

## Data Collection and Preprocessing Phase

Date	15 July 2024
Team ID	740682
Project Title	Polycystic Ovary Syndrome Classification Using Machine Learning
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	"Decoding PCOS: A Comprehensive Classification Approach" aims to classify and analyze different manifestations of polycystic ovary syndrome (PCOS) using machine learning techniques. The project seeks to identify distinguishing features of PCOS subtypes, predict patient outcomes, and assist in personalized treatment plans. Objectives include creating a comprehensive database of PCOS patient profiles, developing a predictive model for PCOS classification, and providing insights into optimal management strategies.

Data Collection Plan	<ul style="list-style-type: none"> <li>• <b>Clinical Databases:</b> Accessing datasets from hospitals, clinics, and medical research institutions specializing in PCOS studies.</li> <li>• <b>Patient Surveys:</b> Conducting surveys among PCOS patients to collect data on symptoms, lifestyle factors, and treatment outcomes.</li> <li>• <b>Online Repositories:</b> Utilizing online databases like PubMed, government health databases, and academic repositories that provide medical data.</li> <li>• <b>Experimental Data:</b> Collecting data from clinical trials and controlled studies where PCOS patients are monitored and treated under specific protocols.</li> </ul>
----------------------	--

### Raw Data Sources Template

Source Name	Description	Location/URL	Format	Size	Access Permissions
Harvesting Brilliance: A Taxanomic 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