**INTERDISCIPLAINARY TRAINING REPORT**

**at**

**Sathyabama Institute of Science and Technology**

**(DEEMED TO BE UNIVERSITY)**

Submitted in partial fulfillment of the requirements for the award of

Bachelor of Engineering Degree in

Computer Science and Engineering

By

**ARUNKUMAR V**

**(Reg. No. 39110086)**



**DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING**

**SCHOOL OF COMPUTING**

**SATHYABAMA INSTITUTE OF SCIENCE AND TECHNOLOGY**

**JEPPIAAR NAGAR, RAJIV GANDHI SALAI,**

**CHENNAI, TAMILNADU – 600119.**

**APRIL 2022**

|  |  |  |
| --- | --- | --- |
|  | **SATHYABAMA**  **INSTITUTE OF SCIENCE AND TECHNOLOGY** (DEEMED TO BE UNIVERSITY) **Accredited with Grade “A” by NAAC**  (Established under Section 3 of UGC Act, 1956)  JEPPIAAR NAGAR, RAJIV GANDHI SALAI  CHENNAI– 600119  [www.sathyabama.ac.in](http://www.sathyabama.ac.in) |  |

# DEPARTMENT OF COMPUTER SCIENCE AND ENGINEERING

**BONAFIDE CERTIFICATE**

This is to certify that this Project Report is the Bonafide work of **ARUNKUMAR V (Reg. No. 39110086)** who carried out the project entitled **“ANALYSIS OF SKILLS REQUUIRED FOR A DATA SCIENTIST USING MACHINE LEARNING”** under my supervision during February 2022 - April 2022.

## Internal Guide

## MS.MADHUMITHA

**Head of the Department**

**Dr. L. Lakshmanan, M.E., Ph.D.,**



## Submitted for Viva voce Examination held on

**Internal Examiner External Examiner**

**DECLARATION**

I, **ARUNKUMAR V** hereby declare that the project report entitled **ANALYSIS OF SKILLS REQUIRED FOR A DATA SCIENTIST USING MACHINE LEARNING** done by me under the guidance of **MS.MADHUMITHA** is submitted in partial fulfillment of the requirements for the award of Bachelor of Engineering Degree in Computer Science and Engineering

**DATE:**

**PLACE: SIGNATURE OF THE CANDIDATE**

**ACKNOWLEDGEMENT**

I am pleased to acknowledge my sincere thanks to **Board of Management** of **SATHYABAMA** for their kind encouragement in doing this project and for completing it successfully. I am grateful to them.

I convey my thanks to **Dr. T. Sasikala M.E., Ph.D.**, **Dean**, School of Computing, **Dr. S. Vigneshwari, M.E., Ph.D. and Dr. L. Lakshmanan, M.E., Ph.D., Heads of the Department** of **Computer Science and 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 **MS.MADHUMITHA** for his valuable guidance, suggestions and constant encouragement paved 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 Computer Science and Engineering** who were helpful in many ways for the completion of the project.

**TRAINING CERTIFICATE**

**ABSTRACT**

The Data Scientist role demands a person to be a jack of many skills. This information can be used to gain insights such as the key skills and tools required and which degree is preferred the most and the categorization of different roles based on the key skills, etc. This information of 7,000 data scientist jobs around the U.S is collected from Indeed website. The information collected consists of: Company Name, Position Name, Location, Job Description, and Number of Reviews of the Company.

By Analysis Of Skills Required For A Data Scientist Using IBM Watson we will:

* Know fundamental concepts and can work on IBM Watson Studio.
* Gain a broad understanding of Stochastic Gradient Descent Classification.

Artificial Intelligence being the trending technology can end up with the best solution for every typical problem. So here we are going to use the rich set of Machine Learning algorithms to predict the position.

By using the supervised learning algorithms of Machine learning such as Stochastic Gradient Descent classification we can predict the position.

**TABLE OF CONTENTS**

**TITLE PAGE NO**

**CHAPTER 1: INTRODUCTION** 8

* 1. OVERVIEW 8
  2. PURPOSE 8

**CHAPTER 2: LITERATURE SURVEY 8-9**

* 1. EXISTING PROBLEM 8
  2. PROPOSED SOLUTION 9

**CHAPTER 3: THEORTICAL ANALYSIS 9-10**

3.1 BLOCK DIAGRAM 9

3.2 HARDWARE/SOFTWARE DESIGNING 10

**CHAPTER 4: EXPERIMENTAL INVESTIGATIONS 11-12**

**CHAPTER 5: FLOW CHART 13**

**CHAPTER 6: RESULTS 14-15**

**CHAPTER 7: ADVANTAGES AND DISADVANTAGES 16**

**CHAPTER 8: APLLICATIONS 16**

**CHAPTER 9: CONCLUSION 17**

**CHAPTER 10: FUTURE SCOPE 17**

**CHAPTER 11: BIBILOGRAPHY 18**

**CHAPTER 12: APPENDIX 19-24**

**CHAPTER 1**

**INTRODUCTION**

* 1. **Overview:**

The Data Scientist role demands a person to be a jack of many skills. This information can be used to gain insights such as the key skills and tools required and which degree is preferred the most and the categorization of different roles based on the key skills, etc. This information of 7,000 data scientist jobs around the U.S is collected from Indeed website. The information collected consists of: Company Name, Position Name, Location, Job Description, and Number of Reviews of the Company.

* 1. **Purpose:**

By Analysis Of Skills Required For A Data Scientist Using IBM Watson we will:

* Know fundamental concepts and can work on IBM Watson Studio.
* Gain a broad understanding of Stochastic Gradient Descent Classification.

**CHAPTER 2**

**LITERATURE SURVEY**

* 1. **Existing Problem:**

Determining the position of different Data Scientist is a difficult task for a human as he has to consider different parameter like skills, tools, degree and come up with a conclusion. By Using the skills, tools, degree we can predict the position of Data Scientist easily

It is quite difficult for a human to work with such a huge and typical data. So here we are coming up with a solution

* 1. **Proposed Solution:**

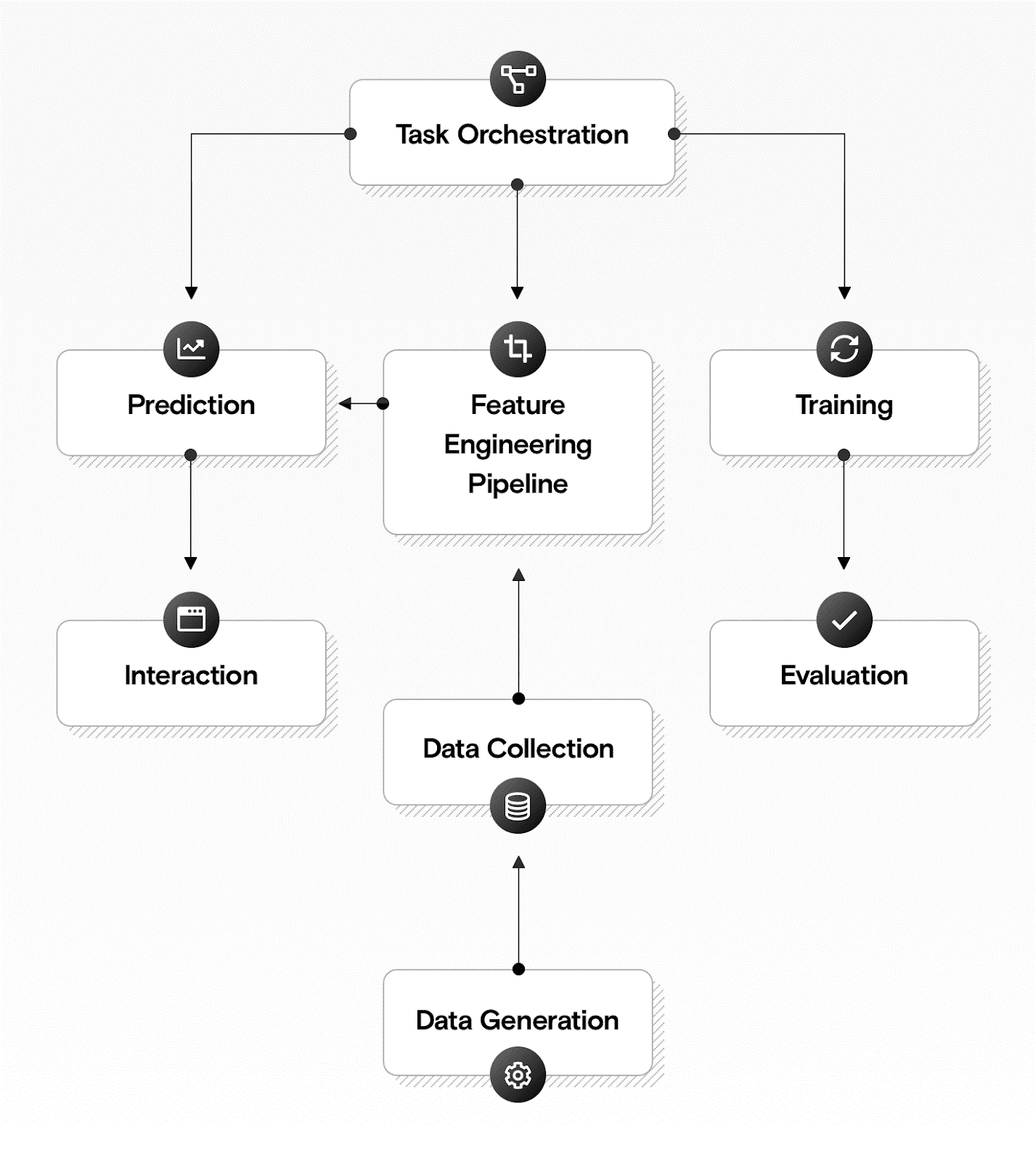
Artificial Intelligence being the trending technology can end up with the best solution for every typical problem. So here we are going to use the rich set of Machine Learning algorithms to predict the position.

By using the supervised learning algorithms of Machine learning such as Stochastic Gradient Descent classification we can predict the position.

**CHAPTER 3**

**THEORTICAL ANALYSIS**

**3.1 Block Diagram:**

****

**3.2 Hardware/software designing:**

**Software specifications:**

|  |  |
| --- | --- |
| **REQUIREMENT** | **SPECIFICATION** |
| Anaconda Navigator | You must have anaconda installed in your device prior to begin. |
| Spyder, Jupyter Notebook, Flask  Frame work | 1. One should have Spyder and Jupyter notebook. 2. One should install flask framework through anaconda prompt for running their web application 3. We need to build the model using Jupyter notebook with all the imported packages. |
| Web browser | For all Web browsers, the following must be enabled:   * cookies * JavaScript |

**Hardware Specifications:**

|  |  |
| --- | --- |
| **REQUIREMENT** | **SPECIFICATIONS** |
| Operating system | Microsoft Windows  UNIX  Linux® |
| Processing | Minimum: 4 CPU cores for one user. For each deployment, a sizing exercise is highly recommended. |
| RAM | Minimum 8 GB. |
| Operating system specifications | File descriptor limit set to 8192 on UNIX and Linux |
| Disk space | A minimum of 7 GB of free space is required to install the software. |

**CHAPTER 4**

**EXPERIMENTAL INVESTIGATIONS**

Analysis or the investigation made while working on the solution:

While working on the solution we investigated on what is IBM cloud, IBM Watson studio, Machine Learning service, Cloud Object Storage. The key role on investigation is collection of the dataset.

**IBM Cloud Account**:

IBM Acquired soft layer, a public cloud platform, to serve as the foundation for its IaaS offering. In October 2016, IBM rolled the soft layer brand under its Blue mix brand of PaaS offerings, giving users to access both IaaS and PaaS resources from a single console. IBM cloud provides a full-stack, public cloud platform with various products in the catalog, including options for compute, storage, networking, end to end developer solutions for app development, testing and deployment, security databases, and cloud native services.

Creating the IBM cloud account by going to the IBM cloud login page and click create on IBM cloud account. Enter our IBM id and an ID is created based on the email that we enter. Completing the remaining fields with our information and click create account by this the account is created.

**Dataset collection**:

The data collection by:

* Articulate the problem early.
* Establish data collection.
* Check our data quickly.
* Format data to make it consistent.
* Reduce data.
* Complete data cleaning.
* Decompose data.
* Take the required fields of data.

**Model Building:**

* Training and testing the model
* Evaluation of Model
* Save the model
* Predicting the output using the model

**CHAPTER 5**

**FLOW CHART**

DATA COLLECTION

DATA PREPROCESSING

IDENTIFY DEPENDENT AND INDEPENDENT VALUES

DATA VISUALISATION

IDENTIFY THE REQUIRED COLUMNS

SPLIT THE DATA INTO TRAIN AND TEST

APPLYING REGRESSION, CLASSIFICATION ALGORITHAM

TEST THE MODEL

SAVE THE MODEL

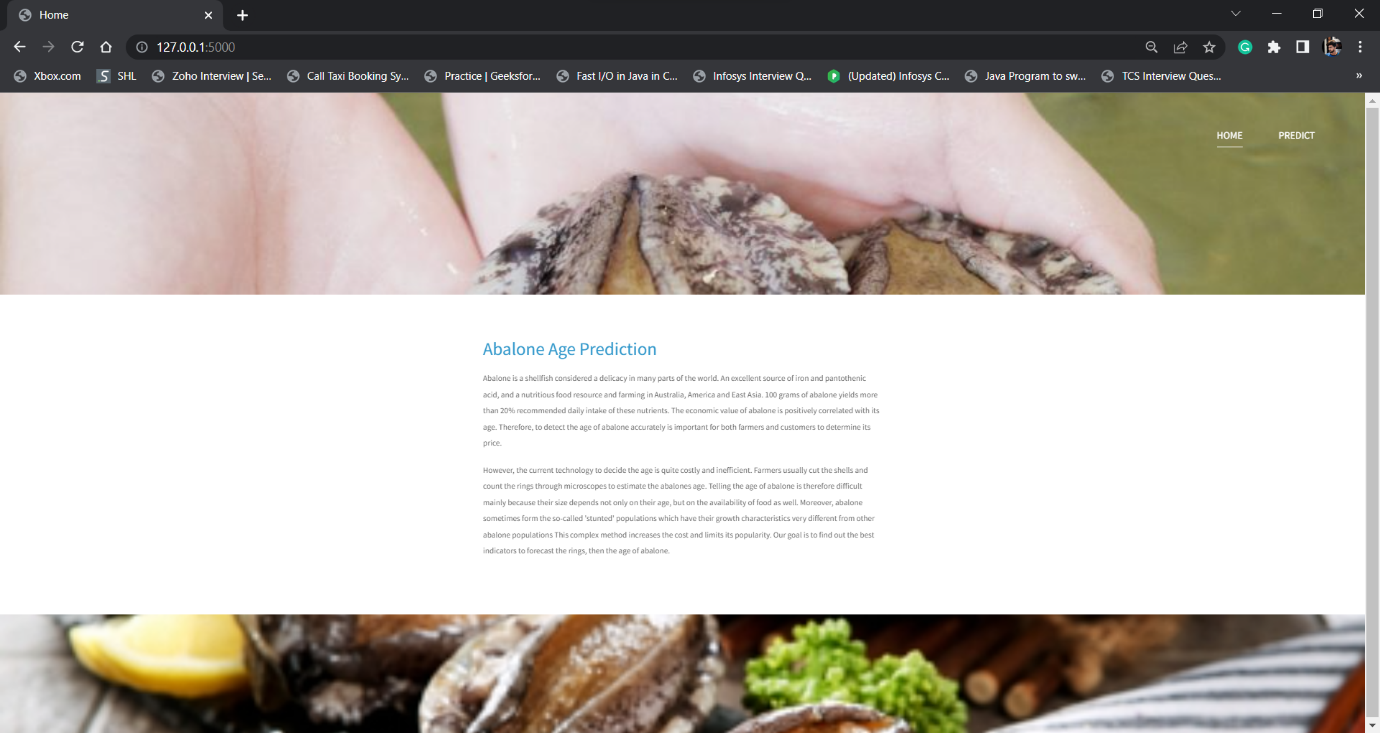
CREATE THE FLASK APPLICATIONAND INTEGRATE WITH MODEL

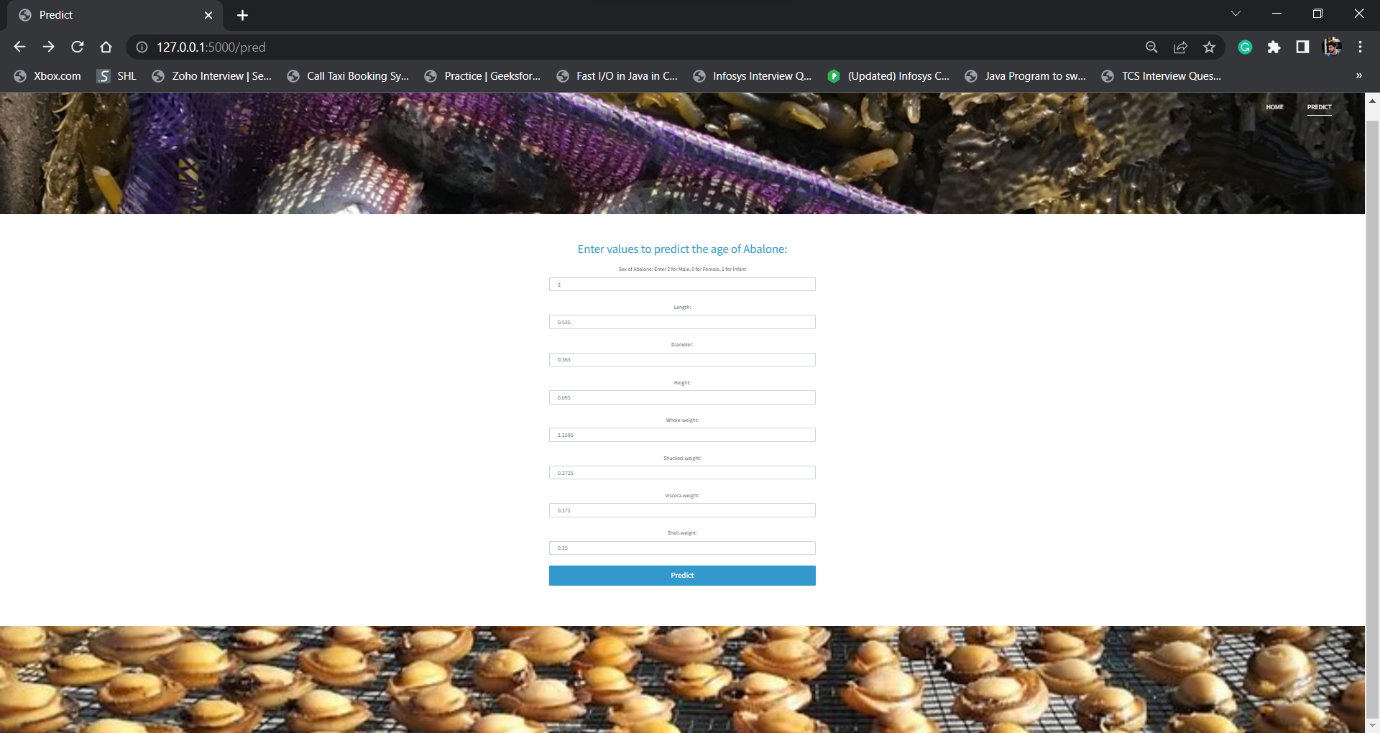
DEPLOY MODEL ON IBM

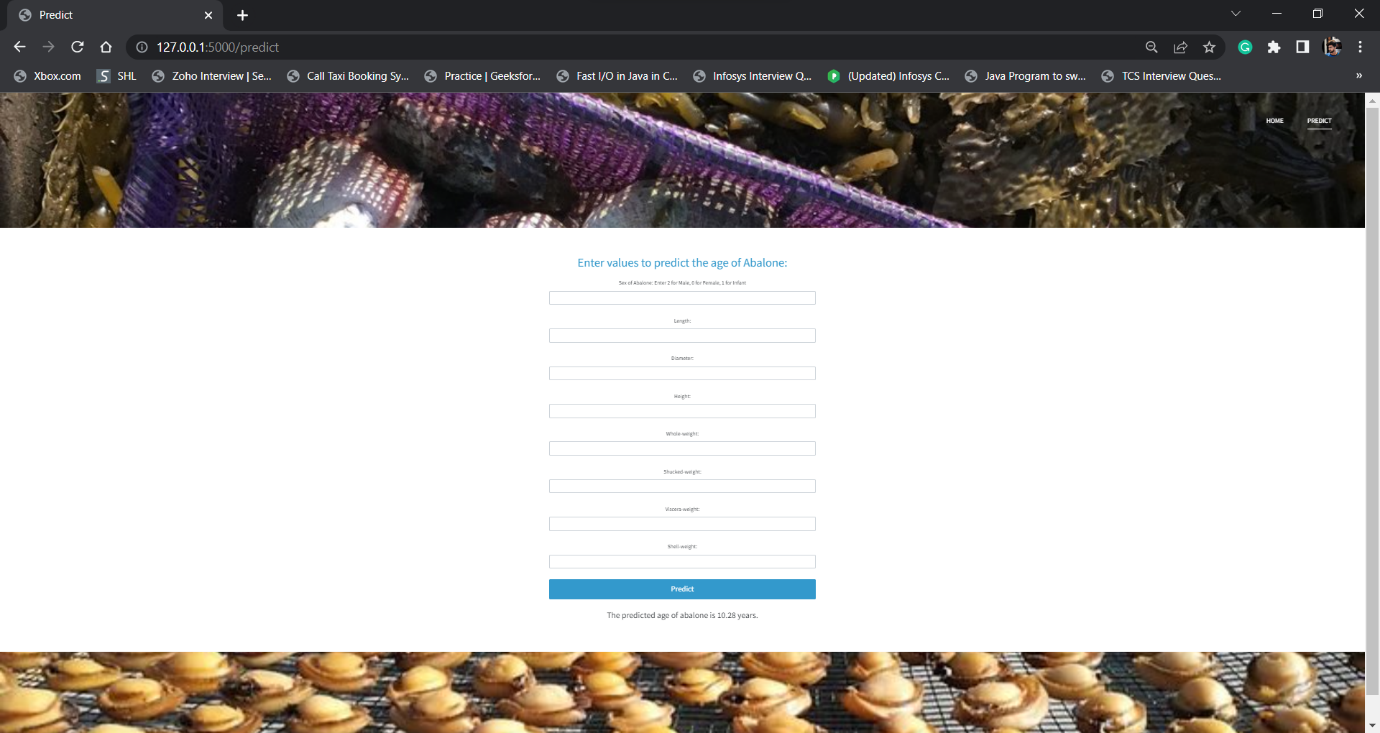
**CHAPTER 6**

**RESULTS**

**Final output of the project:**

****





**CHAPTER 7**

**ADVANTAGES AND DISADVANTAGES**

**Advantages:**

* Lower costs - reduces maintenance due to complete report coverage and a zero-footprint environment.
* Faster results - shortens reporting time due to seamless integration and adaptive authoring.
* High performance data access across all source.

**Disadvantages:**

* The permission level for a user cannot be modified
* Data grouping
* Custom visualizations
* Insights in visualization

Lower c Lower costs—reduces maintenance due to complete rep

High performance **data**

**ac CHAPTER 8**

**APLLICATIONS**

**The areas where this solution can be applied:**

* A brief aside on the motivation behind collecting the dataset. its Abalone is a type of consumable snail whose price varies as per age. The aim is to predict the age of abalone from physical measurements. The age of abalone is traditionally determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope – a boring and time consuming task. Other measurements, which are easier to obtain, are used to predict the age.
* The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope -- a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age.
* The main objective is to determine the age of Abalone from the physical measurements.

**CHAPTER 9**

**CONCLUSION**

**From this entire findings we know fundamental concepts and can work on IBM Watson and machine learning.**

* On the source of this study it appears the future regression systems effort well to forecast the age of abalone.
* The study directs that we do not prerequisite to count the quantity of rings consuming microscopic test. In other disputes, we do not need any laboratory experiment to predict the age of abalones.
* We can predict the age and price of abalone using the very simple physical individualities like weight, height, diameter, and length.

**CHAPTER 10**

**FUTURE SCOPE**

**Enhancements that can be made in the future:**

* The age of abalone is determined by cutting the shell through the cone, staining it, and counting the number of rings through a microscope -- a boring and time-consuming task. Other measurements, which are easier to obtain, are used to predict the age.
* The main objective is to determine the age of Abalone from the physical measurements.

**CHAPTER 11**

**BIBILOGRAPHY**

[1] Abalone. http://en.wikipedia.org/wiki/Abalone.

[2] UCI Machine Learning Repository: Abalone Dataset. http://archive.ics.uci.edu/ml/datasets/Abalone.

[3] Al-Daoud, E. (2009). A comparison between three neural network models for classification problems. Journal of artificial intelligence, 2(2), 56-64.

[4] Bretos, M. (1980). Age determination in the keyhole limpet Fissurella crassa Lamarck (Archaeogastropoda: Fissurellidae), based on shell growth rings. The Biological Bulletin, 159(3), 606-612.

[5] Clark, D., Schreter, Z., & Adams, A. (1996, April). A quantitative comparison of dystal and backpropagation. In Australian Conference on Neural Networks.

[6] Day, R. W., Williams, M. C., & Hawkes, G. P. (1995). A comparison of fluorochromes for marking abalone shells. Marine and Freshwater Research, 46(3), 599-605.

[7] Gurney, L. J., Mundy, C., & Porteus, M. C. (2005). Determining age and growth of abalone using stable oxygen isotopes: a tool for fisheries management. Fisheries Research, 72(2-3), 353- 360.

[8] Naylor, J. R., Manighetti, B. M., Neil, H. L., & Kim, S. W. (2007). Validated estimation of growth and age in the New Zealand abalone Haliotis iris using stable oxygen isotopes. Marine and freshwater research, 58(4), 354-362.

[9] Siddeek, M. S. M., & Johnson, D. W. (1997). Growth parameter estimates for Omani abalone (Haliotis mariae, Wood 1828) using length-frequency data. Fisheries research, 31(3), 169-188.

[10] Takami, H., Oshino, A., Sasaki, R., Fukazawa, H., & Kawamura, T. (2006). Age determination and estimation of larval period in field caught abalone (Haliotis discus hannai Ino 1953) larvae and newly metamorphosed post-larvae by counts of radular teeth rows. Journal of experimental marine biology and ecology, 328(2), 289-301.

**APPENDIX**

**home.html**

<!DOCTYPE html>

<html lang="en">

<head>

<meta charset="UTF-8">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Home</title>

<link rel="stylesheet" href="/static/fontawesome/css/all.min.css"> <!-- https://fontawesome.com/ -->

<link href="https://fonts.googleapis.com/css2?family=Source+Sans+Pro&display=swap" rel="stylesheet">

<!-- https://fonts.google.com/ -->

<link rel="stylesheet" href="/static/css/bootstrap.min.css">

<link rel="stylesheet" href="/static/css/templatemo-video-catalog.css">

<!--

TemplateMo 552 Video Catalog

https://templatemo.com/tm-552-video-catalog

-->

</head>

<body>

<div class="tm-page-wrap mx-auto">

<div class="position-relative">

<div class="potition-absolute tm-site-header">

<div class="container-fluid position-relative">

<div class="row">

<div class="col-5 col-md-8 ml-auto mr-0">

<div class="tm-site-nav">

<nav class="navbar navbar-expand-lg mr-0 ml-auto" id="tm-main-nav">

<button class="navbar-toggler tm-bg-black py-2 px-3 mr-0 ml-auto collapsed" type="button"

data-toggle="collapse" data-target="#navbar-nav" aria-controls="navbar-nav"

aria-expanded="false" aria-label="Toggle navigation">

<span>

<i class="fas fa-bars tm-menu-closed-icon"></i>

<i class="fas fa-times tm-menu-opened-icon"></i>

</span>

</button>

<div class="collapse navbar-collapse tm-nav" id="navbar-nav">

<ul class="navbar-nav text-uppercase">

<li class="nav-item active">

<a class="nav-link tm-nav-link" href="/home">Home<span class="sr-only">(current)</span></a>

</li>

<li class="nav-item">

<a class="nav-link tm-nav-link" href="/pred">Predict</a>

</li>

</ul>

</div>

</nav>

</div>

</div>

</div>

</div>

</div>

<div class="tm-welcome-container tm-fixed-header tm-fixed-header-2">

</div>

<div id="tm-fixed-header-bg"></div> <!-- Header image -->

</div>

<!-- Page content -->

<main>

<div class="container-fluid px-0">

<div class="mx-auto tm-content-container">

<div class="row mt-3 mb-5 pb-3">

<div class="col-12">

<div class="mx-auto tm-about-text-container px-3">

<h2 class="tm-page-title mb-4 tm-text-primary">Abalone Age Prediction</h2>

<p class="mb-4">Abalone is a shellfish considered a delicacy in many parts of the world. An excellent source of iron and pantothenic acid, and a nutritious food resource and farming in Australia, America and East Asia. 100 grams of abalone yields more than 20% recommended daily

intake of these nutrients. The economic value of abalone is positively correlated with its age. Therefore, to detect the age of abalone accurately is important for both farmers and customers to determine its price.</p>

<p class="mb-4">However, the current technology to decide the age is quite costly and inefficient. Farmers usually cut the shells and count the rings through microscopes to estimate the abalones age. Telling the age of abalone is therefore difficult mainly because their size

depends not only on their age, but on the availability of food as well. Moreover, abalone sometimes form the so-called &#39;stunted&#39; populations which have their growth characteristics very different from other abalone populations This complex method increases the cost and

limits its popularity. Our goal is to find out the best indicators to forecast the rings, then the age of abalone.</p>

</div>

</div>

</div>

</div>

<div class="parallax-window" data-parallax="scroll" data-image-src="/static/img/about-2.jpg"></div>

</div>

</main>

</div>

<script src="/static/js/jquery-3.4.1.min.js"></script>

<script src="/static/js/bootstrap.min.js"></script>

<script src="/static/js/parallax.min.js"></script>

</body>

</html>

**Upload.html**

<!DOCTYPE html>

<html lang="en">

<head>

<meta charset="UTF-8">

<meta name="viewport" content="width=device-width, initial-scale=1.0">

<title>Predict</title>

<link rel="stylesheet" href="/static/fontawesome/css/all.min.css"> <!-- https://fontawesome.com/ -->

<link href="https://fonts.googleapis.com/css2?family=Source+Sans+Pro&display=swap" rel="stylesheet">

<!-- https://fonts.google.com/ -->

<link rel="stylesheet" href="/static/css/bootstrap.min.css">

<link rel="stylesheet" href="/static/css/templatemo-video-catalog.css">

</head>

<!--

TemplateMo 552 Video Catalog

https://templatemo.com/tm-552-video-catalog

-->

<body>

<div class="tm-page-wrap mx-auto">

<div class="position-relative">

<div class="potition-absolute tm-site-header">

<div class="container-fluid position-relative">

<div class="row">

<div class="col-5 col-md-8 ml-auto mr-0">

<div class="tm-site-nav">

<nav class="navbar navbar-expand-lg mr-0 ml-auto" id="tm-main-nav">

<button class="navbar-toggler tm-bg-black py-2 px-3 mr-0 ml-auto collapsed" type="button"

data-toggle="collapse" data-target="#navbar-nav" aria-controls="navbar-nav"

aria-expanded="false" aria-label="Toggle navigation">

<span>

<i class="fas fa-bars tm-menu-closed-icon"></i>

<i class="fas fa-times tm-menu-opened-icon"></i>

</span>

</button>

<div class="collapse navbar-collapse tm-nav" id="navbar-nav">

<ul class="navbar-nav text-uppercase">

<li class="nav-item">

<a class="nav-link tm-nav-link" href="/home">Home</a>

</li>

<li class="nav-item active">

<a class="nav-link tm-nav-link" href="/pred">Predict<span class="sr-only">(current)</span></a>

</li>

</ul>

</div>

</nav>

</div>

</div>

</div>

</div>

</div>

<div class="tm-welcome-container tm-fixed-header tm-fixed-header-3">

</div>

<div id="tm-fixed-header-bg"></div> <!-- Header image -->

</div>

<!-- Page content -->

<main>

<center>

<div class="container-fluid px-0">

<div class="mx-auto tm-content-container">

<div class="row mt-3 mb-5 pb-3">

<div class="col-12">

<div class="mx-auto tm-about-text-container px-3">

<h2 class="tm-page-title mb-4 tm-text-primary">Enter values to predict the age of Abalone:</h2>

<form action="{{ url\_for('predict')}}" method="post">

<div class="form-group">

<label for="Sex">Sex of Abalone: Enter 2 for Male, 0 for Female, 1 for Infant</label>

<input type="text" class="form-control" name="Sex" id="Sex">

</div>

<div class="form-group">

<label for="Length">Length:</label>

<input type="text" class="form-control" name="Length" id="Length">

</div>

<div class="form-group">

<label for="Diameter">Diameter:</label>

<input type="text" class="form-control" name="Diameter" id="Diameter">

</div>

<div class="form-group">

<label for="Height">Height:</label>

<input type="text" class="form-control" name="Height" id="Height">

</div>

<div class="form-group">

<label for="Whole-weight">Whole-weight:</label>

<input type="text" class="form-control" name="Whole weight" id="Whole-weight">

</div>

<div class="form-group">

<label for="Shucked-weight">Shucked-weight:</label>

<input type="text" class="form-control" name="Shucked weight" id="Shucked-weight">

</div>

<div class="form-group">

<label for="Viscera-weight">Viscera-weight:</label>

<input type="text" class="form-control" name="Viscera weight" id="Viscera-weight">

</div>

<div class="form-group">

<label for="Shell-weight">Shell-weight:</label>

<input type="text" class="form-control" name="Shell weight" id="Shell-weight">

</div>

<div class="form-group" id="submit">

<center><button type="submit" class="btn btn-primary btn-block btn-large" id="submit" value="Predict"> Predict</button></center>

</div>

<div id="output">

<h3>{{ prediction\_text }}</h3>

</div>

</form>

</div>

</div>

</div>

</div>

</center>

<div class="parallax-window parallax-window-2" data-parallax="scroll" data-image-src="/static/img/contact-2.jpg"></div>

</div>

</main>

</div>

<script src="/static/js/jquery-3.4.1.min.js"></script>

<script src="/static/js/bootstrap.min.js"></script>

<script src="/static/js/parallax.min.js"></script>

</body>

</html>

**App.py**

# -\*- coding: utf-8 -\*-

"""

Created on Thu Feb 4 16:43:40 2021

@author: rincy

"""

import numpy as np

import pickle

from flask import Flask,request, render\_template

app=Flask(\_\_name\_\_,template\_folder="templates")

model = pickle.load(open('model.pkl', 'rb'))

@app.route('/', methods=['GET'])

def index():

return render\_template('home.html')

@app.route('/home', methods=['GET'])

def about():

return render\_template('home.html')

@app.route('/pred',methods=['GET'])

def page():

return render\_template('upload.html')

@app.route('/predict', methods=['GET', 'POST'])

def predict():

input\_features = [float(x) for x in request.form.values()]

features\_value = [np.array(input\_features)]

print(features\_value)

features\_name = ['Sex','Length','Diameter','Height','Whole weight','Shucked weight','Viscera weight','Shell weight']

prediction = model.predict(features\_value)

output=prediction[0]

print(output)

return render\_template('upload.html', prediction\_text='The predicted age of abalone is {} years.'.format((output+1.5)))

if \_\_name\_\_ == '\_\_main\_\_':

app.run(debug=False)