. KNN showed the highest average accuracy (89%, followed by the SVM approach  90%. |
| Kaur et al. 2014 | Animal Health Analysis and Disease Prediction | Disease Dataset | DT, SVM, KNN | More effective in high dimensional spaces.  They are effective in cases where the number of dimensions is greater than the number of samples. | The Algorithms does not perform very well when the data set has more target classes overlapping.  They work by putting data points, above and below the classifying hyperplane there is no probabilistic explanation for the classification. | The accuracy values of SVM were 89%, followed by the KNN approach  85% and DT 90% |
| Balducci et al. 2012 | Livestock Disease Monitoring and prediction | Disease Dataset | Decision Tree | The quality of the predictions made by the model is dependent upon the quality of data being fed to the model to train on.  The rules implemented by decision trees can be displayed in a flow chart-like manner, allowing data scientists and other professionals to explain to the stakeholders the model’s predictions | High time-consuming in its training phase, and this problem can be exaggerated if there are multiple continuous independent variables  Limited performance in regression | They produced varying results with the third split giving the best accuracy of 73.1%. Then Bagging Ensemble was applied to further increase the accuracy and the third split gave accuracy of 75.1%. When compared with the standard DT algorithm it was observed that after applying split feature reduction and bagging ensemble the accuracy was increased by 4.6%. |

**CHAPTER THREE**

**3.0 METHODOLOGY**

**3.1 Introduction**

The research problem is on late identification of animal diseases, the information used in this thesis are from the journals, news and trusted internet sources such as Wikipedia, google scholar, articles and reports. This section of the research describes the steps that were followed to design the algorithm for disease prediction using Python programming frameworks such as jupyter, flask and the python dependencies and machine learning algorithms.

**3.2 General Requirements**

The general requirements include: The user is required to have knowledge of the symptoms obtained from the monitoring of the animal behavior, the user has then chosen the animal’s name from the drop-down list in the user interface, he/she is then required to choose five symptoms by choosing each symptom from the drop-down list.

**3.3 System Design**

Animal\_Disease\_Prediction

CSV File

Data (Analysis And Splitting)

Testing Set

Training Set

Feature Selection

Classificaton Technique

Decision Tree

Algorithm

Random Forest Algorithm

Prediction of the Disease

**3.3 Tools and Specification**

The tools that contributed to the success of the designing implementation of the system are: Flask framework which helps in implementing a machine learning application to be easily extended and deployed as a web application during the creation of the GUI

HTML, CSS with Bootstrap which we used during the development of the user interface.

Dataset, for machine learning models to understand how to perform various actions, training datasets must be fed to the machine learning algorithm, the dataset thus ensured that the interpretation of the data by the model was accurate

**3.4 Implementation of the Animal Disease Prediction System**

For the user to see a display of the disease from the symptoms the GUI provides a clear interactive page that allows the user to input the symptoms from the observed animal behavior and other observations and onclick of the submit button a disease is displayed.

**3.4.1 Procedure**

The user chooses the name of the ill-health animal from a drop-down list in the first selection with a label animal name, the user then is required to choose five symptoms from the checkboxes preceding the animal’s name checkbox. Onclick of the submit button displays the predicted disease at the bottom of the button.

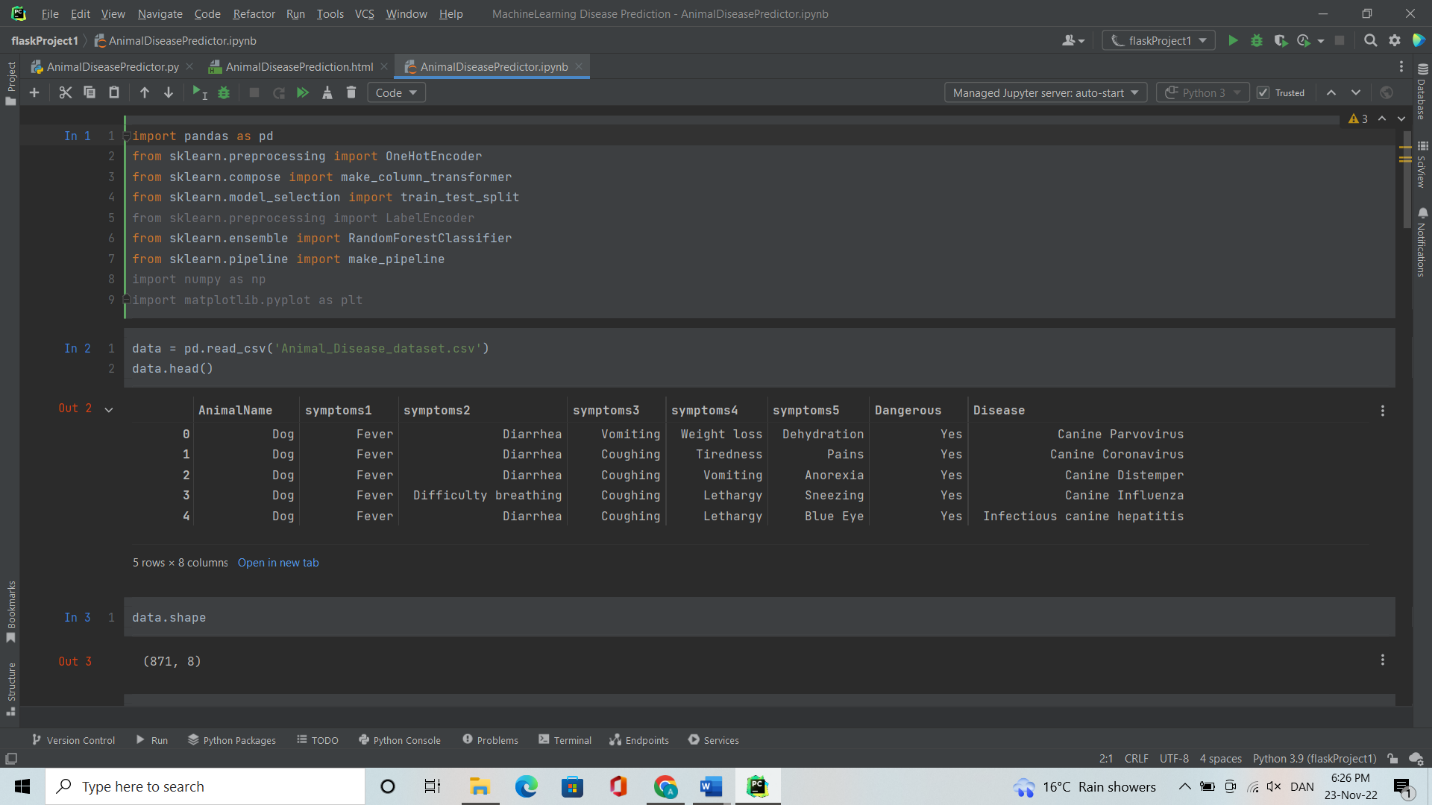
To accurately do the disease prediction from the entered symptoms forest algorithm was written and trained with a dataset which contained animal name with five symptoms and the disease.

**3.5 Used Machine Learning Algorithm**

The python packages used are:

Pandas, which is used for analyzing data. It was also used for data cleansing, data inspection and loading and saving data from the dataset used.

sklearn, this is the python framework that was used for performing classification, regression and clustering algorithms.



Procedure

Importing of the python frameworks and packages is the first step to ensure that the environment is well setup.

Loading the dataset, the dataset to be used to train the model is loaded. The loading of the dataset is done using the following code:

Data= pd.read\_csv(“Input the dataset path”), read\_csv function is a function of pandas used to read the data in csv format.

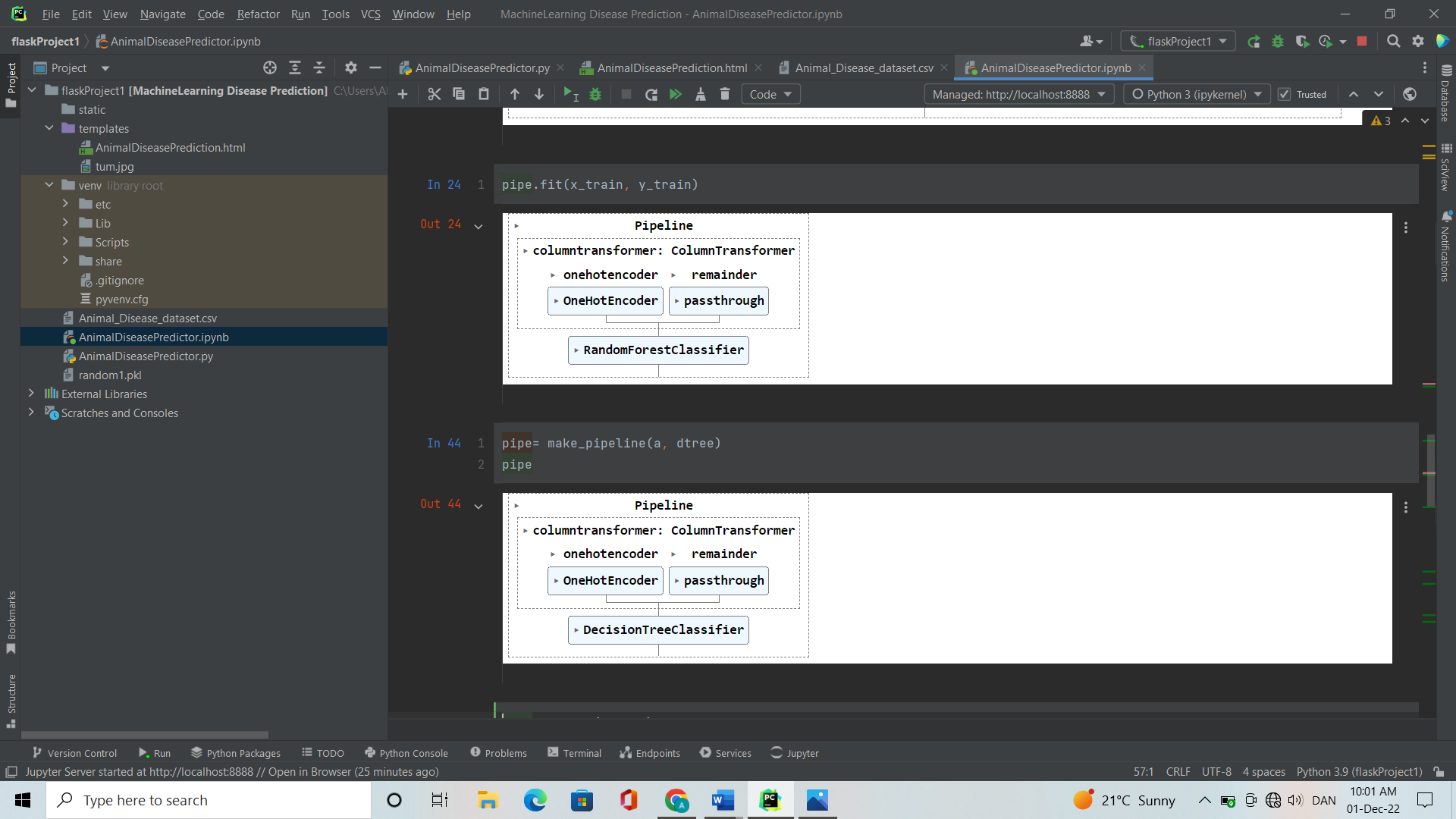
The dataset is then split into training and testing set, first we split our data into input and output. Y is the output and is stored in ‘Disease’ column of the data frame and x contains the other columns and are features or input. The splitting is done with the following code:

x=data.drop(['Dangerous', 'Disease'], axis=1)

y=data.Disease

x\_train,x\_test,y\_train,y\_test=train\_test\_split(x,y, test\_size=.1, random\_state=1)

Training the model is the next step:

The training of the model is done with the following code: 

model = RandomForestClassifier()

pipe.fit(x\_train, y\_train)

model = DecisionTreeClassifier()

pipe.fit(x\_train, y\_train)

model = RandomForestClassifier()

pipe.fit(x\_train, y\_train)

The last step is evaluating the model, we use the predict() to predict the output:

pipe.predict(x\_test)

For classification we use pipeline score attribute.

pipe.score(x\_test, y\_test)

pipe.score(x\_train, y\_train)

**3.6 Data Collection and Analysis**

The data collected was from the downloaded csv dataset file which provided the functional requirements of the model such as the symptoms and the preceding diseases. The dataset contains 484 rows and 50 different diseases with different groups of symptoms

**CHAPTER FOUR**

* 1. **RESULTS AND EXPERIMENTATION**
  2. **Introduction**

This chapter shows the practical part of designing the system.

* 1. **Requirement Specification**
     1. **Functional Requirements**

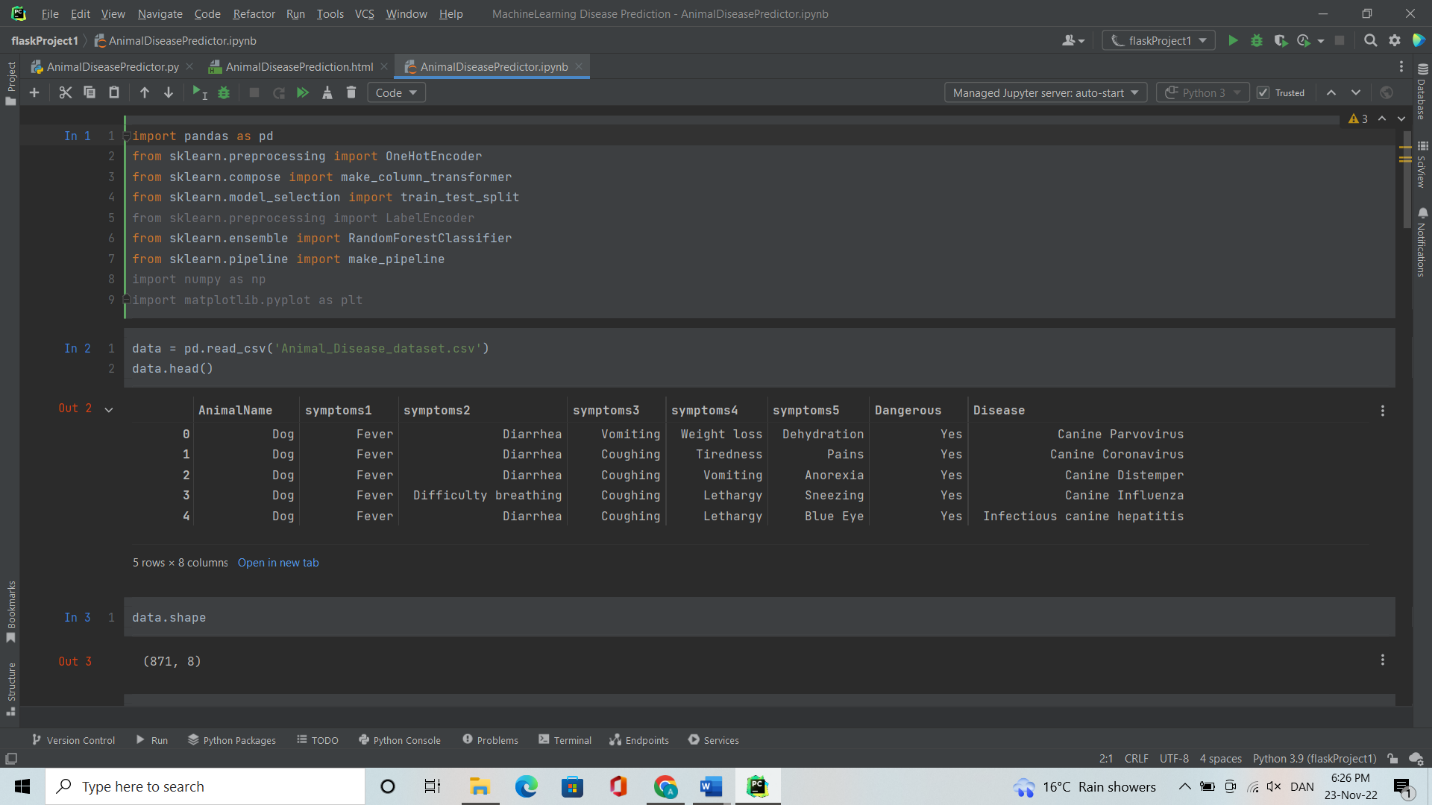
This system requires the user to have knowledge of the symptoms obtained from the close monitoring of the ill-health animal, he/she is then required to choose the animal’s name and five symptoms from the drop-down list in the user interface.

* + 1. **Non-Functional Requirements**

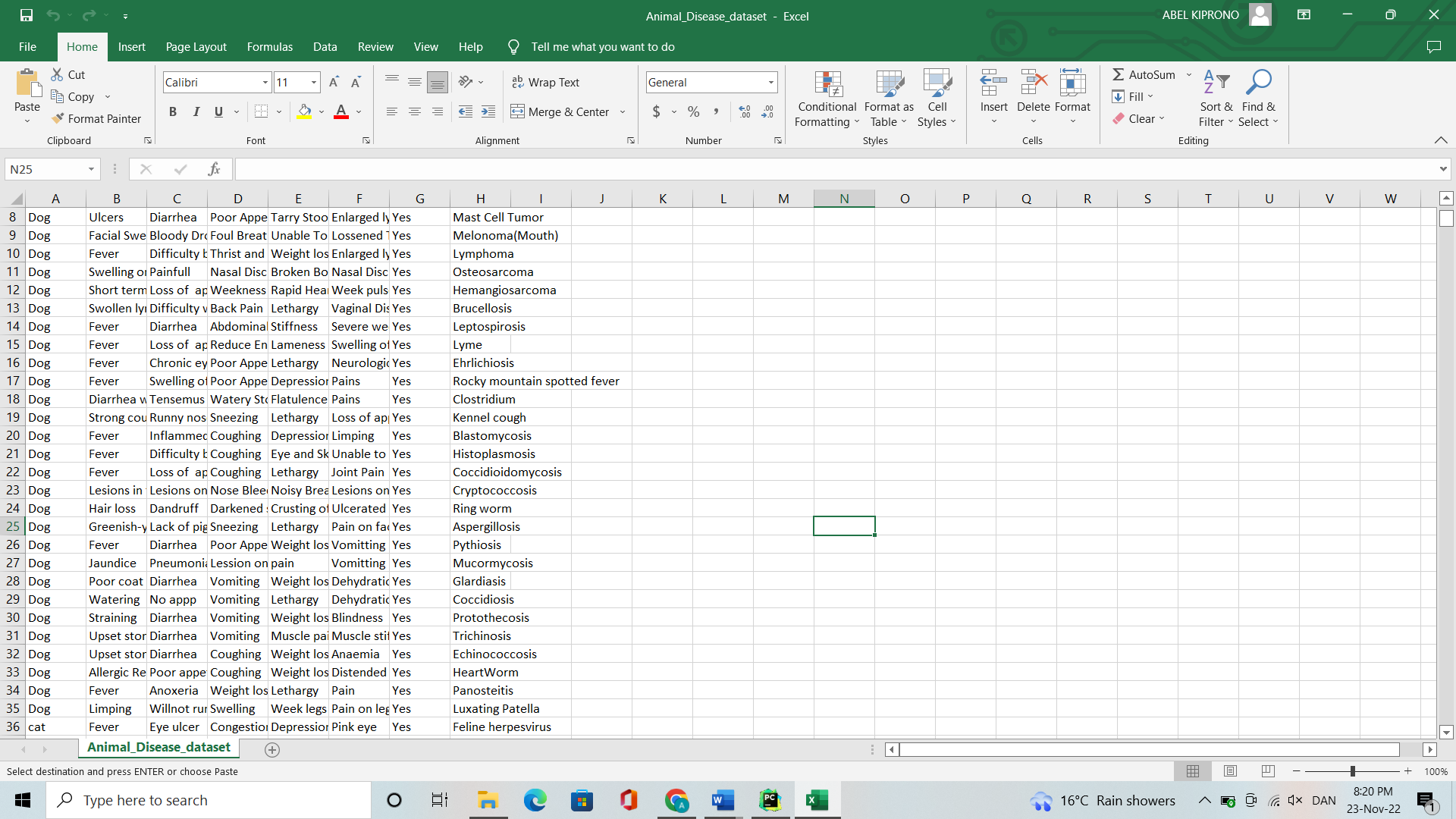
The user must choose the animal’s name before clicking the submit button, the animal’s name is a required choice for the system to predict the possible disease.

* + 1. **Hardware and Software Requirements**

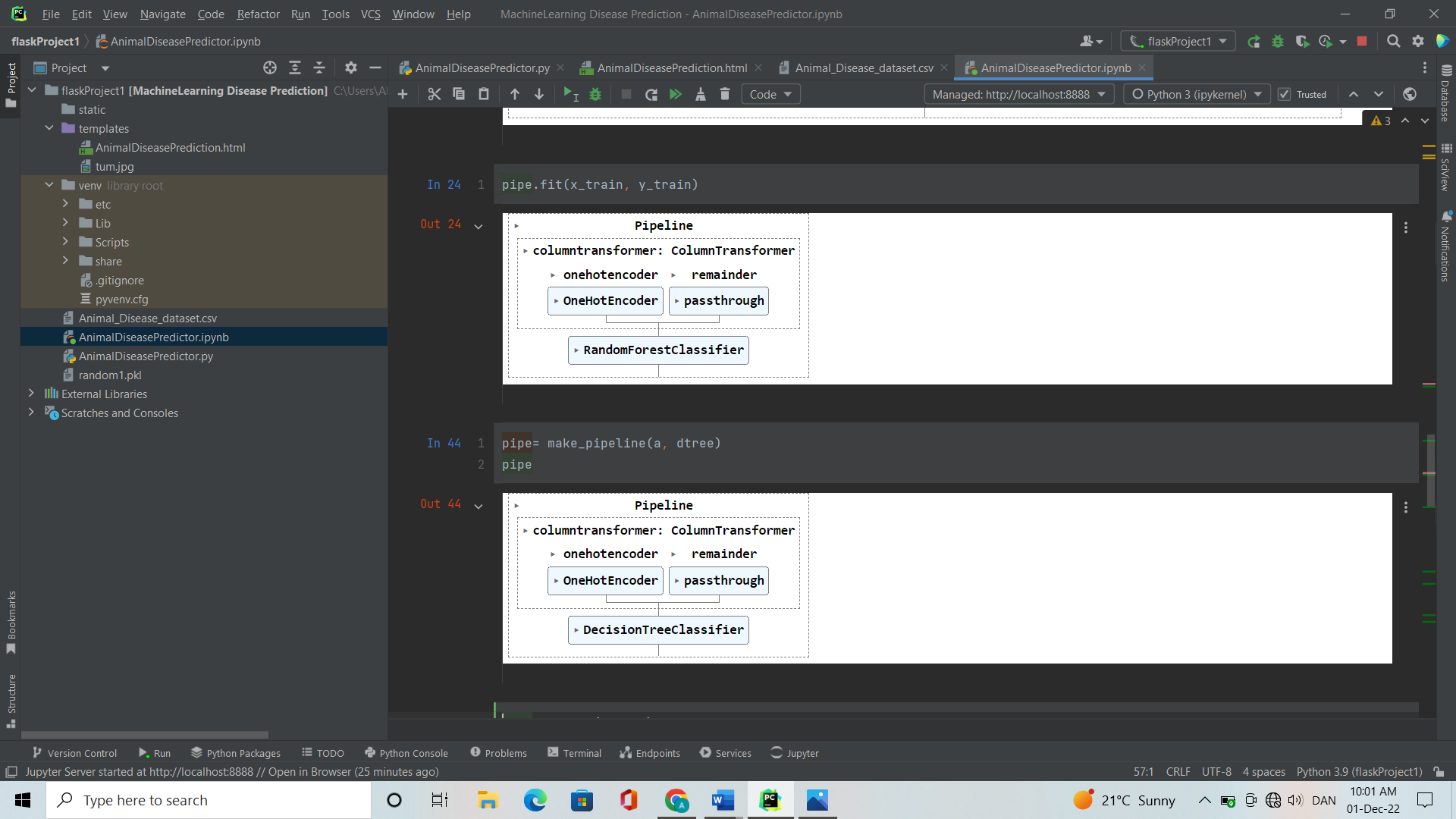
For proper functionality of the system, a Python supporting IDE is required such as the Pycharm Professional Edition which we used during the designing of the system, spyder and the python dependencies and packages installed such as flask, pandas, jupyter with a proper set of the python environment.

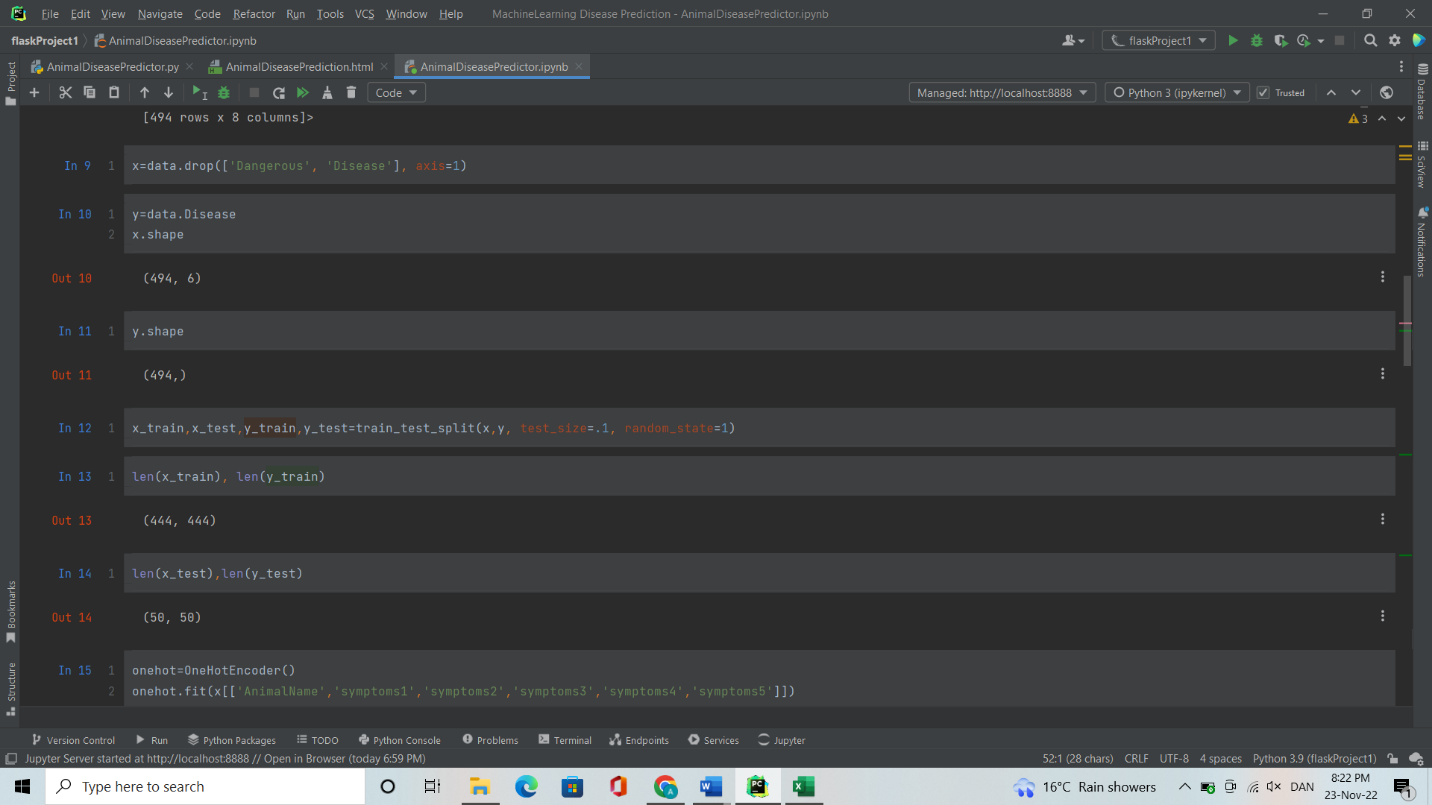
The first procedure was importing the required dependencies and packages such as pandas, sklearn, numpy and pickle module which keeps track of the objects it has already serialized.to avoid repeat serializing.

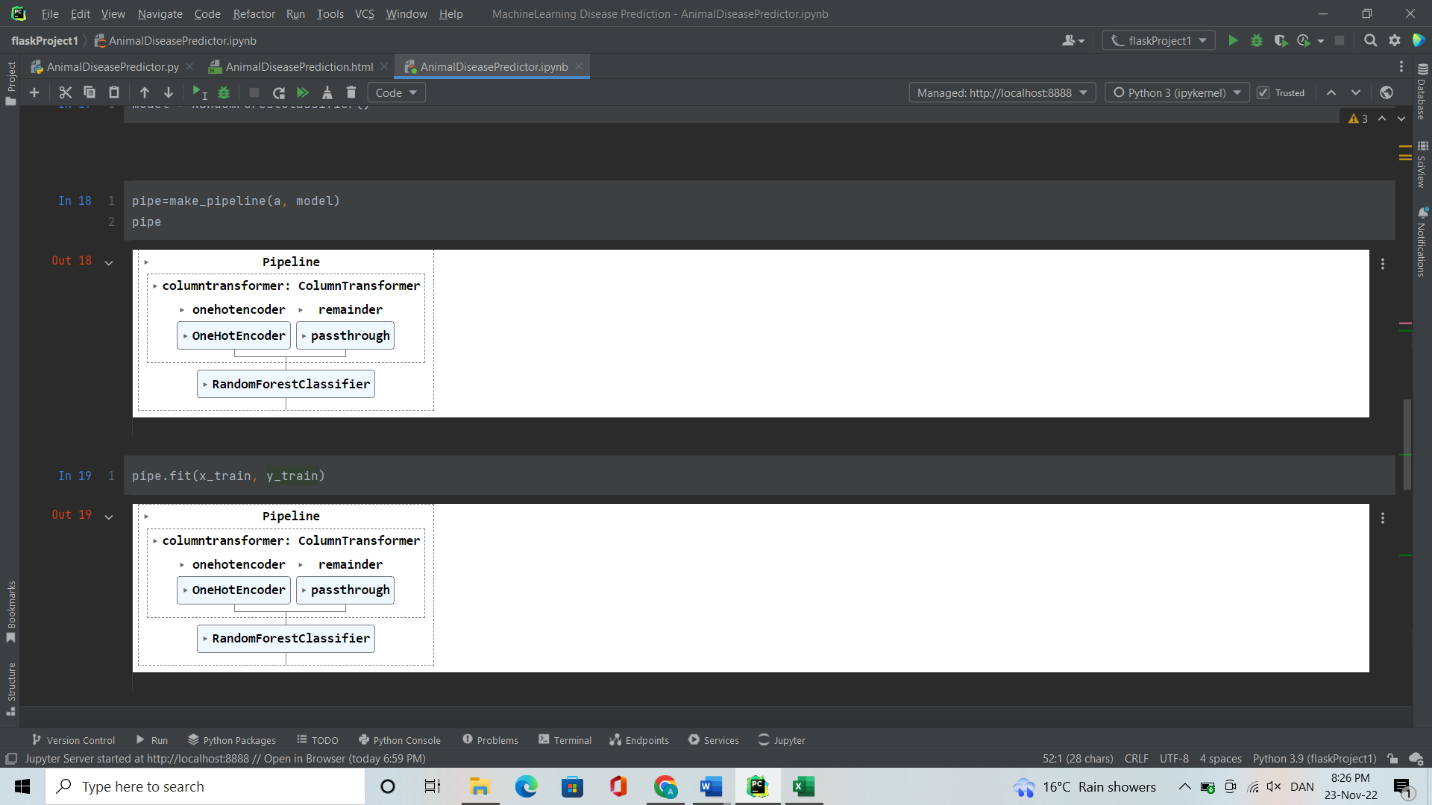
The second was loading the dataset, the Animal\_Disease\_Prediction CSV file.



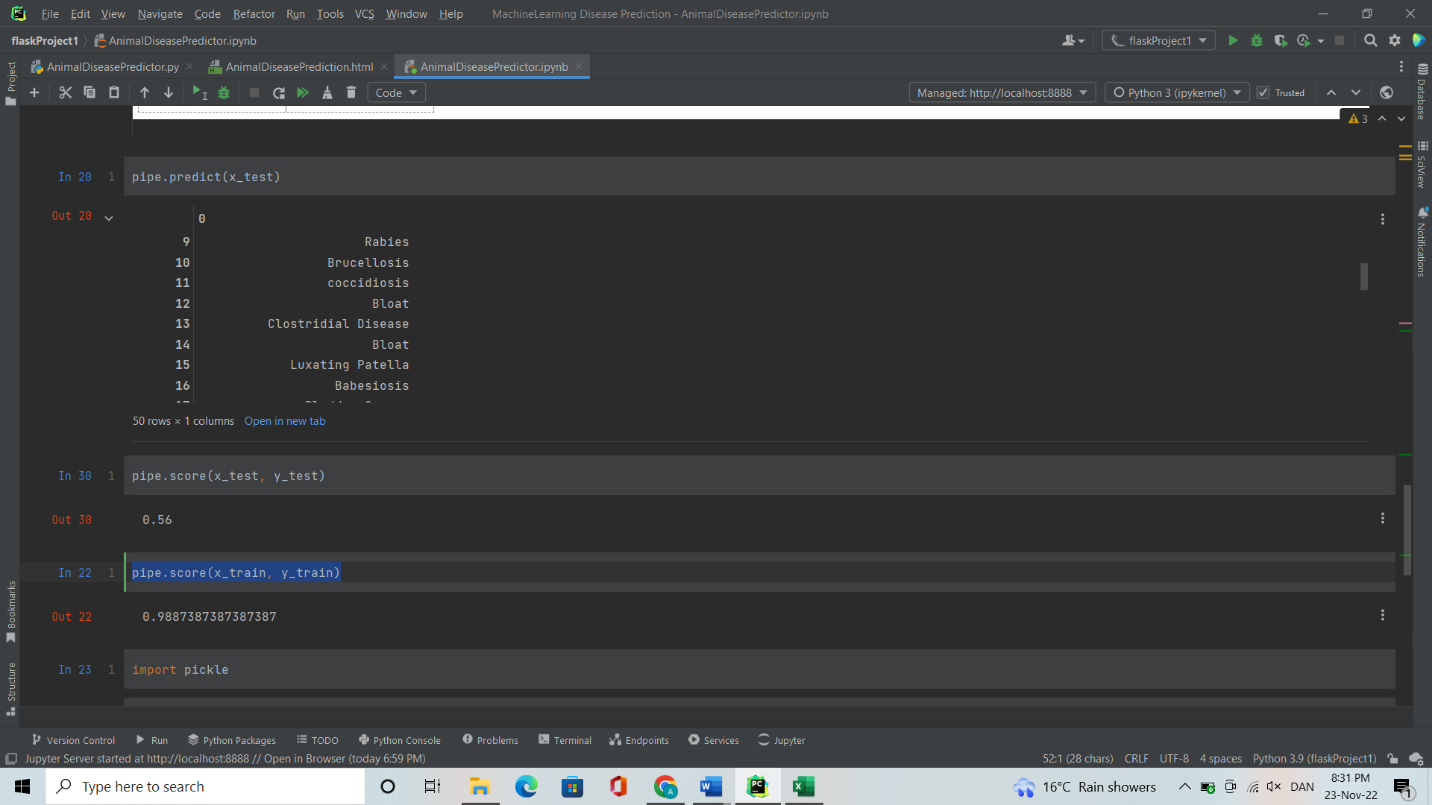
The third and fourth procedure is splitting the dataset and training the dataset, which split the data to training set and testing set, from our data the animal name and the symptoms column are converted to the training set and the last column converted to the test set. This involves training both the Random forest classifier algorithm and the Decision tree classifier algorithm with the animal disease prediction dataset.



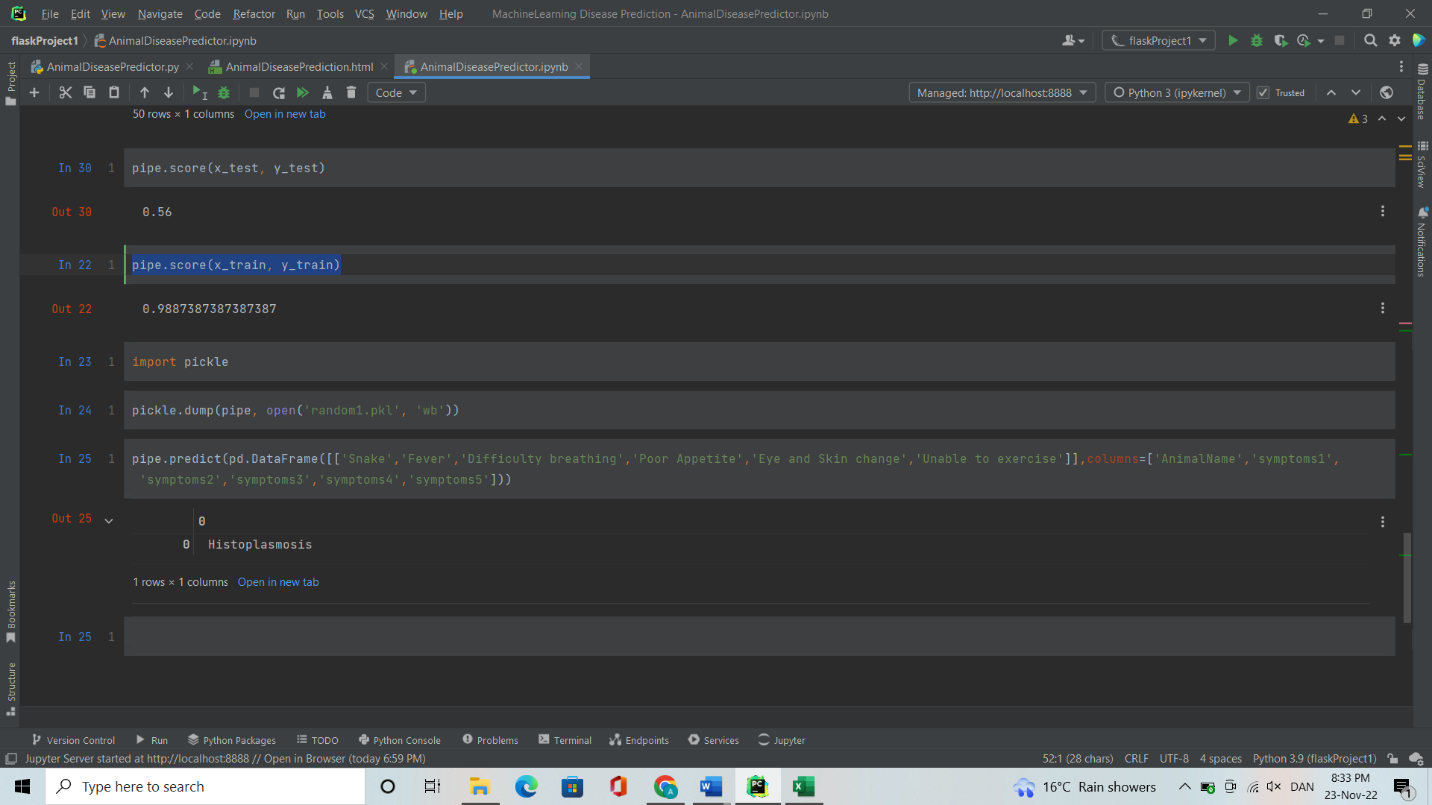




The last procedure was evaluating the model and classification using pipeline score attribute, which uses the predict function, the evaluation of the model brings about 50 different diseases with different phases of symptoms.



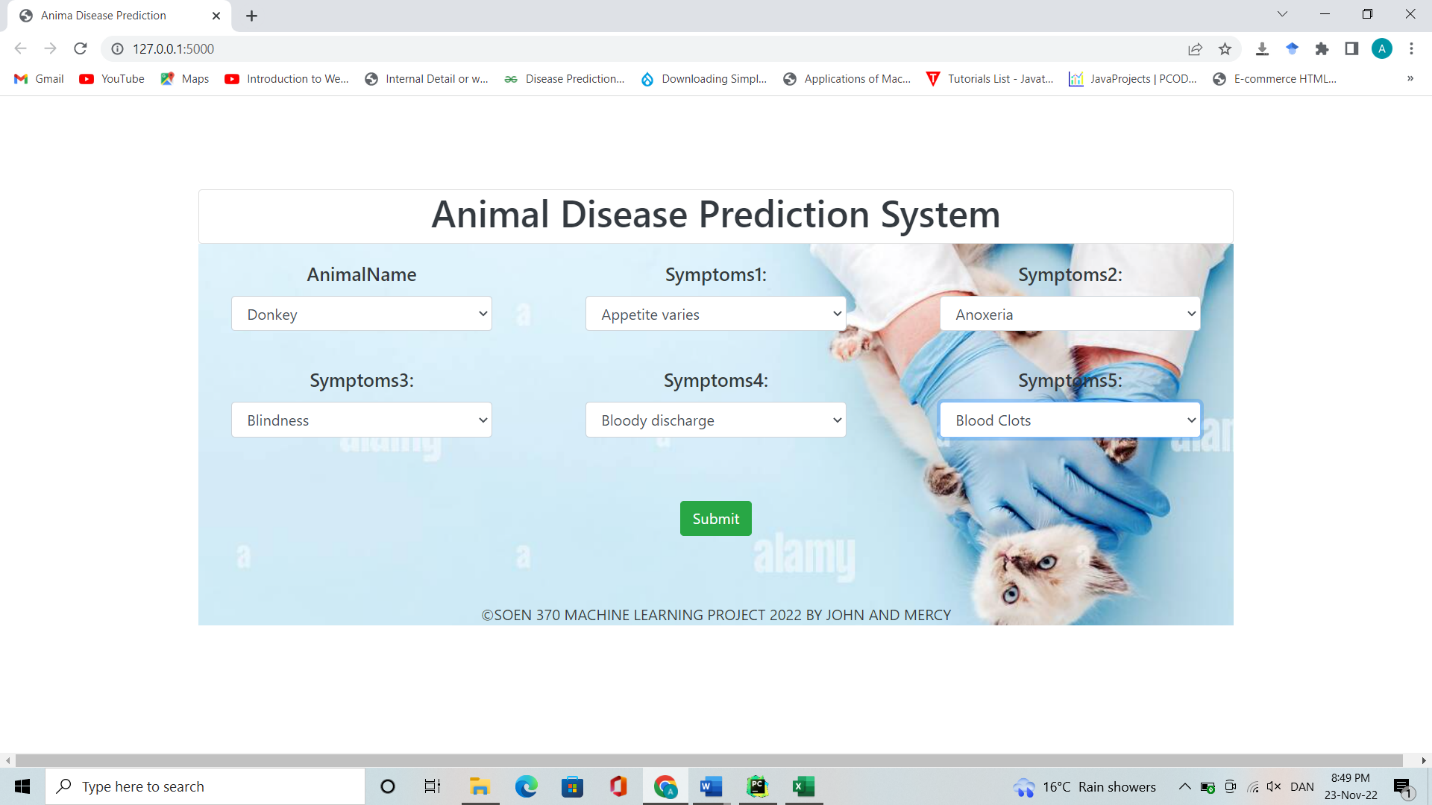
This involves using pipe to check the accuracy and the performance of the algorithms using the pipe score attribute.



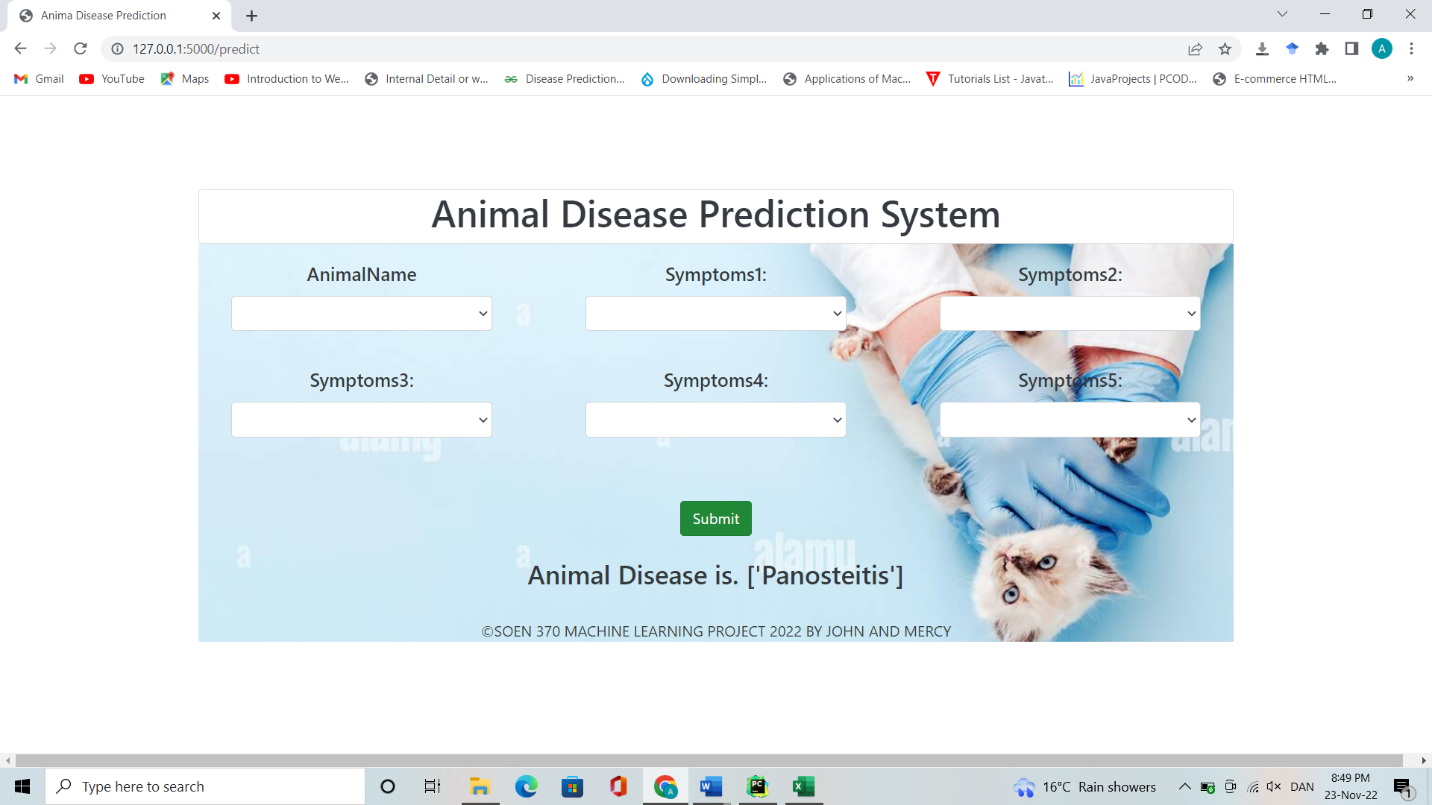
* 1. **Outputs**

After a complete development we made a test of the system by randomly choosing symptoms with different animal names the system successfully displayed the possible animal’s disease from the symptoms.

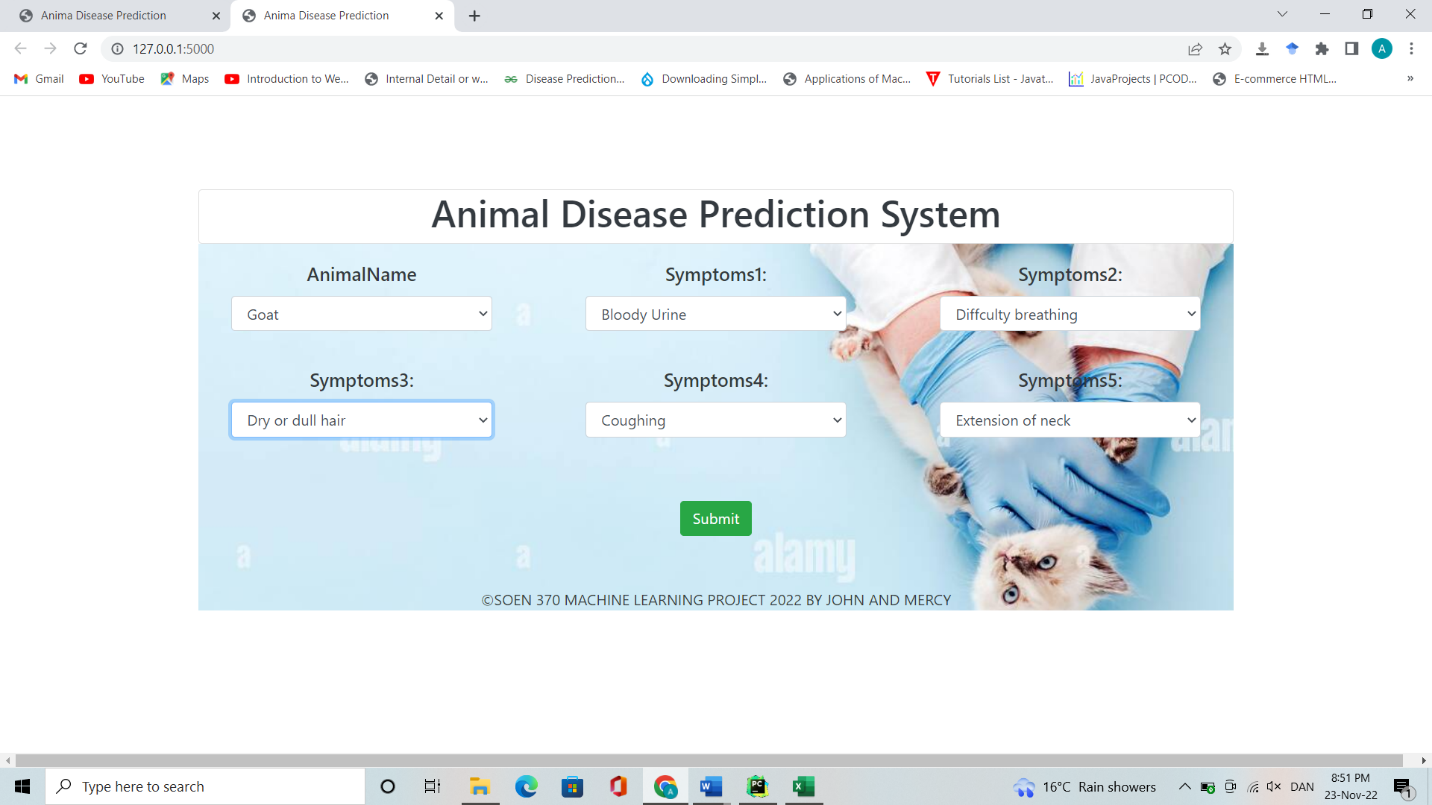
The first sample was a donkey with various symptoms such as

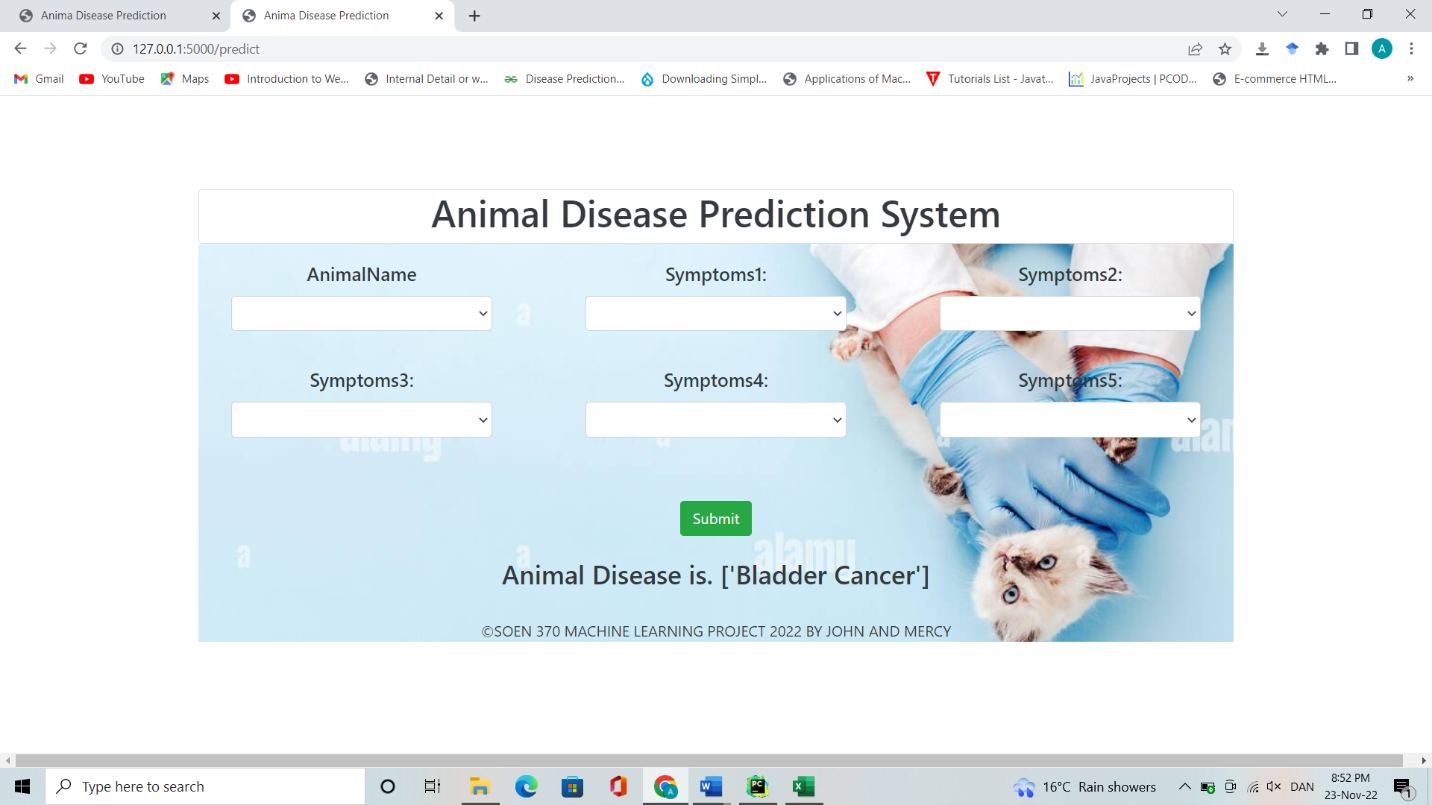


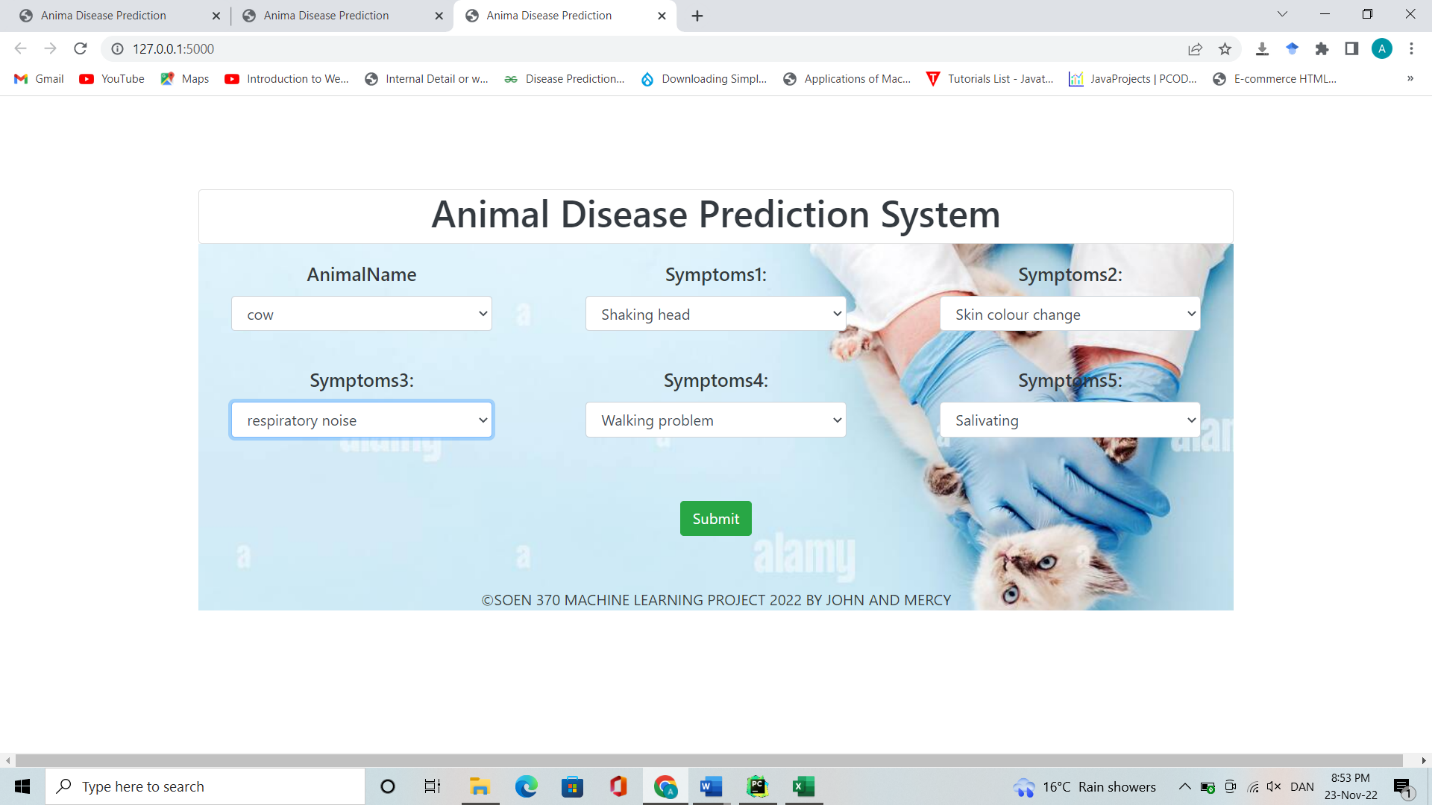
The result of the prediction was panosteitis disease



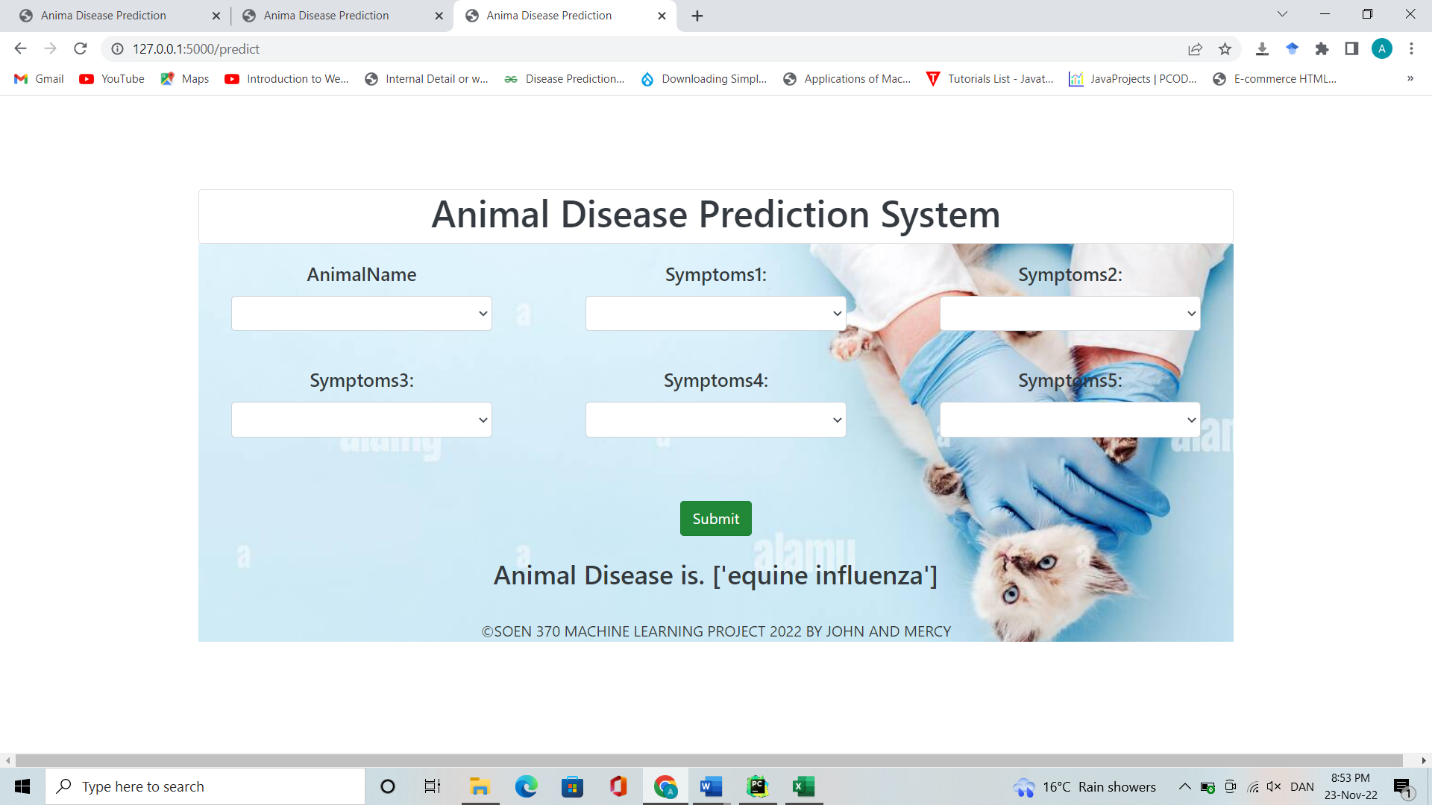
The Second sample was a Goat with various symptoms such as Bloody Urine, Difficulty Breathing, Dry or Dull hair, coughing and Extension of neck

The Possible disease was Bladder Cancer

The last sample of our testing was a cow, the symptoms were, shaking head, skin color change , respiratory noise, walking problem and salivating



The predicted disease was equine influenza which is the most rampant infection in Kenya.



**CHAPTER FIVE**

**5.0 SUMMARY, CONCLUSIONS AND RECOMMENDATIONS**

**5.1 Introduction**

This chapter contains the summary, conclusions and Recommendation from the study findings. The areas to be reviewed and modified to improve the health of animals.

**5.2 Summary**

The study and research on how to improve the health of animals and reducing the occurrence of human-animal contact related diseases through early animal’s disease prediction system was a success. All the research objectives which included: To analyze current developed systems for animal disease prediction, to find out the challenges of the developed animal disease prediction systems, To develop an animal prediction system that accepts symptoms as input to predict animal’s disease and To test and implement the developed animal disease prediction system.

To begin with the first objective and the second objective which is, to analyze current developed systems for animal disease prediction and finding out the challenges of the developed animal disease prediction systems. The analysis of the developed systems was successful. Many of the developed animal prediction systems such as the Empress I launched by the FAO director General Qu Dongyu is limited with the assumptions that the user which are the farmers have the full knowledge of the animal symptoms, we developed a system that will allow the users of the system to choose the disease symptoms from a simple monitoring of the animal behavior. This helps the user with little knowledge of the symptoms to choose the symptoms related to the disease affecting the animal by the carried observation of the animal behavior.

The third objective was to develop an animal prediction system that accepts symptoms as input to predict animal’s disease which was a success, we developed a system that takes the animal symptoms as input, with the python frameworks and dependencies such as the pycharm IDE, flask, and the jupyter notebook which aids in the development of the system. The python packages such as pandas, sklearn and other imported dependencies that were used in the classification of the data and the model. The dataset obtained from the Kaggle.com, contained the data used in the training and predicting the model.

The last objectives is testing and implementation of the developed system. The development of the user interface and the functionality of the algorithm was a success. The developed system works by accepting the symptoms as the user input and predicting the animal’s disease by the analysis of the symptoms. A test of the system was a success, by a sample of the input of the animal’s name and the symptoms which successfully predicts the disease.

**5.3 Conclusions**

The research focused on the development of a system that will help in early disease prediction using the symptoms and signs obtained from observation of the animal behavior and the physical signs. The users of the system which are mostly the livestock farmers would be able to make early prediction and early treatment of the various diseases. The disease control activity would also be of lightweight since early isolation of the infected animals will be of great importance by reducing the spread of the predicted disease.

**5.4 Recommendations**

Animal diseases are many with the limited dataset the few predictions made gives a foundation of a system that can accommodate all the diseases, this will be of a create help to promote the health of animals and reduce the occurrence of human-animal contact related diseases. Therefore, more research is needed to identify more ways of improving the health of animals since We need to prioritize on the animal health sector as said by the FAO director General Qu Dongyu.

**References**

Magwood, S. E. (2010). Preventive veterinary medicine in Canada: study on results of a survey. *The Canadian Veterinary Journal*, *24*(6), 178

Dijkhuizen, A. A., Huirne, R. B., & Jalvingh, A. W. (1995). Economic analysis of animal diseases and their control. *Preventive Veterinary Medicine*, *25*(2), 135-149.

Rabinowitz, P., Scotch, M., & Conti, L. (2009). Human and animal sentinels for shared health risks. *Veterinaria Italiana*, *45*(1), 23.

Kelly, T. R., Machalaba, C., Karesh, W. B., Crook, P. Z., Gilardi, K., Nziza, J., ... & Mazet, J. A. (2020). Implementing One Health approaches to confront emerging and re-emerging zoonotic disease threats: lessons from PREDICT. *One Health Outlook*, *2*(1), 1-7.

Rosa, G. J. (2021). Grand Challenge in Precision Livestock Farming. *Frontiers in Animal Science*, *2*, 650324.

Morsy, S., Dang, T. N., Kamel, M. G., Zayan, A. H., Makram, O. M., Elhady, M., ... & Huy, N. T. (2018). Prediction of Zika-confirmed cases in Brazil and Colombia using Google Trends. *Epidemiology & Infection*, *146*(13), 1625-1627.

Karesh, W. B., Dobson, A., Lloyd-Smith, J. O., Lubroth, J., Dixon, M. A., Bennett, M., ... & Heymann, D. L. (2012). Ecology of zoonoses: natural and unnatural histories. *The Lancet*, *380*(9857), 1936-1945.

Rabinowitz, P. M., Kock, R., Kachani, M., Kunkel, R., Thomas, J., Gilbert, J., ... & Rubin, C. (2013). Toward proof of concept of a one health approach to disease prediction and control. *Emerging Infectious Diseases*, *19*(12).

Neethirajan, S., & Kemp, B. (2021). Digital livestock farming. *Sensing and Bio-Sensing Research*, *32*, 100408.

Peters, D. P., McVey, D. S., Elias, E. H., Pelzel‐McCluskey, A. M., Derner, J. D., Burruss, N. D., ... & Rodriguez, L. L. (2020). Big data–model integration and AI for vector‐borne disease prediction. *Ecosphere*, *11*(6), e03157.

Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, *29*, 100367.

Singh, R. K., & Sharma, V. C. (2015). Ensemble Approach for Zoonotic Disease Prediction Using Machine Learning Techniques.

Smith, P. G., & Bradley, R. (2013). Bovine spongiform encephalopathy (BSE) and its epidemiology. *British medical bulletin*, *66*(1), 185-198.

Kim, H. G., Yang, C. J., & Yoe, H. (2012). Design and implementation of livestock disease forecasting system. *The Journal of Korean Institute of Communications and Information Sciences*, *37*(12), 1263-1270.

McGonigle, P., & Ruggeri, B. (2014). Animal models of human disease: challenges in enabling translation. *Biochemical pharmacology*, *87*(1), 162-171.

Zhang, S., Su, Q., & Chen, Q. (2021). Application of machine learning in animal disease analysis and prediction. *Current Bioinformatics*, *16*(7), 972-982.

Ma, S., Yao, Q., Masuda, T., Higaki, S., Yoshioka, K., Arai, S., ... & Itoh, T. (2021). Development of noncontact body temperature monitoring and prediction system for livestock cattle. *IEEE Sensors Journal*, *21*(7), 9367-9376.

Mazet, J. A., Clifford, D. L., Coppolillo, P. B., Deolalikar, A. B., Erickson, J. D., & Kazwala, R. R. (2009). A “one health” approach to address emerging zoonoses: the HALI project in Tanzania. *PLoS medicine*, *6*(12), e1000190.

Buttke, D. E. (2011). Toxicology, environmental health, and the “One Health” concept. *Journal of Medical Toxicology*, *7*(4), 329-332.

Denayer, T., Stöhr, T., & Van Roy, M. (2014). Animal models in translational medicine: Validation and prediction. *New Horizons in Translational Medicine*, *2*(1), 5-11.

Morota, G., Ventura, R. V., Silva, F. F., Koyama, M., & Fernando, S. C. (2018). Big data analytics and precision animal agriculture symposium: Machine learning and data mining advance predictive big data analysis in precision animal agriculture. *Journal of animal science*, *96*(4), 1540-1550.

Pavan, W., Fraisse, C. W., & Peres, N. A. (2011). Development of a web-based disease forecasting system for strawberries. *Computers and electronics in agriculture*, *75*(1), 169-175.

Mak, S., Klinkenberg, B., Bartlett, K., & Fyfe, M. (2010). Ecological niche modeling of Cryptococcus gattii in British Columbia, Canada. *Environmental Health Perspectives*, *118*(5), 653-658.

Cheng, T. L., Reichard, J. D., Coleman, J. T., Weller, T. J., Thogmartin, W. E., Reichert, B. E., ... & Frick, W. F. (2021). The scope and severity of white‐nose syndrome on hibernating bats in North America. *Conservation Biology*, *35*(5), 1586-1597.

Guan, Y., Chen, H., Li, K. S., Riley, S., Leung, G. M., Webster, R., ... & Yuen, K. Y. (2007). A model to control the epidemic of H5N1 influenza at the source. *BMC Infectious Diseases*, *7*(1), 1-8.

Bischoff, K., Priest, H., & Mount-Long, A. (2010). Animals as sentinels for human lead exposure: a case report. *Journal of Medical Toxicology*, *6*(2), 185-189.

Dowsett, R., & Shannon, M. (1994). Childhood plumbism identified after lead poisoning in household pets. *New England Journal of Medicine*, *331*(24), 1661-1662.

Mutluoglu, M., Uzun, G., Eroglu, M., & Ay, H. (2012). Domestic animals as a warning sign for carbon monoxide poisoning. *Pediatric emergency care*, *28*(6), 596.

Anyamba, A., Chretien, J. P., Small, J., Tucker, C. J., Formenty, P. B., Richardson, J. H., ... & Linthicum, K. J. (2009). Prediction of a Rift Valley fever outbreak. *Proceedings of the National Academy of Sciences*, *106*(3), 955-959.

Jore, S., Viljugrein, H., Brun, E., Heier, B. T., Borck, B., Ethelberg, S., ... & Hofshagen, M. (2010). Trends in Campylobacter incidence in broilers and humans in six European countries, 1997–2007. *Preventive veterinary medicine*, *93*(1), 33-41.

Wålinder, R., Riihimäki, M., Bohlin, S., Hogstedt, C., Nordquist, T., Raine, A., ... & Elfman, L. (2011). Installation of mechanical ventilation in a horse stable: effects on air quality and human and equine airways. *Environmental health and preventive medicine*, *16*(4), 264-272.

Gray, D. J., Williams, G. M., Li, Y., Chen, H., Forsyth, S. J., Li, R. S., ... & McManus, D. P. (2009). A cluster-randomised intervention trial against Schistosoma japonicum in the Peoples' Republic of China: bovine and human transmission. *PLoS one*, *4*(6), e5900.

Benjamin, M., & Yik, S. (2019). Precision livestock farming in swine welfare: a review for swine practitioners. *Animals*, *9*(4), 133.

Gauthier, R., Largouët, C., & Dourmad, J. Y. (2022). Prediction of litter performance in lactating sows using machine learning, for precision livestock farming. *Computers and Electronics in Agriculture*, *196*, 106876.

Neethirajan, S. (2020). The role of sensors, big data and machine learning in modern animal farming. *Sensing and Bio-Sensing Research*, *29*, 100367.

Wang, L. D., Chen, H. G., Guo, J. G., Zeng, X. J., Hong, X. L., Xiong, J. J., ... & Zhou, X. N. (2009). A strategy to control transmission of Schistosoma japonicum in China. *New England Journal of Medicine*, *360*(2), 121-128.

Schelling, E., Bechir, M., Ahmed, M. A., Wyss, K., Randolph, T. F., & Zinsstag, J. (2007). Human and animal vaccination delivery to remote nomadic families, Chad. *Emerging infectious diseases*, *13*(3), 373.

Cleaveland, S., Kaare, M., Tiringa, P., Mlengeya, T., & Barrat, J. (2003). A dog rabies vaccination campaign in rural Africa: impact on the incidence of dog rabies and human dog-bite injuries. *Vaccine*, *21*(17-18), 1965-1973.

Mahmud, M. S., Zahid, A., Das, A. K., Muzammil, M., & Khan, M. U. (2021). A systematic literature review on deep learning applications for precision cattle farming. *Computers and Electronics in Agriculture*, *187*, 106313.

Rowe, E., Dawkins, M. S., & Gebhardt-Henrich, S. G. (2019). A systematic review of precision livestock farming in the poultry sector: Is technology focussed on improving bird welfare?. *Animals*, *9*(9), 614.

Bao, J., & Xie, Q. (2022). Artificial intelligence in animal farming: A systematic literature review. *Journal of Cleaner Production*, *331*, 129956.

Collins, Á. B., Floyd, S., Gordon, S. V., & More, S. J. (2022). Prevalence of Mycobacterium bovis in milk on dairy cattle farms: An international systematic literature review and meta-analysis. *Tuberculosis*, 102166.

Cravero, A., Pardo, S., Sepúlveda, S., & Muñoz, L. (2022). Challenges to Use Machine Learning in Agricultural Big Data: A Systematic Literature Review. *Agronomy*, *12*(3), 748.

Farooq, M. S., Riaz, S., Abid, A., Umer, T., & Zikria, Y. B. (2020). Role of IoT technology in agriculture: A systematic literature review. *Electronics*, *9*(2), 319.

Jeong, S., Jeong, H., Kim, H., & Yoe, H. (2013). Cloud computing based livestock monitoring and disease forecasting system. *International Journal of Smart Home*, *7*(6), 313-320.

Oberhelman, R. A., Gilman, R. H., Sheen, P., Cordova, J., Zimic, M., Cabrera, L., ... & Perez, J. (2006). An intervention-control study of corralling of free-ranging chickens to control Campylobacter infections among children in a Peruvian periurban shantytown. *The American journal of tropical medicine and hygiene*, *74*(6), 1054-1059.