Data Collection plan and Raw Data Sources Report

Date	25 june 2024
Team ID	739849
Project Title	Doctor's annual salary 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 endeavor.

Data Collection plan:

Section	Description
Project Overview	The machine learning project aims to predict a doctor's annual salary based on various factors. Using a dataset with features such as age, gender, years of experience, specialty, education, location, and work characteristics, the objective is to build a model that accurately estimates annual salary. This will help healthcare institutions, doctors, and researchers make informed decisions.
Data Collection Plan	Search for datasets related to physician salaries, employment, and demographic information. \n - Utilize publicly available databases, surveys, market research reports, and online job portals. \n - Prioritize datasets with comprehensive and diverse information on various factors influencing salaries.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from public databases like the Bureau of Labor Statistics, Medscape Physician Compensation Report, Doximity Physician Compensation Report, state medical boards, and online job portals such as Glassdoor, Indeed, and LinkedIn. These sources provide extensive information on doctor salaries and related factors.

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises applicant details (age,gender, specialty, sub-specialties, education, practice settings, location,type of practice, regional demand, medicare)	"C:\Users\navyasri\OneDrive\Desktop\soumyasri project\Dataset"	CSV	15 kB	Public