

ANIMAL MONITORING BASED ON MACHINE LEARNING

Submitted in partial fulfillment of the requirements for the award of Bachelor of
Engineering / Technology degree in
ELECTRONICS AND COMMUNICATION ENGINEERING

By

THIYANESHWARAN R (37130409)

STENISON J (37130397)



**DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING
SCHOOL OF ELECTRICAL AND ELECTRONICS ENGINEERING**

SATHYABAMA

INSTITUTE OF SCIENCE AND TECHNOLOGY

**(DEEMED TO BE UNIVERSITY) JEPPIAAR NAGAR, RAJIV
GANDHI SALAI, CHENNAI - 600 119**

Accredited with Grade "A" by NAAC

March 2021



SATHYABAMA

**INSTITUTE OF SCIENCE AND TECHNOLOGY
(DEEMED TO BE UNIVERSITY)**



Accredited with Grade “A” by NAAC | 12B Status by UGC | Approved by AICTE

Jeppiaar Nagar, Rajiv Gandhi Salai, Chennai-600 119

www.sathyabama.ac.in

DEPARTMENT OF ELECTRONICS AND COMMUNICATION ENGINEERING

BONAFIDE CERTIFICATE

This is to certify that this Project Report is the bonafide work of **THIYANESHWARAN R (37130409), STENISON J(37130397)** who carried out the project entitled “ **Animal Monitoring based on Machine Learning** ” under our supervision from **AUG 2020** to **MARCH 2021**.

Internal Guide

Dr. R.NARMADHA, M.E., Ph.D.

Head of the Department

Dr. T. RAVI, M.E., Ph.D.

■

Submitted for Viva voce Examination held on _____

Internal Examiner

External Examiner

DECLARATION

We **THIYANESHWARAN R (37130409), STENISON J (37130397)** hereby declare that the Project Report entitled “**Animal Monitoring based on Machine Learning** ” done by us under the guidance of **Dr. R.NARMADHA** at **SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY** is submitted In partial fulfilment of the requirements for the award of Bachelor Of Engineering degree in Electronics and Communication Engineering.

DATE: (1)

PLACE: (2)

SIGNATURE OF THE CANDIDATE

ACKNOWLEDGEMENT

We are pleased to acknowledge our sincere thanks to the Board of Management of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. We are grateful to them.

I convey my thanks to **Dr. N. M. NANDHITHA, M.E., Ph.D.**, Prof. & Dean, School of Electrical and Electronics Engineering and **Dr. T. RAVI, M.E., Ph.D.**, Head of the Department, Dept. of Electronics and Communication Engineering for providing me necessary support and details at the right time during the progressive reviews.

I would like to express my sincere and deep sense of gratitude to my Project Guide **Dr. R.NARMADHA, M.E., Ph.D.**, for her valuable guidance, suggestions and constant encouragement paved the way for the successful completion of my project work.

I wish to express my thanks to all Teaching and Non-teaching staff members of the Department of **Electronics and Communication Engineering** who were helpful in many ways for the completion of the project.

ABSTRACT

In this paper, we approach a new method of Machine learning Algorithm Decision Tree for the prediction of wild animals such as elephants in where there is a density of high population and cultivation. Safety and human cost is a most valuable thing. Risking human lives is not acceptable. So, it is a need to find an outcome for reducing the mortality of lives and cultivation. The proposed system overcomes the existing animal classification methods and gives us a broad view of the animals. The Basics of Machine Learning gives us a broad view of input data which is given and helps us to process the data for future purposes. The predicted data can be viewed by the user in the shell of python. The feature of the project is approved by software implementation. Selecting the required variables to classify the animals. Building a model with efficient runtime, generating required metrics, and accuracy. Creating and assigning the data for the classification of animals. Using Sklearn library to implement the decision tree algorithm.

TABLE OF CONTENTS

S.NO	DESCRIPTION	PAGE NO
Chapter 1	INTRODUCTION	1
	1.1 INTRODUCTION TO AI	2
	1.2 IMPORTANCE TO AI	3
	1.3 MACHINE LEARNING	3
	1.3.1 MACHINE LEARNING APPROACHES	3
	1.3.1.1 UNSUPERVISED LEARNING	3
	1.3.1.2 REINFORCEMENT LEARNING	3
	1.3.1.3 SUPERVISED LEARNING	4
	1.3.1.4 UNSUPERVISED LEARNING	5
	1.3.1.5 REINFORCEMENT LEARNING	6
	1.3.2 DECISION TREE	6
	1.3.2.1 THE OBJECTIVES OF DECISION TREE	6
	1.3.2.2 ADVANTAGES OF DECISION TREE	7
	1.4 SOFTWARE SPECIFICATION	7
	1.4.1 PYTHON	8
	1.4.2. EXCEL	8
Chapter 2	LITERATURE SURVEY	10
	2.1 INTRODUCTION	10
	2.1.1 MULTIMODAL BLENDING FOR	10
	2.1.2 ANIMAL RECOGNITION SYSTEM CNN	11
	2.1.3 IDENTIFICATION OF ANIMAL DEEP CNN	11
	2.1.4 OBJECT DETECTION TRAINED MODELS	12
	2.2 LITERATURE SUMMARY	13
Chapter 3	AIM AND SCOPE OF PRESENT INVESTIGATION	14
	3.1 SEVEN STEPS	14
	3.1.1 DEFINE THE TARGET OF THE MATTER	14

3.1.2 DATA GATHERING	14
3.1.3 DATA PREPARATION	15
3.1.4 EXPLORATIVE DATA ANALYSIS	15
3.1.5 BUILDING A MODEL	15
3.1.5.1 3 STAGES IN BUILDING THE MODEL	15
3.1.6 EVALUATING THE MODEL	15
3.1.7 PREDICTIONS	16
3.2 SOFTWARE DESCRIPTION PYTHON	16
3.2.1 PYTHON 2 VS. PYTHON 3	18
3.2.2 PYTHON APPLICATIONS	19
3.2.2.1 INTERNET APPLICATIONS	19
3.2.2.2 DESKTOP GRAPHICAL USER INTERFACE	19
APPLICATIONS	19
3.2.2.3 SOFTWARE PACKAGE DEVELOPMENT	19
3.2.2.4 Scientific And Numeric	19
3.2.2.5 BUSINESS APPLICATIONS	19
3.2.2.6 CONSOLE PRIMARILY BASED APPLICATION	20
3.2.2.7 AUDIO OR VIDEO PRIMARILY BASED	20
APPLICATIONS	20
3.2.2.8 3D CAD APPLICATIONS	20
3.2.2.9 ENTERPRISE APPLICATIONS	20
3.2.2.10 APPLICATIONS FOR PICTURES	20
	20
Chapter 4 MATERIALS AND METHODS	21
4.1 DECISION TREE ALGORITHM	21
4.2 LOGISTIC REGRESSION	21
4.3 SVM	22
4.4 RANDOM FOREST	23
4.5 DATASET	24
4.6 TRAINING AND SPLITTING	25
4.7 DATAFLOW / DATA VISUALIZATION	25
4.8 PRECISION / ACCURACY CHECK	27

Chapter 5	RESULT AND DISCUSSION	28
	5.1 FEATURES OF ANIMAL SPECIFICATION	28
	5.2 PREDICTED DATA ON SHELL	29
Chapter 6	SUMMARY AND CONCLUSION	30
	REFERENCES	31
	APPENDICES - SOURCE CODE	32

TABLE OF FIGURES

S.NO	DESCRIPTION	PAGE NO
1.3.1.3	Supported Vector Machine	4
1.3.2	Decision Tree	8
2.2.1	Proposed System	13
3.2	Python	16
4.1	Decision tree algorithm	21
4.2	LOGISTIC REGRESSION	22
4.4	Random Forest	24
4.5	CSV Characteristics	25
4.2.1 - 5	Dataset	26
4.2.7	Heatmap	26
4.3	Accuracy	27
5.1	Features of Animal Specification	28
5.2	Predicted Data on shell	29

CHAPTER 1

INTRODUCTION

Machine learning is a subsystem of artificial intelligence that makes systems involuntarily learn and progress from experience without being programmed. Machine learning focuses on the growth of computer programs that can access data and utilize it to learn for themselves. The learning process starts with interpretation of data, like examples, previous model, or suggestions, in order to take improved decisions in the future. The principal aim is to permit the computer's train themselves without human intrusion or assistance and corrects its mistakes themselves through this learning.

Classification of animals is a test that each animal needs to confront, it's acceptable to realize that there are a few data science procedures that can forestall this sort of misfortune. For such predictions we have used decision Tree to classify the animals in an advance manner.

1.1 INTRODUCTION TO AI

An Artificial intelligence (AI), is the intelligence that is illustrated by machines, different from the natural intelligence displayed by humans and animals. Most popular AI textbooks define the scope as the study of "intelligent agents": any device that perceives its environment and takes actions that increases its chance of successfully gaining its goals. Similarly, the term "artificial intelligence" is often used to refer to the machines (or computers) that mimic "cognitive" functions that humans associate with the human mind, such as "learning" and "problem-solving".

I have taken all the data sets except a type of animal as the independent variable. Based on that we have predicted the outcome of charges by using linear regression and random forest respectively.

1.2 IMPORTANCE TO AI :

The importance of AI and its future elements are acknowledged for quite while currently. they're being looked upon as tools and techniques to form this world a stronger place. And it's simply not that you simply got to head to these fancy school gadgets to be able to use them. you'll be able to merely go searching and that i am just about positive most of your tasks area unit created sleek by the utilization of AI.

Its importance lies in creating our lives easier. These technologies area unit a good plus to humans and so area unit programmed to cut back human effort the maximum amount as doable. they have an inclination to possess the aptitude of operating in an automatic fashion and so manual intervention is that the last item that might be asked for or seen whereas operational elements related to this technology.

These machines tend to hurry up your tasks and processes along side a secure level of exactness and accuracy and so this can be what makes them a helpful and necessary tool. with the exception of creating the planet associate error-free place with their easy and everyday techniques, these technologies and applications aren't solely associated with our general and everyday lives. it's conjointly impacting and holds importance for alternative domains in addition.

1.3 MACHINE LEARNING :

Machine learning (ML) is that the study of pc algorithms that improve mechanically through expertise. it's seen as a set of AI. Machine learning algorithms build a model supported sample information, referred to as "training data", to form predictions or choices while not being expressly programmed to try to to therefore.

Machine learning algorithms area unit employed in a good sort of applications, like email filtering and pc vision, wherever it's tough or impossible to develop typical algorithms to perform the required tasks.

A set of machine learning is closely associated with machine statistics, that focuses on creating predictions victimization computers; however not all machine learning is applied math learning. The study of mathematical optimisation delivers ways, theory, and application domains to the sector of machine learning. data processing could be a connected field of study, that specialize in alpha information analysis through unsupervised learning. In its application across business issues, machine learning is additionally mentioned as prognosticative analytics.

1.3.1 MACHINE LEARNING APPROACHES :

Machine learning approaches are historically divided into 3 broad classes, reckoning on the character of the "signal" or "feedback" accessible to the training system:

Supervised learning: the pc is conferred with example inputs and their desired outputs, given by a "teacher", and therefore the goal is to be told a general rule that maps inputs to outputs.

1.3.1.1 UNSUPERVISED LEARNING: No labels are given to the training algorithmic rule, going it on its own to seek out structure in its input. unattended learning is a goal in itself (discovering hidden patterns in data) or a way towards associate finish (feature learning).

1.3.1.2 REINFORCEMENT LEARNING: A malicious program interacts with a dynamic atmosphere during which it should perform a definite goal (such as driving a vehicle or taking part in a game against associate opponent). because it navigates its downside area, the program is provided feedback that is analogous to rewards, that it tries to maximise.

Other approaches are developed that do not work showing neatness into this three-fold categorization, and typically quite one is employed by an equivalent machine learning system. for instance topic modelling, spatiality reduction, or meta-learning.

As of 2020, deep learning has become the dominant approach for abundant in progress add the sector of machine learning.

1.3.1.3 SUPERVISED LEARNING:

Supervised learning algorithms build a mathematical model of a group of knowledge that contains each the inputs and therefore the desired outputs. knowledge|the info|the information} is understood as coaching data and consists of a group of coaching examples. every coaching example has one or a lot of inputs and therefore the desired output, conjointly referred to as a superior signal. within the mathematical model, every coaching example is pictured by associate array or vector, typically known as a feature vector, and therefore the coaching information is pictured by a matrix. Through repetitious optimisation of associate objective perform, supervised learning algorithms learn a perform which will be wont to predict the output related to new inputs. associate best perform can enable the algorithmic rule to properly confirm the output for inputs that weren't a region of the coaching information. associate algorithmic rule that improves the accuracy of its outputs or predictions over time is alleged to possess learned to perform that task.

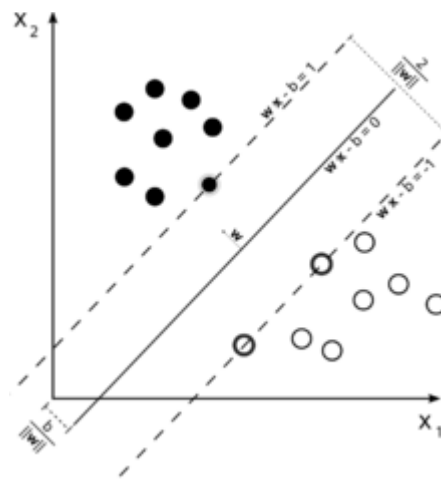


Fig 1.3.1.3: Supported Vector Machine

Types of supervised learning algorithms embrace active learning, classification, and regression.

Classification algorithms are used once the outputs are restricted to a restricted

set of values, and regression algorithms are used once the outputs

could have any numerical worth inside a spread. As an example, for a classification rule that filters emails, the input would be an incoming email, and also the output would be the name of the folder within which to file the e-mail.

Similarity learning is a vicinity of supervised machine learning closely associated with regression and classification, however the goal is to find out from examples employing a similarity perform that measures however similar or connected 2 objects are. it's applications within the ranking, recommendation systems, visual identity pursuit, face verification, and speaker verification.

1.3.1.4 UNSUPERVISED LEARNING:

Unsupervised learning algorithms take a collection of knowledge that contains solely inputs, and spectrum line within the information, like grouping or cluster of knowledge points. The algorithms, therefore, learn from check information that has not been tagged, classified, or categorised. rather than responding to feedback, unsupervised learning algorithms determine commonalities within the information and react supported the presence or absence of such commonalities in every new piece of knowledge. A central application of unsupervised learning is within the field of density estimation in statistics, like finding the likelihood density perform. although unsupervised learning encompasses alternative domains involving summarizing and explaining information options.

Cluster analysis is that the assignment of a collection of observations into subsets (called clusters) so observations inside constant cluster are similar per one or a lot of predesigned criteria, whereas observations drawn from completely different clusters are dissimilar. {different|totally completely different|completely different} cluster techniques create different assumptions on the structure of the information, usually outlined by some similarity metric and evaluated, as an example, by internal compactness, or the similarity between members of constant cluster, and separation, the distinction between clusters. alternative strategies are supported

calculable density and graph property.

1.3.1.5 REINFORCEMENT LEARNING:

Reinforcement learning is a region of machine learning involved with however computer code agents have to be compelled to take actions in associate degree setting to maximise some notion of additive reward. thanks to its generality, the sphere is studied in several alternative disciplines, like scientific theory, management theory, research, scientific theory, simulation-based improvement, multi-agent systems, swarm intelligence, statistics, and genetic algorithms. In machine learning, the setting is often described as Markov Decision Process (MDP). several reinforcement learning algorithms use dynamic programming techniques. Reinforcement learning algorithms don't assume data of precise|a particular|a precise|a certain|a definite|an explicit} mathematical model of the MDP and are used once exact models are impracticable. Reinforcement learning algorithms are employed in autonomous vehicles or in learning to play a game against a personality's opponent.

1.3.2 DECISION TREE

Decision Trees are a non-parametric supervised learning technique used for each classification and regression tasks. Non-parametric means it doesn't build assumptions, so we are able to acquire smart prediction results which are accurate.

1.3.2.1 THE OBJECTIVES OF DECISION TREE

The goal is to form a model that predicts the worth of a target variable by learning straightforward call rules inferred from the info options. the decision rules are usually in sort of if-else statements. The deeper the tree, the additional complicated the foundations and fitter the model.

A decision tree may be a graphical illustration of all the attainable solutions to a call. These choices are supported by some conditions and might be simply explained. A decision tree may be a tree-like graph with nodes representing the place wherever we tend to decide associate attribute and raise a question; edges represent the answers to the question, and also the leaves represent the particular output or category label. they're utilized in non-linear call-making with a straightforward linear surface.

1.3.2.2 ADVANTAGES OF DECISION TREE

- Most correct learning algorithms
- Runs expeditiously on massive databases
- Simple to know, interpret, visualize.
- Decision trees implicitly perform variable screening or feature choice.
- Can handle each numerical and categorical data. can even handle multi-output issues.
- Decision trees need comparatively very little effort from users for knowledge preparation.
- Nonlinear relationships between parameters don't have an effect on tree performance
- Methods for equalization error in unbalanced knowledge sets

DISADVANTAGES OF DECISION TREE

- Decision-tree learners will produce over-complex trees that don't generalize the info well. this can be known as overfitting. call trees will be unstable as a result of tiny variations within the knowledge may lead to a totally completely different tree being generated. this can be known as variance, that must be down

by ways like sacking and boosting.

- Greedy algorithms cannot guarantee to come the globally best call tree. this will be slaked by coaching multiple trees, wherever the options and samples square measure every which way sampled with replacement. call tree learners produce biased trees if some categories dominate. it's thus suggested to balance the info set before fitting with the decision tree.

Building Decision Tree

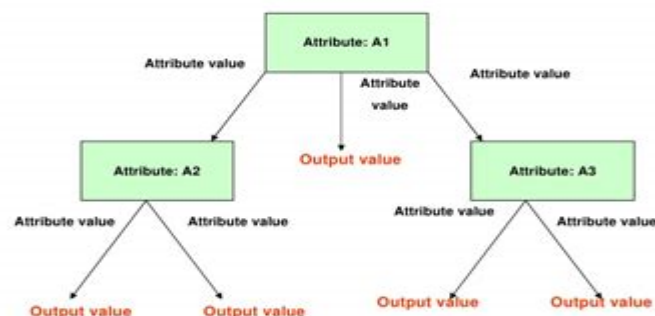


Fig 1.3.2: Decision Tree

1.4 SOFTWARE SPECIFICATION

Following are the software used:

1.4.1 PYTHON: Python has been the go-to language for programmers who does data analytics because it is easy to use, easy to understand and it has numerous statistical and mathematical libraries.

1.4.2. EXCEL: Excel is the most used spreadsheet software ever since it is developed by Microsoft. It stores the data in rows and columns and is used to perform data manipulation using mathematical functions.

Python packages used:

1.4.1.1 NUMPY: Numpy is a library for Python, adding support for large, multi-dimensional arrays and matrices, besides that it has a large collection of high-level mathematical functions such as mean, median, etc., to operate on these arrays.

1.4.1.2. PANDAS: Pandas is an open-source library file for python to function data analysis and manipulate data. It is used to scan the collected datasets and turn them into data frames.

1.4.1.3. NLTK(NATURAL LANGUAGE TOOLKIT): is a suite of libraries and programs for symbolic and statistical natural language processing for English written in the Python programming language.

1.4.1.4. SKLEARN(SCIKIT-LEARN): is a machine learning library for Python. It has many algorithms like random forests, support vector machine, and k-means, and it also supports Python libraries like NumPy and SciPy.

1.4.1.5. CSV: The CSV type file intricately determines the classes to scan and print tabular data with file format naming CSV i.e., to rename the data in the format preferably given by Excel or to scan the data that is produced by excel.

1.4.1.6. RANDOM FOREST: A random forest is an algorithm used for classification. It sub-samples the data and sets decision trees for each. Then voting is done among the trees and then classifies the data accordingly. It uses averaging to increase accuracy and reduce over-fitting.

1.4.1.7. MATPLOTLIB: Matplotlib is a plotting library for python. It is used to plot different types of graphs.

CHAPTER 2

LITERATURE SURVEY

2.1 INTRODUCTION

Many literatures have been carefully examined, and numerous findings have been analyzed. These findings aided us in effectively applying the process. Many facts must be understood in order to draw conclusions, and these conclusions are based on those facts.

2.1.1 MULTIMODAL BLENDING FOR HIGH-ACCURACY INSTANCE RECOGNITION.

Despite the rich information provided by sensors such as the Microsoft Kinect in the robotic perception setting, the problem of detecting object instances remains unsolved, even in the tabletop setting, where segmentation is greatly simplified. Existing object detection systems often focus on textured objects, for which local feature descriptors can be used to reliably obtain correspondences between different views of the same object. We examine the benefits of dense feature extraction and multimodal features for improving the accuracy and robustness of an instance recognition system. By combining multiple modalities and blending their scores through an ensemble based method in order to generate our final object hypotheses, we obtain significant improvements over previously published results on two RGB-D datasets. On the Challenge dataset, our method results in only one missed detection (achieving 100% precision and 99.77% recall). On the Willow dataset, we also make significant gains on the prior state of the art (achieving 98.28% precision and 87.78% recall), resulting in an increase in F-score from 0.8092 to 0.9273

2.1.2 ANIMAL RECOGNITION SYSTEM BASED ON CONVOLUTIONAL NEURAL NETWORK.

In this paper, the performances of well-known image recognition methods such as Principal Component Analysis (PCA), Linear Discriminant Analysis (LDA), Local Binary Patterns Histograms (LBPH) and Support Vector Machine (SVM) are tested and compared with proposed convolutional neural network (CNN) for the recognition rate of the input animal images. In our experiments, the overall recognition accuracy of PCA, LDA, LBPH and SVM is demonstrated. Next, the time execution for animal recognition process is evaluated. The all experimental results on created animal database were conducted. This created animal database consist of 500 different subjects (5 classes/ 100 images for each class). The experimental result shows that the PCA features provide better results as LDA and LBPH for large training set. On the other hand, LBPH is better than PCA and LDA for small training data set. For proposed CNN we have obtained a recognition accuracy of 98%. The proposed method based on CNN outperforms the state of the art methods.

2.1.3 IDENTIFICATION OF ANIMAL SPECIES IN CAMERA-TRAP IMAGES USING VERY DEEP CONVOLUTIONAL NEURAL NETWORKS

In this paper, non intrusive monitoring of animals in the wild is possible using camera trapping framework, which uses cameras triggered by sensors to take a burst of images of animals in their habitat. However camera trapping framework produces a high volume of data (in the order on thousands or millions of images), which must be analyzed by a human expert. In this work, a method for animal species identification in the wild using very deep convolutional neural networks is presented. Multiple versions of the Snapshot Serengeti dataset were used in order to probe the ability of the method to cope with different challenges that camera-trap images demand. The method reached 88.9% of accuracy in Top-1 and 98.1% in Top-5 in the evaluation set using a residual network topology. Also, the results show that the proposed method outperforms previous approximations and proves that recognition in camera-trap images can be automated.

2.1.4 OBJECT DETECTION WITH DISCRIMINATIVELY TRAINED PART-BASED MODELS.

We describe an object detection system based on mixtures of multiscale deformable part models. Our system is able to represent highly variable object classes and achieves state-of-the-art results in the PASCAL object detection challenges. While deformable part models have become quite popular, their value had not been demonstrated on difficult benchmarks such as the PASCAL datasets. Our system relies on new methods for discriminative training with partially labeled data. We combine a margin sensitive approach for data-mining hard negative examples with a formalism we call latent SVM. A latent SVM is a reformulation of MI-SVM in terms of latent variables. A latent SVM is semi-convex and the training problem becomes convex once latent information is specified for the positive examples. This leads to an iterative training algorithm that alternates between fixing latent values for positive examples and optimizing the latent SVM objective function.

2.2 SUMMARY OF LITERATURE REVIEW

The proposed system overcomes the existing methods and replaces it by best classification algorithms. The Feature Extraction gives us a broad view about the input data which is given and help us to process the data for pre-processing. The given system has overcomes the errors and has higher efficiency than the current classifying algorithms. By the usage algorithms the animal is predicted and it can be viewed by the user in the shell of python. The FIG 5.2 determines the flow chart of decision tree manipulation.

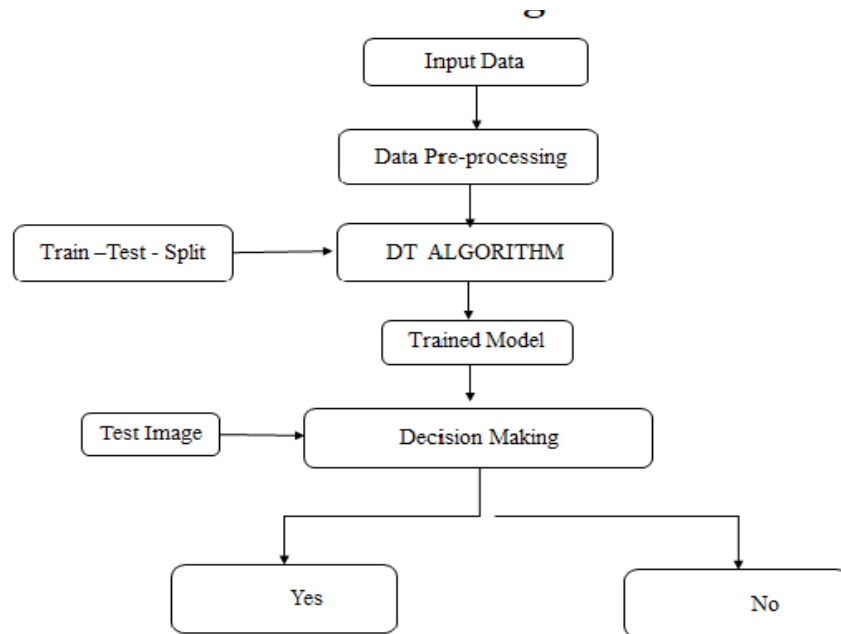


Fig 2.2.1 Proposed System

acoustic monitoring (PAM). In collaboration with conservation efforts, we construct a large labelled dataset of passive acoustic recordings of the African Forest Elephant via crowdsourcing, comprising thousands of hours of recordings in the wild. Using state-of-the-art techniques in AI

CHAPTER 3

AIM AND SCOPE OF PRESENT INVESTIGATION

In Machine learning, seven steps are the essential methodology to predict the result. Before selecting the formula we tend to should perform the steps mentioned as follows,

3.1 SEVEN STEPS:

- Define the target Of the matter
- Data Gathering
- Data Preparation
- Exploratory information Analysis
- Building A Model
- Evaluating The Model
- Predictions

3.1.1 DEFINE THE TARGET OF THE MATTER :

The objective performance is one among the foremost basic elements of a machine learning system, therein it provides the essential, formal specification of the matter. for a few objectives, the best parameters may be found specifically known because of the analytic solution.

3.1.2 DATA GATHERING:

Data gathering is that the method of gathering and measurement data from innumerable completely different sources. To use the information we collect to develop sensible computing (AI) and machine learning solutions, it should be collected and hold in an exceeding method that produces sense for the business downside at hand.

3.1.3 DATA PREPARATION:

Data preparation (also stated as "data pre-processing") is that the method of remodelling {raw information|data|information} so data scientists and analysts will run it through machine learning algorithms to uncover insights or create predictions.

3.1.4 EXPLORATIVE DATA ANALYSIS:

Exploratory Data Analysis (EDA) is Associate in Nursing approach in analysing datasets to summarize their main characteristics, typically with visual strategies. EDA is employed for seeing what the information will tell us before the modelling task. explorative information Associate in Nursing analysis techniques are devised as an aid during this state of affairs.

3.1.5 BUILDING A MODEL:

Machine learning consists of algorithms which will modify analytical model building. victimisation algorithms that iteratively learn from data, machine learning models facilitate computers to search out hidden insights from huge data while not being expressly programmed wherever to appear.

3.1.5.1 3 STAGES IN BUILDING THE MODEL:

- Model Building,
- Model Testing
- Applying Model

3.1.6 EVALUATING THE MODEL:

The model analysis aims to estimate the generalization accuracy of a model on future (unseen/out-of-sample) data. Strategies for evaluating a model's performance are divided into two categories: particularly,

- Holdout
- Cross-Validation.

Both strategies use a take a look at set (i.e data not seen by the model) to judge model performance.

models in machine learning are,

- supervised,
- unsupervised,
- reinforcement learning.

3.1.7 PREDICTIONS:

Prediction suggests that Machine Learning is the output of a function or an algorithm rule once it's been trained on a historical data set and applied to new data once stated the chance of a selected outcome, like whether or not a client can churn in thirty days.

3.2 SOFTWARE DESCRIPTION PYTHON :



Fig 3.2 Python

Python may be a generally taken, interactive, object-oriented, and high-level programming language. It had been created by Guido van Rossum and run through 1985- 1990. Like Perl, Python ASCII text file is additionally obtainable underneath the antelope General Public License (GPL). This tutorial offers enough understanding on Python programming language Python may be a high-level, taken, interactive and object-oriented scripting language. Python is meant to be extremely decipherable. It uses English keywords often wherever as alternative languages use punctuation, and it's fewer syntactic constructions than alternative

languages.

Python may be an essential one for college students operating and dealing with professionals to become an excellent coder specially once they are working in net Development Domain. It will be able to list down a number of the key blessings of learning Python: Python is taken taken Python is processed at runtime by the interpreter. You do not have to be compelled to compile your program before running it. this can be like PERL and PHP. Python is Interactive – you'll truly sit at a Python prompt and act with the interpreter on to write your programs. Python is Object-Oriented – Python supports Object-Oriented vogue or technique of programming that encapsulates code at intervals objects. Python may be a Beginner's Language – Python may be a nice language for the beginner-level programmers and supports the event of a large vary of applications from straightforward text process to WWW browsers to games

Python may be a general purpose, dynamic, high-level, and taken programing language. It supports Object headed programming approach to develop applications. it's straightforward and simple to find out and provides immeasurable high-level information structures. Python is straightforward to find out however powerful and versatile scripting language, that makes it engaging for Application Development. Python's syntax and dynamic writing with its taken nature create it a perfect language for scripting and speedy application development. Python supports multiple programming pattern, as well as object-oriented, imperative, and practical or procedural programming designs. Python isn't supposed to figure in a very explicit space, like net programming. that's why it's called utile programming language as a result of it will be used with net, enterprise, 3D CAD, etc. We don't have to be compelled to use information varieties to declare variable as a result of it's dynamically written therefore we are able to write `a=10` to assign associate whole number worth in associate whole number variable. Python makes the event and debugging quick as a result of there's no compilation step enclosed in Python development, and edit-test-debug cycle is incredibly quick.

3.2.1 PYTHON 2 VS. PYTHON 3

In most of the programming languages, whenever a replacement version releases, it supports the options and syntax of the present version of the language, therefore, it's easier for the comes to modify within the newer version. However, within the case of Python, {the two|the two} versions Python 2 and Python three are noticeably totally different from one another. A list of variations between Python two and Python three are given below: Python two uses `print` as a press release and used as `print "something"` to print some string on the console. On the opposite hand, Python three uses `print` as a operate and used as `print("something")` to print one thing on the console. Python two uses the operate `raw_input()` to simply accept the user's input. It returns the string representing the worth, that is written by the user. To convert it into the whole number, we'd like to use the `int()` operate in Python. On the opposite hand, Python three uses `input()` operate that mechanically taken the sort of input entered by the user. However, we are able to forged this worth to any sort by mistreatment primitive functions (`int()`, `str()`, etc.). In Python two, the implicit string sort is code, whereas, in Python three, the implicit string sort is Unicode. Python three does not contain the `xrange()` operate of Python two. The `x range()` is that the variant of `range()` operate that returns a `x range` object that works like Java iterator. The `range()` returns an inventory as an example the operate `range(0,3)` contains zero, 1, 2. There is additionally a tiny low amendment created in Exception handling in Python three. It defines a keyword as that is critical to be used. we'll discuss it in Exception handling section of Python programming tutorial.

Python was developed by Guido van Rossum within the late eighties and early nineties at the National analysis Institute for arithmetic and computing within the Holland. Python springs from several alternative languages, as well as first rudiment, Modula-3, C, C++, Algol-68, Small Talk, and operating system shell and alternative scripting languages. Python is proprietary. Like Perl, Python ASCII text file is currently obtainable underneath the antelope General Public License (GPL). Python is currently maintained by a core development team at the institute, though Guido van Rossum still holds a significant role in directive its progress.

3.2.2 PYTHON APPLICATIONS

Python is thought for its general purpose nature that produces it applicable in nearly every domain of software package development. Python as an entire may be employed in any sphere of development. Here, we have a tendency to square measure specifying applications areas wherever python may be applied.

3.2.2.1 INTERNET APPLICATIONS

We can use Python to develop internet applications. It provides libraries to handle net protocols like HTML and XML, JSON, Email process, request, BeautifulSoup, Feed parser etc. It conjointly provides Frameworks like Django, Pyramid, Flask etc to style and develop internet primarily based applications.

Some necessary developments are: Python Wiki Engines, Pocoo, Python Blog Software etc.

3.2.2.2 DESKTOP GRAPHICAL USER INTERFACE APPLICATIONS

Python provides Tk graphical user interface library to develop computer program in python primarily based application. another helpful toolkits wx Widgets, Kivy, pyqt that square measure useable on many platforms. The Kivy is fashionable for writing multitouch applications.

3.2.2.3 SOFTWARE PACKAGE DEVELOPMENT

Python is useful for software package development method. It works as a support language and may be used for build management and management, testing etc.

3.2.2.4 Scientific And Numeric

Python is fashionable and wide employed in scientific and numeric computing. Some helpful library and package square measure SciPy, Pandas, IPython etc. SciPy is cluster of packages of engineering, science and arithmetic.

3.2.2.5 BUSINESS APPLICATIONS

Python is employed to create Business applications like ERP and e-commerce systems. Tryton may be a high level application platform.

3.2.2.6 CONSOLE PRIMARILY BASED APPLICATION

We can use Python to develop console primarily based applications. For example: IPython.

3.2.2.7 AUDIO OR VIDEO PRIMARILY BASED APPLICATIONS

Python is awing to perform multiple tasks and may be accustomed develop transmission applications. a number of real applications are: TimPlayer, cplay etc.

3.2.2.8 3D CAD APPLICATIONS

To create CAD application social dancing may be a real application that provides full options of CAD.

3.2.2.9 ENTERPRISE APPLICATIONS

Python may be accustomed produce applications which may be used inside associate Enterprise or a company. Some real time applications are: OpenErp, Tryton, Picalo etc.

3.2.2.10 APPLICATIONS FOR PICTURES

Using Python many application may be developed for image. Applications developed are: VPython, Gogh, imgSeek etc.

CHAPTER 4

ALGORITHM

4.1 DECISION TREE ALGORITHM

In general, Decision tree is a predictive modelling tool which can be applied across many areas. We are using a decision tree algorithm which is a supervised learning algorithm unlike other supervised learning algorithms out there. The goal or reason of using decision algorithms is to create a training model which can further be used to predict the class of the variable by learning decision rules from the training data. This can be used for both regression tasks and classification.

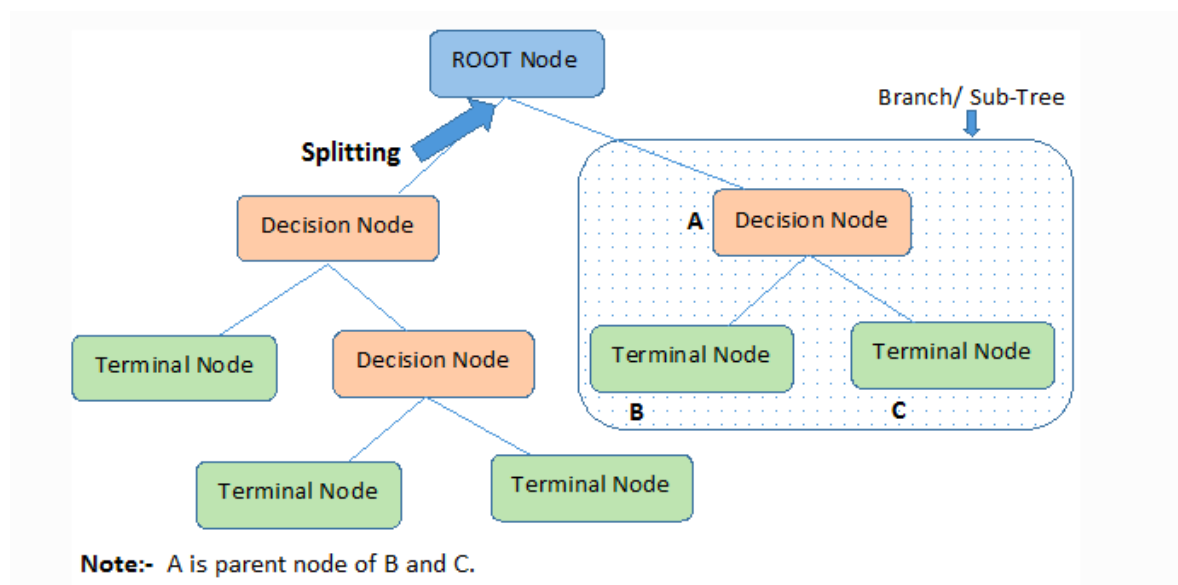


Fig. 4.1 Decision tree algorithm

4.2 LOGISTIC REGRESSION

Logistic regression algorithm is a supervised learning classification algorithm which used to predict the probability of an acquired variable. Nature of the resulting object or the dependent variable is dichotomous, which refers to having only 2 possible classes. In easy words, the dependent variable is binary having information coded as either One (stands for success/yes) or Zero (stands for

failure/no).

Mathematically, a logistic regression model/algorithm predicts $P(Y=1)$ as a function of X . It is one of the simplest Machine Learning algorithms that can be used for various problems such as spam detection, Diabetes prediction, cancer detection, etc. Before diving to the implementation of logistic regression, we must be aware of following assumptions about the same. We should choose a large sample size for logistic regression model

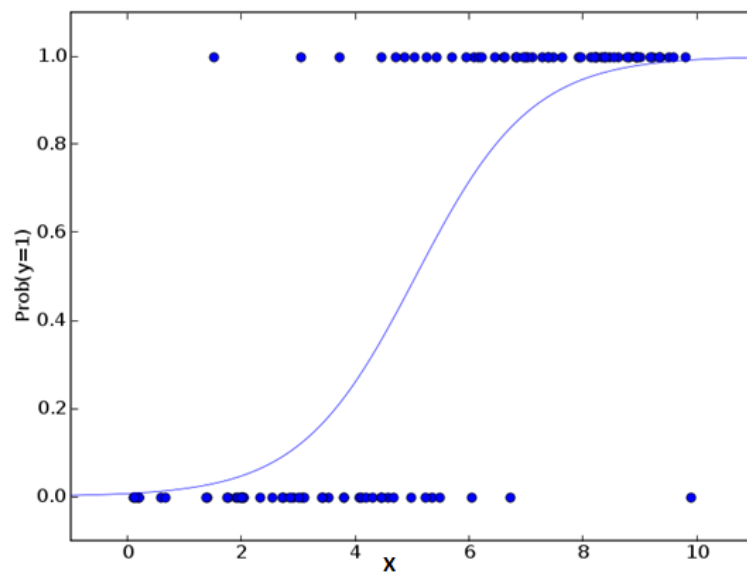


Fig 4.2 Logistic Regression

In the case of binary logistic regression, the target variables must be binary always and the desired outcome is represented by factor level 1. There should not be any multicollinearity in the model, which means the independent variables must be independent of each other. We must include some meaningful variables in our model. We should choose a large sample size for the logistic regression.

4.3 SVM

Support vector machines (SVMs) are powerful yet flexible supervised machine learning algorithms that are used both for classification and regression. But generally, they are used in classification problems. In 1960s, SVMs were first introduced but later they got refined in 1990s. SVMs have their unique way of implementation as compared to other machine learning algorithms SVMs have their unique way of implementation as SVMs. They are extremely popular because

of their ability to handle multiple categorical and continuous variables. SVM models are representation of different classes in a hyperplane in multidimensional space. The hyperplane will be generated iteratively by SVM so that the error can be minimized. The goal of SVM is to divide datasets into classes to find a MMH.

The followings are important concepts in SVM

Support Vectors, Data points that are closest to the hyperplane are called support vectors. The separating lines will be defined with the help of these data.

Hyperplane, As we can see in the diagram above, it is a decision plane or space which is divided between a set of objects which has different classes.

Margin, It may be defined as the gap between 2 lines on the closet data points of different classes.

It can be calculated as the perpendicular distance from line to support vectors.

A small margin is considered as a bad margin and a large margin is considered as a good margin

4.4 RANDOM FOREST

Random forest is a supervised learning algorithm that is used for both classifications as well as regression. However, it is mainly used for classification problems. As we all know a forest is made up of more trees means a more robust forest. Similarly, the random forest algorithm creates a decision tree on the data samples and then gets the prediction from each of them and selects the solution through voting. It is an ensemble method that is better than a single decision tree because it reduces overfitting by average result.

Working of Random Forest:

- *Step One* – start with the selection of random samples from a given dataset.
- *Step Two* – This algorithm will construct a decision tree for every sample. Then it will get the prediction result from every tree.
- *Step Three* – voting will be performed for every predicted result.
- *Step Four* – select the most voted prediction result as the final prediction result.

The following diagram will illustrate its work,

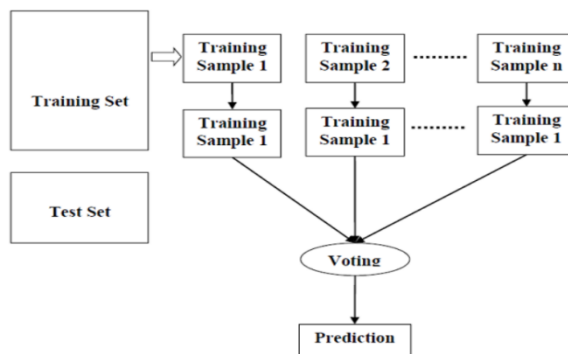


Fig 4.4 Random Forest

4.5 CSV CHARACTERISTICS

Collection of dataset which is fed into the Machine Learning algorithm to train our model (Validation data set).

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
1	animal_name	hair	feathers	eggs	milk	airborne	aquatic	predator	toothed	backbone	breathes	venomous	fins	legs	tail	domestic	catsize	class_type
2	aardvark	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1
3	antelope	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
4	bass	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
5	bear	1	0	0	1	0	0	1	1	1	1	0	0	4	0	0	1	1
6	boar	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
7	buffalo	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
8	calf	1	0	0	1	0	0	0	1	1	1	0	0	4	1	1	1	1
9	carp	0	0	1	0	0	1	0	1	1	0	0	1	0	1	1	0	4
10	catfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
11	cavy	1	0	0	1	0	0	0	1	1	1	0	0	4	0	1	0	1
12	cheetah	1	0	0	1	0	0	1	1	1	1	0	0	4	1	0	1	1
13	chicken	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2
14	chub	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	0	4
15	clam	0	0	1	0	0	0	1	0	0	0	0	0	0	0	0	0	7
16	crab	0	0	1	0	0	1	1	0	0	0	0	0	4	0	0	0	7
17	crayfish	0	0	1	0	0	1	1	0	0	0	0	0	6	0	0	0	7
18	crow	0	1	1	0	1	0	1	0	1	1	0	0	2	1	0	0	2
19	deer	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
20	dogfish	0	0	1	0	0	1	1	1	1	0	0	1	0	1	0	1	4
21	dolphin	0	0	0	1	0	1	1	1	1	1	0	1	0	1	0	1	1
22	dove	0	1	1	0	1	0	0	0	1	1	0	0	2	1	1	0	2
23	duck	0	1	1	0	1	1	0	0	1	1	0	0	2	1	0	0	2
24	elephant	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
25	flamingo	0	1	1	0	1	0	0	0	1	1	0	0	2	1	0	1	2
26	flea	0	0	1	0	0	0	0	0	0	1	0	0	6	0	0	0	6
27	frog	0	0	1	0	0	1	1	1	1	1	0	0	4	0	0	0	5
28	frog	0	0	1	0	0	1	1	1	1	1	1	0	4	0	0	0	5
29	fruitbat	1	0	0	1	1	0	0	1	1	1	0	0	2	1	0	0	1
30	giraffe	1	0	0	1	0	0	0	1	1	1	0	0	4	1	0	1	1
31	girl	1	0	0	1	0	0	1	1	1	1	0	0	2	0	1	1	1
32	gnat	0	0	1	0	1	0	0	0	0	1	0	0	6	0	0	0	6

Fig.4.5 CSV characteristics

4.6 TRAINING AND SPLITTING

Train-Test Split Evaluation. The train-test split is a technique for evaluating the performance of a machine learning algorithm. It can be used for classification or regression problems and can be used for any supervised learning algorithm. The procedure involves taking a data set and dividing it into two subsets.

4.7 DATAFLOW / DATA VISUALIZATION

The data is split into x correlation matrix excluding the class type for purpose of visualization of the dataflow using the **Matplotlib** (A library in python which is used to plot different types of graphs) . The data flow graph can be visually seen through images mentioned below (FIG 4.2.1..... FIG 4.2.6) , the image in FIG 4.2.8 describes the heat map of the given data .

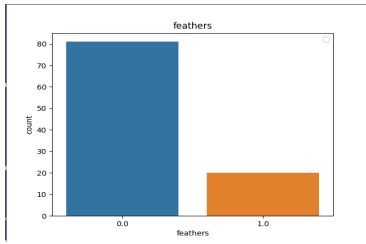


Fig.4.2.1 Dataset

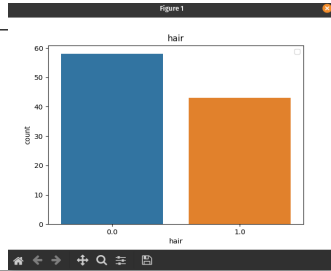
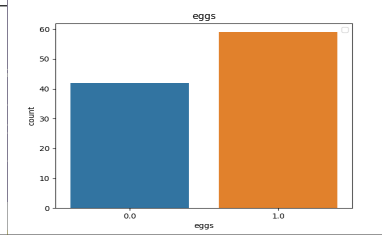
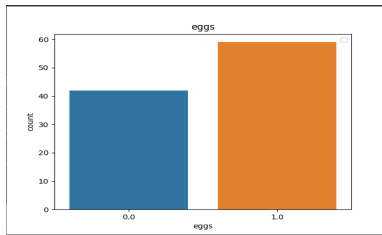
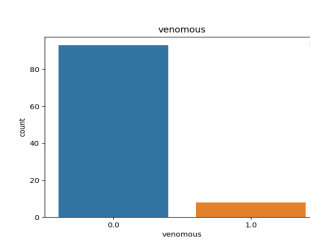
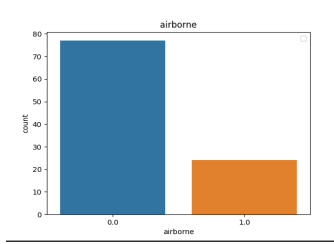


Fig 4.2.2 Dataset

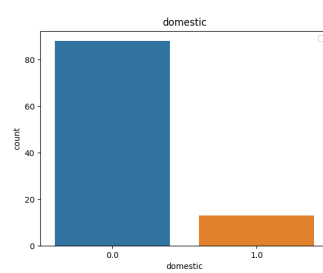
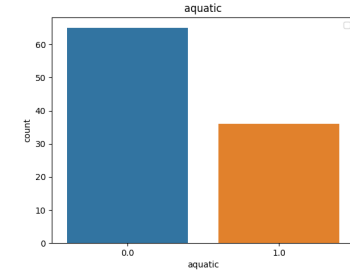
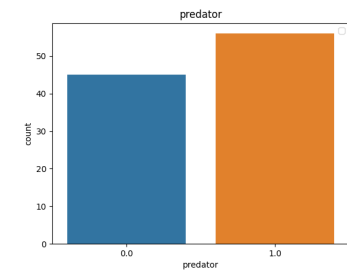


Fig 4.2.3 Dataset

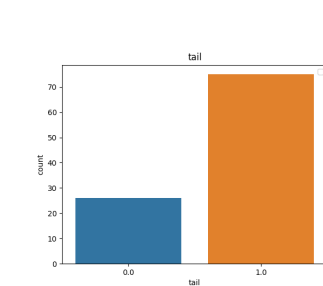
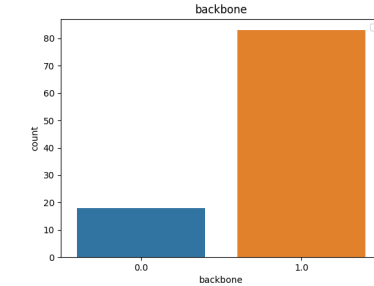
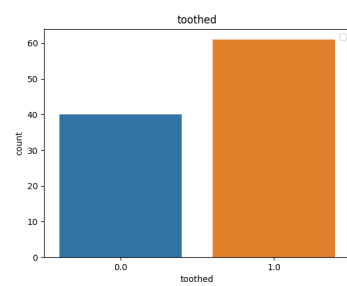


Fig 4.2.4 Dataset

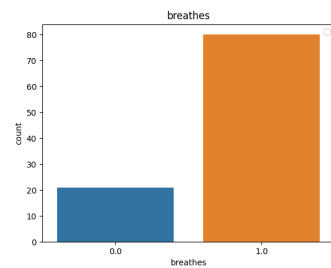


Fig 4.2.5 Dataset

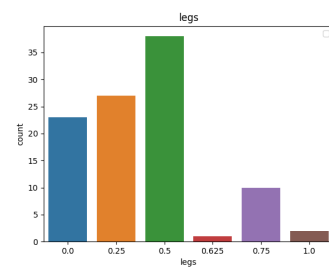


Fig 4.2.6 Graph chart

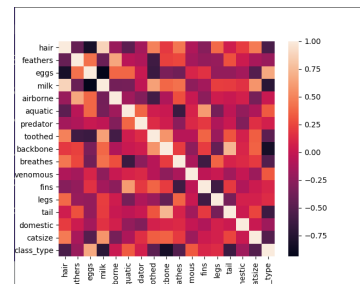


Fig 4.2.7 Heat Map

4.8 PRECISION / ACCURACY CHECK

Since Machine Learning is based on predicting we cannot expect it to give us a correct answer for anything, So accuracy is important. By using the accuracy score we can trust the predicted data. Here in this project we are calculating the accuracy by comparing the Y Test data and Predicted data. The Comparison is achieved by Sklearn metrics (classification report and accuracy_score libraries).

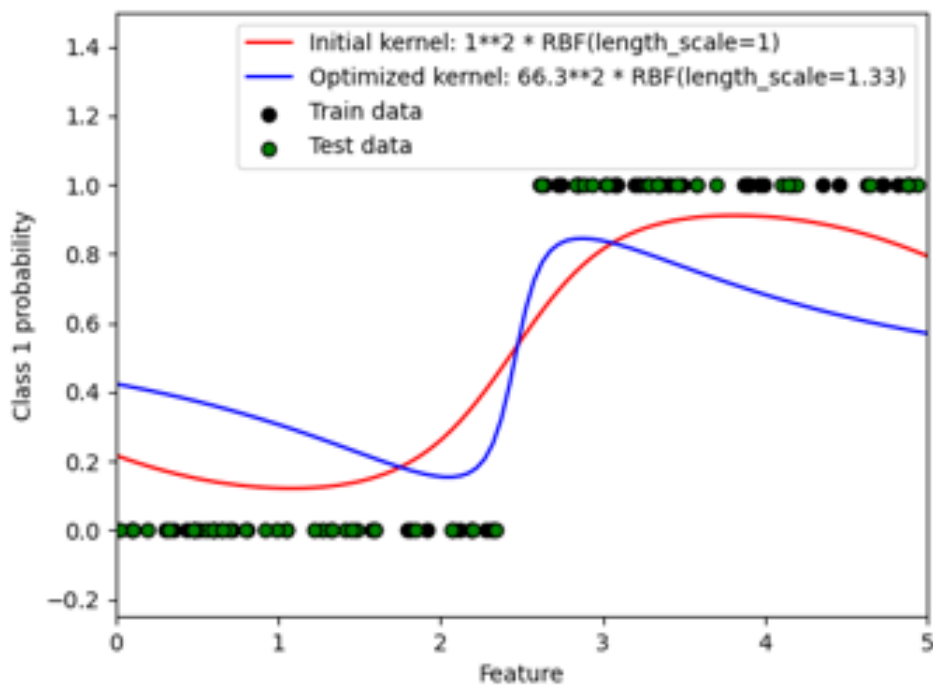


Fig.4.3 Accuracy

CHAPTER 5

RESULT

When the prototype was put into working, it worked as expected. The Animals test data is predicted with the help of Decision tree algorithm. Based on the input Dataset the developed model works and classifies input. Here the Animal is identified by class type values. By the help of Classified data the Animal is predicted for future purposes. The manual methods are replaced by Machine Learning Models.

5.1 FEATURES OF ANIMAL SPECIFICATION



```
print(model.predict([[0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0]]))
# **original value - 4, predicted value - 4, 100%correct**

print(model.predict([[0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]]))
# **original value - 7, predicted value - 7, 100% correct**

print(model.predict([[1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 6, 0, 0, 0]]))
# **original value - 6, predicted value - 6, 100% correct**

print(model.predict([[1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 4, 1, 0, 1]]))
```

Fig 5.1 Features of animal specification

5.2 PREDICTED DATA ON SHELL

```
Prediction:
[5 7 1 2 7 4 6 1 5 1 6 1 2 6 1 1 4 1 1 2 7]

Classification Report:

              precision    recall  f1-score   support

     1              1.00        1.00        1.00         8
     2              1.00        1.00        1.00         3
     3              0.00        0.00        0.00         1
     4              1.00        1.00        1.00         2
     5              0.50        1.00        0.67         1
     6              1.00        0.75        0.86         4
     7              0.67        1.00        0.80         2

 accuracy              0.90         21
 macro avg              0.74        0.82        0.76         21
 weighted avg          0.90        0.90        0.89         21

Accuracy Score: 0.9047619047619048

Final Output:
[4]
[7]
[6]
[1]
[2]
```

Fig 5.2 Predicted Data on shell

we consider applying machine learning to the analysis and compression of audio signals in the context of monitoring elephants in sub-Saharan Africa. Earths biodiversity is increasingly under threat by sources of anthropogenic change (e.g. resource extraction, land use change, and climate change) and surveying animal populations is critical for developing conservation strategies. However, manually monitoring tropical forests or deep oceans is intractable. For species that communicate acoustically, researchers have argued for placing audio recorders in the habitats as a cost effective and non- invasive method, a strategy known as passive systems and can be implemented on future.

CHAPTER 6

SUMMARY AND CONCLUSION

When the prototype was put into working, it worked as expected. The Patient's Condition is predicted with the help of KNN algorithm. Based on the input Dataset the developed model works and classifies input. Here the health condition is identified by EEG values. By the help of Classified data the health is predicted for future purposes. The manual methods are replaced by Machine Learning Models and Cost is reduced.

This work can be further extended by sending an alert in the form of a message when the animal is detected to the nearby forest office. Furthermore it can be used to reduce human wildlife conflict and also animal accidents. This project uses Decision Tree algorithm to classify wild animals. The algorithm classifies animals efficiently with a good number of accuracy and also the data of the classified animal is displayed for a better result so that it can be used for other purposes such as detecting wild animals entering into human habitat and to prevent wildlife poaching and even human animal conflict.

REFERENCES

1. Xie, Z., A. Singh, J. Uang, K.S. Narayan and P. Abbeel. Multimodal blending for high-accuracy instance classification. *Intelligent Robots and Systems*. pp. 2214-2221.(2003)
2. Ahonen, T., Hadid, A., Pietikainen, and M.: data description with local binary patterns: Application to face recognition, pp .2037- 2041 (2006).
3. Burghardt, T., Calic, J.: Real-time tracking of animals. *Neural Network Applications in Electrical Engineering*. pp. 27 (2006).
4. Felzenszwalb, P.F., Girshick, R.B., McAllester, D., Ramanan, D.: Object detection with discriminatively trained part-based models, pp. 1627-1645 (2010).
5. M. Benco, R. Hudec, P. Sykora and A. Satnik. Accurate wild animal recognition using PCA, LDA and LBPH, pp. 62–67 , 2016.
6. WU, J. L. and W. Y. MA. A Deep Learning Framework for Coreference Resolution Based on Convolutional Neural Network. 11th International Conference on Semantic Computing (ICSC). San Diego pp. 61–64, 2017.
7. P.M. Vitousek, H.A. Mooney, J. Lubchenco, J. Melillo, "Human domination of Earth's ecosystems", *Science*, vol. 277, no. 5325, pp. 494- 499, 1997.
8. B.J. Godley, J. Blumenthal, A. Broderick, M. Coyne, M. Godfrey, L. Hawkes, M. Witt, "Satellite tracking of sea turtles: Where have we been and where do we go next?", *Endangered Species Research*, vol.4, no.1-2, pp.3- 22, 2008.
10. Kamencay, P., T. Trnovszky, M. Benco, R. Hudec, P. Sykora and A. Satnik. Accurate wild animal recognition using PCA, LDA and LBPH, In *ELEKTRO* pp. 62–67 , 2016 .
9. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: *CVPR*. pp. 770-778 (2016)

APPENDICES

SOURCE CODE

```
import numpy as np # linear algebra
import pandas as pd # data processing, CSV file I/O (e.g. pd.read_csv)
from warnings import simplefilter
from colorama import Fore, Back, Style #Colors
simplefilter(action='ignore')

# # **Loading the data**
data = pd.read_csv(r"data.csv")
print(data)

# # **Dropping the animal_name column**
data.drop("animal_name", axis=1, inplace=True)
print(data)

# # **Data processing**
# **checking for null values**
print(data.isnull().sum())

# **scaling the data**

from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler()
data_scaled = scaler.fit_transform(data.drop('class_type',axis=1))
scaler
print(data_scaled)
X = pd.DataFrame(data_scaled, columns=data.columns.drop('class_type'))
y = data['class_type']

print(X)

print(y)

# # **Data Visualization**
import matplotlib.pyplot as plt
import seaborn as sns
```

```

print(X)

#""Starts
# **Class_type distribution**
# **class distribution of all columns in X**
for col in X.columns[:-1]:
    sns.countplot(X[col])
    plt.title(col)
    plt.legend()
    plt.show()

# **data correlation matrix**
data_corr = data.corr()
sns.heatmap(data_corr)
plt.show()

# **X correlation matrix**
x_corr = X.corr()
sns.heatmap(x_corr)
plt.show()

# **histogram of all columns in data**
print(data.hist(bins=20))

# **histogram of X columns**
print(X.hist(bins=20))
#Ends""

# # **Splitting the data into training and testing sets**
from sklearn.model_selection import train_test_split
x_train, x_test, y_train, y_test = train_test_split(X, y, test_size=0.2,
random_state=28)

print(x_train, "\n")

# # **Model Building and training**
from sklearn.tree import DecisionTreeClassifier
model = DecisionTreeClassifier()
model.fit(x_train, y_train)

```

```

# # **Predictions**
pred = model.predict(x_test)
print("Prediction: ")
print(pred, "\n")

# # **Precision, Recall and Accuracy check**
from sklearn.metrics import classification_report, accuracy_score
print("Classification Report: ")
print(classification_report(y_test,pred))

print(Fore.GREEN + "Accuracy Score: ", accuracy_score(y_test,pred))
print(Style.RESET_ALL)
print("\n")

data2 = pd.read_csv(r"data.csv")
# # **Lets test on random values from the data**
print(data2, "\n")

print(Fore.GREEN + "Final Output: ")
print(model.predict([[0, 0, 1, 0, 0, 1, 1, 1, 1, 0, 0, 1, 0, 1, 0, 0]])) #2nd col of the
data
# **original value - 4, predicted value - 4, 100%correct**

print(model.predict([[0, 0, 1, 0, 0, 0, 0, 0, 0, 1, 0, 0, 0, 0, 0, 0]])) #99th col of data
# **original value - 7, predicted value - 7, 100% correct**

print(model.predict([[1, 0, 1, 0, 1, 0, 0, 0, 0, 1, 1, 0, 6, 0, 0, 0]])) #97th col of the
data
# **original value - 6, predicted value - 6, 100% correct**

print(model.predict([[1, 0, 0, 1, 0, 0, 0, 1, 1, 1, 0, 0, 4, 1, 0, 1]])) #4th col of the data
# **original value - 1, predicted value - 1, 100% correct**
print("\n")

```