

Data Collection and Preprocessing Phase Report

Date	7 July 2024
Team ID	6
Project Title	Abalone Age Prediction
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report, ensuring meticulous data curation and integrity for informed decision-making in every analysis and decision-making endeavour.

Data Collection Plan:

Section	Description
Project Overview	The machine learning project aims to predict Abalone Age based on its features. Using a dataset with features such Sex, Diameter, Length, Height, Whole weight, Shucked weight, Shell weight, Viscera weight, Rings, the objective is to build a model that accurately Predicts Abalone Age.
Data Collection Plan	<ul style="list-style-type: none"> ● Search for datasets related to Abalone Age. ● Prioritize datasets with diverse information.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, one of the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as Sex, Diameter, Length, Height, Whole weight, Shucked weight, Shell weight, Viscera weight, Rings.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size
Kaggle	Dataset contains features like Sex, Diameter, Length, Height, Whole weight, Shucked weight, Shell weight, Viscera weight, Rings.	https://www.kaggle.com/datasets/rodolfomendes/abalone-dataset	CSV	59 KB

