

## Data Collection and Preprocessing Phase

|               |                                                                                                 |
|---------------|-------------------------------------------------------------------------------------------------|
| Date          | 9 July 2024                                                                                     |
| Team ID       | SWTID1720104839                                                                                 |
| Project Title | <b>Human Resource Management:<br/>Predicting Employee Promotions<br/>Using Machine Learning</b> |
| Maximum Marks | 2 Marks                                                                                         |

### Data Collection Plan & Raw Data Sources Identification Template

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 Template

| Section          | Description                                                                                                                                                                                                                                                                                          |
|------------------|------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Project Overview | <p>The machine learning project builds the employee promotion prediction model based on the employee information provided.</p> <p>Using the dataset that features department, Education, No of trainings, Age, Previous year rating, length of service, KPI, Awards won, Average training score.</p> |

|                             |                                                                                                                                                                                                  |
|-----------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
|                             | The objective of the project is to build a model that classifies whether the employee is promoted or not.                                                                                        |
| Data Collection Plan        | Searching for datasets in Kaggle and identifying the relevant dataset needed.                                                                                                                    |
| Raw Data Sources Identified | <p>Kaggle : popular platform for hosting dataset and sharing knowledge.</p> <p>Kaggle provides a vast collection of public datasets.</p> <p>You can search for terms like Employee Promotion</p> |

### Raw Data Sources Template

| Source Name | Description                                                                                                                                            | Location/URL                                                                                                                                                                        | Format | Size  | Access Permissions |
|-------------|--------------------------------------------------------------------------------------------------------------------------------------------------------|-------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--------|-------|--------------------|
| Kaggle      | <ul style="list-style-type: none"> <li>Employee ID</li> <li>Department</li> <li>Education</li> <li>No of trainings</li> <li>Age of Employee</li> </ul> | <a href="https://drive.google.com/file/d/14eQR1VWHwuomPaXdKuIkZEpSstvav8Db/view?usp=sharing">https://drive.google.com/file/d/14eQR1VWHwuomPaXdKuIkZEpSstvav8Db/view?usp=sharing</a> | CSV    | 3.6MB | Public             |

|  |                                                                                                                                                                        |  |  |  |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|
|  | <ul style="list-style-type: none"><li>• Previous year rating</li><li>• Length of service</li><li>• KPI</li><li>• Awards won</li><li>• Average training score</li></ul> |  |  |  |  |
|--|------------------------------------------------------------------------------------------------------------------------------------------------------------------------|--|--|--|--|