

DataCollectionandPreprocessingPhase

Date	06-06-2024
TeamID	740055
Project Title	DETECTION OF PHISHING WEBSITES FROM URLS
Maximum Marks	2 Marks

Data Quality Report:

TheDataQualityReportwillsummarizedataqualityissuesfromtheselectedsource,including severity levels and resolution plans. It will aid in systematically identifying and rectifying data discrepancies.

Data Quality Report:

Data Source	Data Quality Issue	Severity	Resolution Plan
Kaggle Dataset	url name,having ip address ,url length,ssl,domain,subdomain, columns.	Moderate	Usemean/median imputation.

Kaggle Dataset	Categorical data in the dataset	Moderate	Encoding has to be done in the data.
----------------	---------------------------------	----------	--------------------------------------