**Project of Introduction of Data Science**

**Delivered To :**

: : :Doctor Ayesha Hakim : : :

**Project Name : Abalone Dataset**

**Group 7 Members :**

* Muhammad Asfand yar wali (2021-uam-4632)
* Muaviya (2021-uam-4621)
* Zagham Rasool (2021-uam-4631)
* Mudassir Shabbir (2021-uam-4610)
* M.Tahir (2021-uam-4635

What is Abalone ?

Abalone is a common name for a group of small to very large sea snails, marine gastropod mollusks in the family Haliotidae.

These snails have a large, flattened, ear-shaped shell with a row of holes along the outer edge.

The inner surface of the shell is iridescent and is highly prized for its beauty and used in jewelry and decorative

items.

Abalone is also a popular food source in many cultures, particularly in Asia and North America.

The meat is considered a delicacy and is often used in sushi, salads, and other dishes.

Because of its popularity as a food and the high demand for its shells,

many species of abalone have been overfished and are now endangered.

About Dataset:

The Abalone dataset is a popular machine learning dataset that contains measurements of physical characteristics of abalone, a type of sea snail. The dataset is often used as a benchmark for regression tasks in machine learning.

The dataset includes the following features or variables for each abalone:

1. Sex: categorical variable (M for male, F for female, and I for infant)
2. Length: continuous variable representing the longest shell measurement in mm
3. Diameter: continuous variable representing the diameter of the shell in mm
4. Height: continuous variable representing the height of the shell in mm
5. Whole weight: continuous variable representing the weight of the whole abalone in grams
6. Shucked weight: continuous variable representing the weight of the meat in grams
7. Viscera weight: continuous variable representing the weight of the gut (after bleeding) in grams
8. Shell weight: continuous variable representing the weight of the shell in grams
9. Rings: integer variable representing the age of the abalone (the number of rings on the shell)

The goal of the dataset is to predict the age of the abalone (i.e., the number of rings)

based on its physical characteristics.

This is a regression task, as the target variable (age) is a continuous variable.

The dataset contains 4,177 instances and has been preprocessed to remove any missing values and to

transform the categorical variable (sex) into a set of binary variables (one-hot encoding).

**Project report**

Introduction

Problem statement

The objective of this project is to develop a machine learning model to predict the age of abalone based on its physical

characteristics. The model will be trained using the Abalone dataset

**Abalone Data set:**

The Abalone dataset is a popular machine learning dataset that contains measurements of physical characteristics of abalone, a type of sea snail. The dataset is often used as a benchmark for regression tasks in machine learning.

Objectives:

In this project, we will learn how to build a sequential model using TensorFlow in

Python to predict the age of an abalone. You may wonder what is an abalone.

The answer to this question is that it is a kind of snail. Generally, the age of an Abalone is determined by the physical examination of the abalone but this is a tedious task which is why we will try to build a regressor that can predict the age of abalone using some features which are easy to determine. You can download the abalone dataset from here.

**Importing Libraries and Dataset**

Python libraries make it easy for us to handle the data and perform typical and complex tasks with a single line of code.

* Pandas – This library helps to load the data frame in a 2D array format and has
* multiple functions to perform analysis tasks in one go.
* Numpy – Numpy arrays are very fast and can perform large computations in a very short time.
* Matplotlib/Seaborn – This library is used to draw visualizations.
* Sklearn – This module contains multiple libraries are having pre-implemented functions to perform tasks from data preprocessing to model development and evaluation.