

Data Collection and Preprocessing Phase

Date	03 June 2024
Team ID	739676
Project Title	Harvesting Brilliance: A Taxanomic Tale of Pumpkin Seeds Varieties
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Template


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 endeavor.

Data Collection Plan Template

Section	Description
Project Overview	"Harvesting Brilliance: A Taxanomic Tale of Pumpkin Seeds Varieties" aims to classify and analyze different varieties of pumpkin seeds using machine learning techniques. The project seeks to identify distinguishing features of various seed types, predict their growth potential, and assist in breeding programs. Objectives include creating a comprehensive database of pumpkin seed varieties, developing a predictive model for seed classification, and providing insights into optimal growth conditions.
Data Collection Plan	Data for this project will be collected from multiple sources: - Agricultural Research Databases: Accessing datasets from agricultural research institutions and universities that specialize in pumpkin seed studies. - Field Surveys: Conducting surveys in pumpkin farms to collect data on seed varieties, growth conditions, and yield. - Online Repositories: Utilizing online databases like Kaggle and public government databases that provide agricultural data. - Experimental Data: Collecting data from controlled experiments in greenhouses and laboratories where different pumpkin seed varieties are grown and monitored.

Raw Data Sources Identified	<p>1. Agricultural Research Databases:
 - Source: University agricultural departments and research institutions.
 - Details: Comprehensive data on pumpkin seed varieties, including genetic information, growth patterns, and yield statistics
</p> <p>2. Field Surveys:
 - Source: Pumpkin farms across different regions va
 - Details: Real-world data on seed varieties, planting methods, soil conditions, climate impact, and harvest yields
</p> <p>3. Online Repositories:
 - Source: Kaggle, USDA (United States Department of Agriculture), FAO (Food and Agriculture Organization)
 - Details: Large datasets with information on pumpkin seed genetics, market availability, and global distribution
</p> <p>4. Experimental Data:
 - Source: Greenhouse and laboratory experiments
 - Details: Controlled data on growth conditions, nutrient requirements, pest resistance, and phenotypic characteristics of different pumpkin seed varieties.</p>
-----------------------------	---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------

Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Harvesting Brilliance: A Taxonomic Tale of Pumpkin Seeds Varieties	This dataset contains comprehensive information on various pumpkin seed varieties, including genetic details, growth patterns, and yield statistics collected from multiple research institutions.	 Pumpkin_Seeds_Data set.xlsx	CSV	50GB	Public