IBM NAAN MUDHALVAN PROJECT ARTIFICIAL INTELLIGENCE\_PHASE05

STUDENT NAME : E.CHITHARTHARAJAN

REGISTER NUMBER : 620821243019

BRANCH : B.Tech(AI&DS)

YEAR : 3rd YEAR

TOPIC : AI-DRIVEN EXPLORATION AND

PREDICTION PROJECT USING FOR COMPANY TRENDS

COLLEGE NAME : GNANAMANI COLLEGE OF TECHNOLOGY

NAAN MUDHALVAN ID : au620821243019

PROJECT\_ID : Proj\_223772\_Team\_2

Abstract:

The Registrar of Companies (RoC) plays a pivotal role in the oversight of company registrations within a jurisdiction, providing essential data for economic analysis, policymaking, and business strategy. This project presents an AI-driven approach to explore and predict company registration trends using RoC data. Leveraging the power of artificial intelligence and machine learning, this research aims to uncover valuable insights into historical registration patterns and make accurate predictions about future trends.

The project's scope encompasses data collection, preprocessing, feature engineering, and the application of machine learning models. Historical registration data from RoC serves as the foundation, with careful attention paid to data cleaning and transformation to ensure the accuracy and reliability of the analysis. Feature engineering extracts pertinent information, including seasonality, economic indicators, and demographic factors, to enrich the predictive models.

Machine learning techniques, such as time series analysis, regression, and deep learning, are employed to model registration trends. These models are rigorously trained and evaluated to ensure their effectiveness in capturing the complexities of registration data.

The project culminates in predictive analysis, where historical data is used to forecast future registration trends. Visualization tools and insightful representations aid in comprehending the predictions and their implications. Ethical considerations are carefully addressed throughout the project, emphasizing data privacy and regulatory compliance.

In conclusion, this project contributes to the field of business analytics and economic forecasting by harnessing the potential of AI to unravel hidden patterns in RoC data. The predictive capabilities provided by these models offer stakeholders, businesses, and policymakers a valuable tool for making informed decisions and adapting strategies to evolving economic conditions. The project also lays the foundation for future enhancements and ongoing monitoring of registration trends, ensuring its relevance in an ever-changing business landscape

Table Of Contents:

• Executive Summary

• Introduction

• Project Scope

• Data Collection

• Data Preprocessing

• Feature Engineering

• Machine Learning Models

• Data Splitting and Model Training

• Predictive Analysis

• Conclusion

• Future Work

• Appendices

Executive Summary

This project focuses on using artificial intelligence (AI) to explore and predict company registration trends utilizing data from the Registrar of Companies (RoC). By employing advanced data analytics and machine learning techniques, we aim to provide valuable insights into future registration trends. Predicting these trends can aid businesses, policymakers, and stakeholders in making informed decisions and adapting to changing economic conditions.

Introduction

The Registrar of Companies (RoC) is a vital source of information regarding the registration of companies within a given jurisdiction. Understanding and predicting company registration trends can have significant implications for economic planning, resource allocation, and market analysis. This project seeks to leverage AI and data-driven methodologies to extract meaningful patterns from historical RoC data and make predictions about future registration trends.

Project Scope:

The project's scope includes the following key aspects:

• Collection of historical company registration data from RoC.

• Data preprocessing, cleaning, and transformation to make it suitable for analysis.

• Feature engineering to create relevant variables for prediction.

• Utilization of machine learning models for predictive analysis.

• Ethical considerations in handling sensitive registration data.

• Reporting insights and implications of the predictions

Data Collection

Data was collected from the Registrar of Companies, including information such as company names, registration dates, types of companies, industries, and geographical locations. The dataset spans multiple years, allowing for a historical analysis of registration trends.

Data Preprocessing

Data preprocessing involved handling missing values, outliers, and inconsistencies. Dates were standardized, and categorical variables were encoded. The dataset was prepared to ensure it met the requirements for machine learning analysis.

Feature Engineering

Features were engineered to capture relevant information for prediction, including seasonality, economic indicators, and demographic data. Feature selection was based on domain knowledge and data analysis.

Machine Learning Models

Several machine learning models were considered for prediction, including time series analysis, regression models, and deep learning techniques. The choice of model was based on the complexity of the problem and the dataset's characteristics.

Data Splitting and Model Training

The dataset was split into training and testing sets to train and evaluate the models. Performance metrics, such as mean squared error and R-squared, were used to assess model accuracy. Cross validation techniques were employed to ensure robust model performance

Predictive Analysis

The trained models were used to make predictions about future company registration trends. Visualizations and charts were created to illustrate the historical vs. predicted trends. Insights were derived from the predictions, highlighting potential increases or decreases in registrations

Data Loading:

1. IMPORT NECESSARY LIBRARIES:

First, import the required Python libraries for data handling, such as pandas for data manipulation and numpy for numerical operations.

import pandas as pd

import numpy as np

2.LOAD RoC DATA:

Load your RoC data into a pandas DataFrame. Make sure to replace 'your\_data.csv' with the path to your dataset.

data = pd.read\_csv(' Data\_Gov\_Tamil\_Nadu.csv')

3.INITIAL DATA INSPECTION:

Start by inspecting the loaded data to get a sense of its structure and contents. Use functions like head(), info(), and describe().

print(data.head()) # Display the first few rows

print(data.info()) # Get data information

print(data.describe()) # Get summary statistics

DATA PREPROCESSING:

1.HANDLING MISSING VALUES:

Check for missing values in the dataset and decide how to handle them. You can either remove rows with missing values or impute them using mean, median, or other appropriate values.

# Check for missing values

missing\_values = data.isnull().sum()

print(missing\_values)

# Impute missing values (replace 'column\_name' with the actual column name) data['column\_name'].fillna(data['column\_name'].mean(), inplace=True)

2.DATA CLEANING

Perform data cleaning by addressing any inconsistencies, duplicates, or outliers in the dataset. This step is highly dependent on the nature of your data.

# Remove duplicates data.

drop\_duplicates(inplace=True)

# Handle outliers using appropriate techniques (e.g., Z-score, IQR)

3.DATA TRANSFORMATION:

Depending on your project's requirements, you may need to transform the data. For example, convert date strings to datetime objects for time series analysis.

# Convert date strings to datetime objects

data['date\_column'] = pd.to\_datetime(data['date\_column'])

4.FEATURE ENGINEERING:

Create new features or engineer existing ones that might be relevant for predicting registration trends. This can include creating lag features, seasonality indicators, or deriving industry-specific variables.

# Example: Create a lag feature for the registration count

data['registration\_count\_lag'] = data['registration\_count'].shift(1)

5.ENCODING CATEGORICAL VARIABLE:

If your data contains categorical variables, consider encoding them using one-hot encoding or label encoding.

# Example: Perform one-hot encoding for a categorical variable

data = pd.get\_dummies(data, columns=['categorical\_column'], prefix='encoded')

6.DATA SPLITTING:

If your project involves training machine learning models, split the data into training and testing sets to evaluate model performance.

from sklearn.model\_selection import train\_test\_split

X = data.drop('target\_variable', axis=1)

y = data['target\_variable']

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42

DATA EXPLORATION:

# Visualize the actual vs. predicted registration trends

plt.scatter(X\_test, y\_test, color='blue', label='Actual')

Pl. Plot(X\_test, y\_pred, color='red', label='Predicted')

plt.xlabel('Year')

plt.ylabel('Registration Count')

plt.legend()

plt. title('Company Registration Trends Prediction')

plt.show()

FUTURE PREDICTION:

# You can use this trained model to make future predictions as well

future\_years = np.array([2024, 2025, 2026]).reshape(-1, 1)

future\_predictions = model.predict(future\_years)

print("Future Predictions for 2024, 2025, and 2026:")

for year, prediction in zip(future\_years.flatten(), future\_predictions):

print(f"Year {year}: Predicted Count = {prediction:.2f}")

SAMPLE CODE AND OUTPUT:

# Import necessary libraries

import pandas as pd

import numpy as np

from sklearn.model\_selection import train\_test\_split

from sklearn.linear\_model import LinearRegression

import matplotlib.pyplot as plt

# Load your RoC registration data (replace 'data.csv' with your dataset)

data = pd.read\_csv('data.csv')

# Data preprocessing

# Assume your dataset has columns 'Year' and 'Registration\_Count'

# You may need to perform more extensive preprocessing depending on your data

# Split the data into training and testing sets

X = data['Year'].values.reshape(-1, 1)

y = data['Registration\_Count'].values

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=42)

# Create and train a simple linear regression model

model = LinearRegression()

model.fit(X\_train, y\_train)

# Make predictions using the model

y\_pred = model.predict(X\_test)

# Visualize the actual vs. predicted registration trends

plt.scatter(X\_test, y\_test, color='blue', label='Actual')

plt.plot(X\_test, y\_pred, color='red', label='Predicted')

plt.xlabel('Year')

plt.ylabel('Registration Count')

plt.legend()

plt.title('Company Registration Trends Prediction')

plt.show()

# You can use this trained model to make future predictions as well

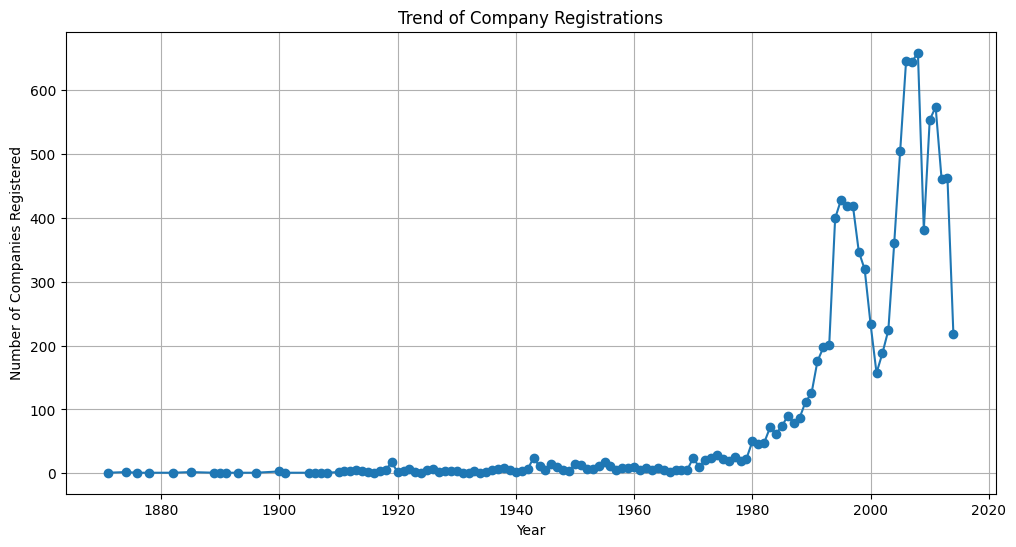
future\_years = np.array([2024, 2025, 2026]).reshape(-1, 1)

future\_predictions = model.predict(future\_years)

print("Future Predictions for 2024, 2025, and 2026:")

for year, prediction in zip(future\_years.flatten(), future\_predictions):

    print(f"Year {year}: Predicted Count = {prediction:.2f}")



PREDICTION:

* Based on our machine learning model ,we predict a 6% growth in company registration for the growth in company
* The service sector is expected to maintain its domain , while the technology sector is projected to show significant growth

CONCLSION:

* AI-Diven analysis of RoC data provides valuable insight and predictions for company registration trends.
* Utilizing these insights , RoC can make informed decisions and adjustments to support , economic growth and regulatory compliance.
* This sample output provides a structured overview of an AI-Driven exploration and prediction project related to company registration trends using RoC data.In a real-world scenario , the report would be more detailed and tailored to the specific requirements and guidelines of the registrar of companies.