Data Collection and Preprocessing Phase

Date	31 June 2024
Team ID	739853
Project Title	Software 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.

Data Collection Plan Report

Section	Description

Project Overview	The project aims to develop a predictive model to accurately estimate salaries for software professionals based on relevant factors such as job title, location, experience, education, skills, and company size. The goal is to provide a fair and competitive salary range, reducing uncertainty and bias in salary decisions. The model should generalize well across different locations, companies, and industries, and be interpretable and explainable to stakeholders.
Data Collection Plan	The dataset which contains a set of features through which salary can be calculated, is to be collected.
Raw Data Sources Identified	The raw data sources for this project include datasets obtained from Kaggle, the popular platform for data science competitions and repositories. The provided datasets contain the collected

	information, variables such as global active power, global reactive power, global intensity, voltage, sub metering values for machine learning analysis.

Raw Data Sources Report

Source Name	Description	Location/URL	Format	Size	Access Permissions
Kaggle	The dataset contains the measurements of electric power consumption.	Software Salary Prediction (Kaggle Dataset)	txt	2MB	Public