

Data Collection and Preprocessing Phase

Date	21 June 2024
Team ID	739832
Project Title	Startup Prophet
Maximum Marks	2 Marks

Data Collection Plan & Raw Data Sources Identification Report:

Elevate the 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	"Startup Prophet leverages machine learning to predict success rates of tech startups based on various metrics, aiding investors in decision-making processes. The project aims to offer actionable insights through data analysis and predictive modeling, optimizing investment strategies in the tech industry."
Data Collection Plan	<ul style="list-style-type: none"> Search for datasets related to is_commerce, is_otherstate, has_VC, has_angel, has_roundA, has_roundB, has_roundC, has_roundD, is_top500, funding_rounds, milestones, relationships Prioritize datasets with more success rates.

Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, the popular platforms for data science competitions and repositories. The provided sample data represents a subset of the collected information, encompassing variables such as, is_commerce, is_otherstate, has_VC, has_angel, has_roundA, has_roundB, has_roundC, has_roundD, is_top500, funding_rounds, milestones, relationships details for machine learning analysis.
-----------------------------	--

Raw Data Sources Report:

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle Dataset	The dataset comprises details of (is_commerce, is_otherstate, has_VC, has_angel, has_roundA, has_roundB, has_roundC, has_roundD, is_top500, funding_rounds, milestones, relationships)	https://www.kaggle.com/datasets/manishkc06/start-up-success-prediction	CSV	66kb	Public