Data Collection and Preprocessing Phase

Date	15 MARCH2024
TeamID	LTVIP2024TMID25011
ProjectTitle	EarlyPredictionOfChronicKidneyDiseaseUsingMachineLearning
MaximumMarks	2Marks

DataCollectionPlan&RawDataSourcesIdentificationTemplate

Elevate your data strategy with the Data Collection plan and the Raw Data Sources report,ensuringmeticulousdatacurationandintegrityforinformeddecision-makingineveryanalysisanddecision-making endeavor.

DataCollectionPlanTemplate

Section	Description
ProjectOverview	It is possible with such approaches to train an ML model andrecognize the signs of chronic kidney disease, however, it isnecessarytoincludethedoctorsintotheMLmodeldevelopmenttoconsider the real healthcare setting at the moment of the modelbuilding.
DataCollectionPlan	Identifythenecessaryfeatures(variables)relatedtoCKDdiagnosis(e.g.,serumcreatinine,bloodpressure, urineprotein). Determinethesample sizerequiredforreliablemodeltraining.
Raw DataSourcesIdentified	TherawdatasourcesforthisprojectincludedatasetsobtainedfromKaggle &UCI.Theprovidedsampledatarepresentsasubsetof thecollectedinformation,encompassingvariablessuchasage,bp,rbc ,pcc,hemo etc.. details for machine learninganalysis.

RawDataSourcesReport:

SourceName	Description	Location/URL	Format	Size	AccessPermissions
KaggleDataset	**This dataset contains detailed healthinformation for 1,659 patientsdiagnosed with Chronic KidneyDisease (CKD). The dataset includesdemographic details, lifestyle factors,symptoms, quality of life scores,environmental exposures, and healthbehaviors. Each patient is uniquelyidentifiedbyaPatientID,andthedataincludes a confidential columnindicatingthedoctorincharge.**	https://drive.google.com/file/d/1mPl4yaTKuKZ3017YfYC19Ni7Y964eCNI/view?usp=sharing	CSV	1.0MB	Public