

AI-DRIVEN EXPLORATION AND PREDICTION OF COMPANY REGISTRATION TRENDS WITH THE REGISTRAR OF COMPANIES

PHASE-3: LOADING AND PREPROCESSING DATASET

INTRODUCTION: The design phase of the "AI-Driven Exploration and Prediction of Company Registration Trends with the Registrar of Companies" project is a pivotal step in the development journey. This phase is where we formulate the overarching structure and strategy that will guide the implementation of AI-driven solutions for exploring and predicting company registration trends based on data sourced from the Registrar of Companies. At this juncture, the project team collaborates to determine the key elements that will shape the project's direction. This involves decisions on data handling, AI model selection, system architecture, user interface, and the overall workflow. It's a moment where we transition from abstract concepts to concrete plans.

DATASET:

Dataset can be taken from the below link,

Dataset link: <https://tn.data.gov.in/resource/company-master-data-tamil-nadu-upto-28th-february-2019>

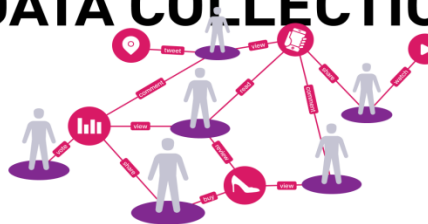
The development phase for a project like "AI-Driven Exploration and Prediction of Company Registration Trends with Register of Companies" typically involves several key steps:

1. **Planning:** Define the project scope, objectives, and goals. Identify the data sources (Register of Companies data), tools, and technologies needed for the project.

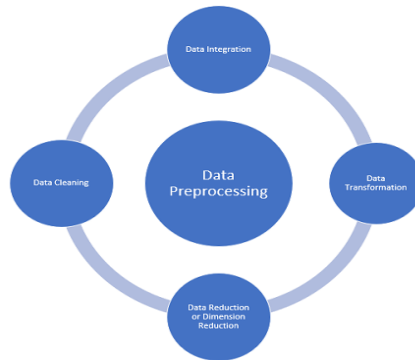


2. **Data Collection:** Gather the relevant data from the Register of Companies, which might include information on new company registrations, historical trends, and any other related data sources.

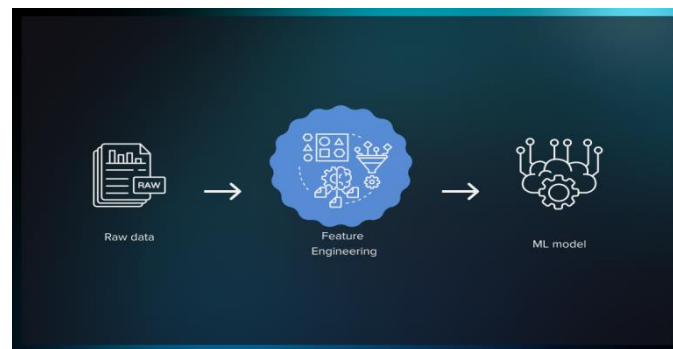
DATA COLLECTION



3. **Data Preprocessing:** Clean and preprocess the data to ensure it's in a usable format. This may involve data cleaning, normalization, and handling missing values.



4. **Feature Engineering:** Create relevant features and variables that will be used by the AI models for exploration and prediction. This might include variables such as registration dates, company types, geographic locations, and more.



5. **Model Selection:** Choose the appropriate AI and machine learning models for exploration and prediction. Common choices might include regression models, time series analysis, or deep learning models, depending on the nature of the prediction task.
6. **Training and Testing:** Train the selected models on historical data and test their performance to ensure they can make accurate predictions.
7. **AI Integration:** Implement the AI models into the project's workflow to enable real-time or batch predictions.
8. **Visualization and Exploration:** Develop data visualization tools and techniques to explore trends and insights in the data. This could include charts, graphs, and interactive dashboards.
9. **Prediction and Forecasting:** Use the trained AI models to make predictions about future company registration trends. Continuously update and retrain models as new data becomes available.

10. Validation and Evaluation: Continuously monitor the performance of the AI models and refine them as needed. Use appropriate metrics to evaluate the accuracy and effectiveness of predictions.

11. Deployment: Deploy the AI-driven system in a production.



PROGRAM:

```
#import necessary libraries
import pandas as pd
import numpy as np
from sklearn.model_selection import train_test_split
from sklearn.ensemble import RandomForestRegressor
from sklearn.metrics import mean_squared_error
import matplotlib.pyplot as plt

# Load the CSV data
data = pd.read_csv('/content/drive/MyDrive/DataGov_TamilNadu.csv',
encoding = "ISO-8859-1")

# Data Preprocessing and Feature Engineering (customize this part based
on your data)
data['DATE_OF_REGISTRATION'] =
pd.to_datetime(data['DATE_OF_REGISTRATION'])
data['year'] = data['DATE_OF_REGISTRATION'].dt.year
```