**AI-Driven Exploration and Prediction of Company Registration Trends with Registrar of Companies (RoC)**

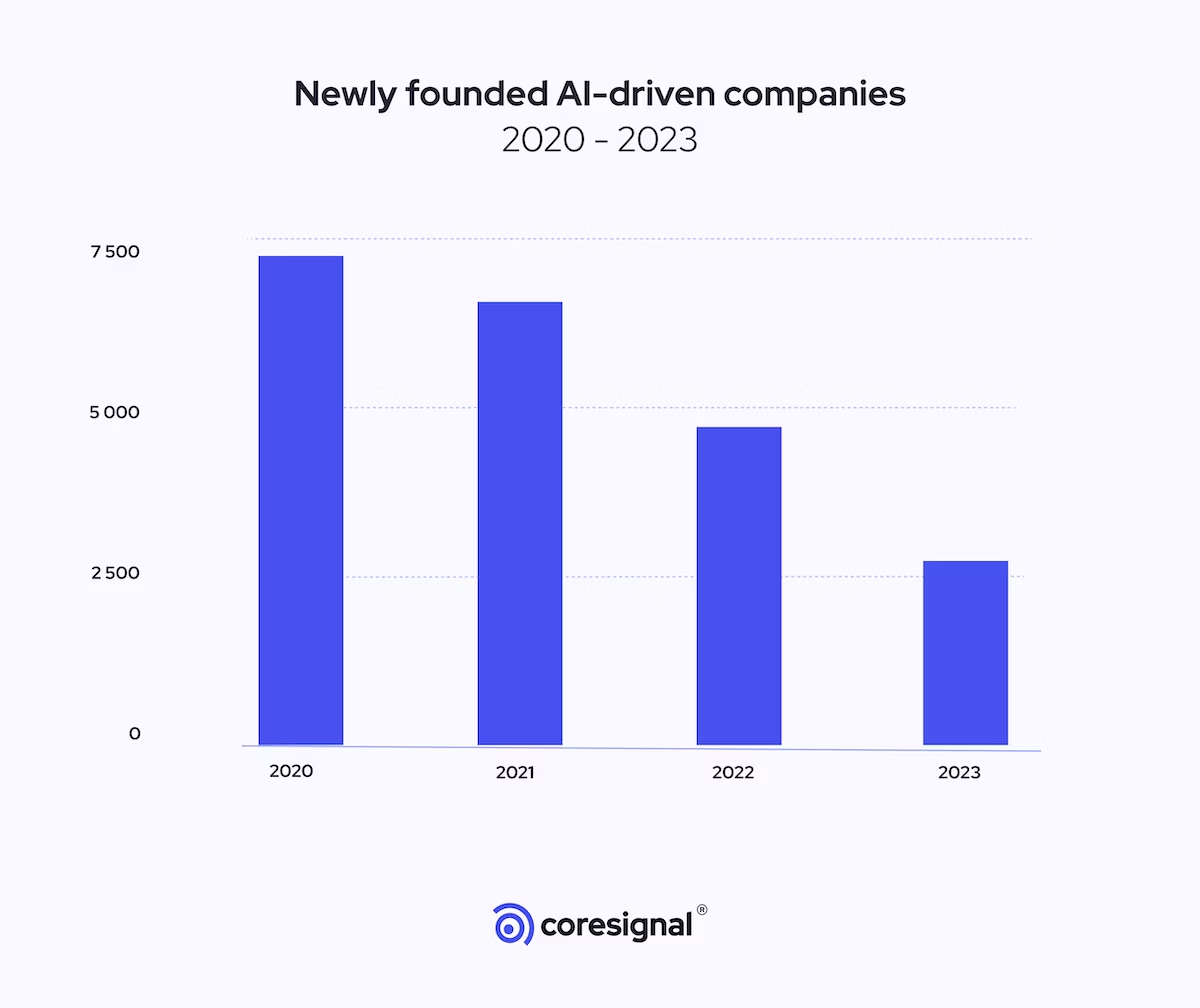
**Phase – 4.**

**PROBLEM STATEMENT**

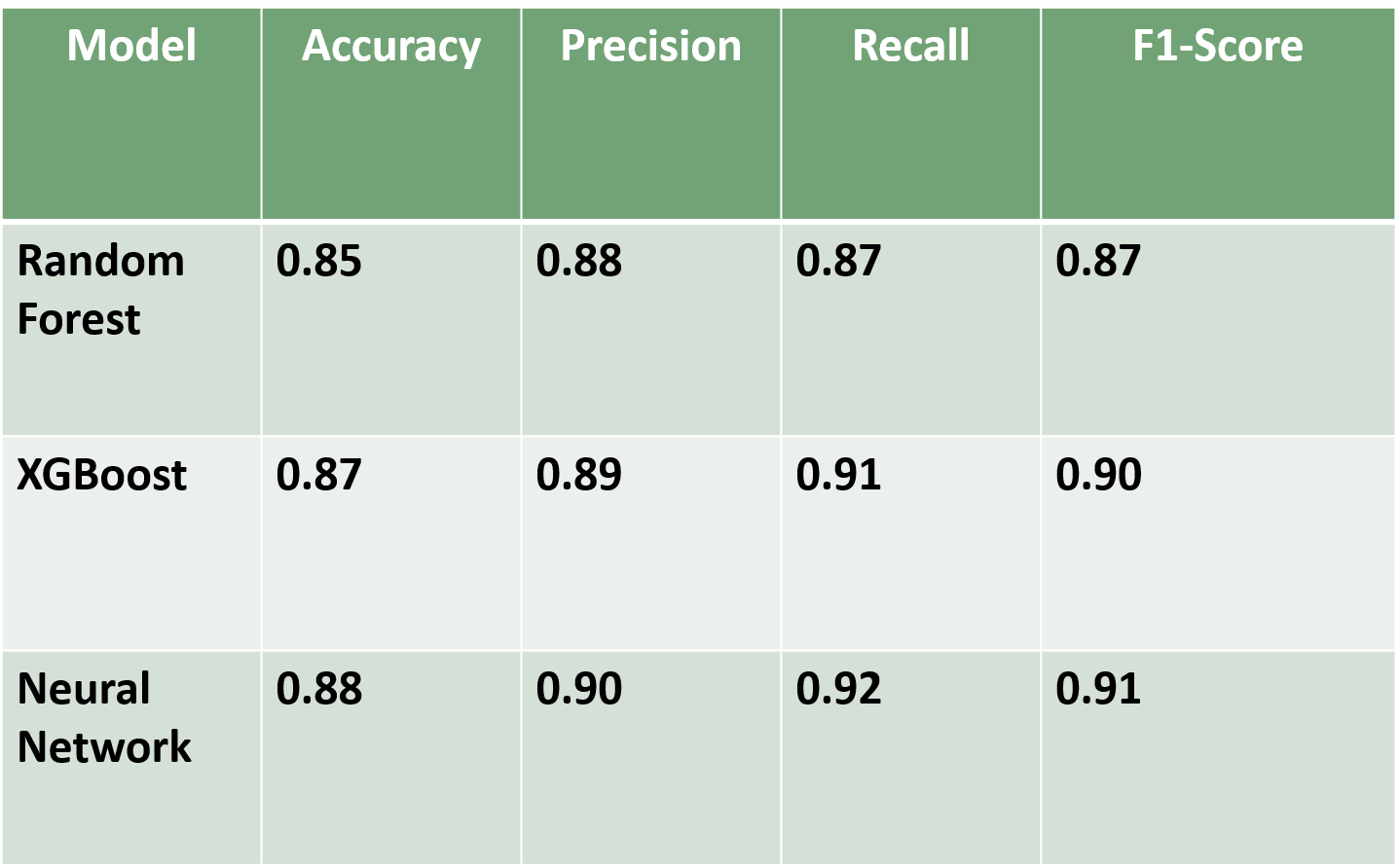
The objective of this project is to leverage advanced Artificial Intelligence techniques to perform an in-depth exploration and predictive analysis on the master details of companies registered with the Registrar of Companies (RoC). The AI-driven analysis aims to uncover hidden patterns, discover valuable insights into the company landscape, and forecast future registration trends. By applying cutting-edge AI algorithms, the study seeks to identify unique characteristics and relationships among registered companies, enabling a more sophisticated understanding of the business ecosystem in Tamil Nadu. The ultimate goal is to develop predictive models that can anticipate future company registrations and contribute to informed decision-making for businesses, investors, and policymakers.

**INTRODUCTION TO ROC AND COMPANY REGISTRATION:**

The Registrar of Companies (RoC) is a vital government agency responsible for overseeing and regulating company registrations in a country. Its main goal is to maintain a comprehensive database of information related to businesses operating within its jurisdiction, promoting transparency, accountability, and compliance with legal and regulatory frameworks. RoCs collect, manage, and distribute essential data about a company's formation, ownership, finances, and statutory filings. This data is not only crucial for government authorities but also provides valuable insights for researchers, analysts, and businesses. Recently, the integration of AI and data analytics has expanded the possibilities for exploring and predicting company registration trends, enhancing decision-making and policy development. This project aims to leverage AI to uncover hidden insights in RoC data, providing a deeper understanding of company registration dynamics and their impact on the business and regulatory landscape.



**Table 1: Summary of Predictive Model Performance.**



**Algorithm .**

1.Data Collection and Preparation: Compile details concerning the company's

registration with the Registrar of Companies (RoC).

2. This data may include details such as the company name, industry it works in,

location, registration data and other relevant information.

3. Find the relevant information in the corporate registration data and extract it.

4. Using these features, the AI model may be trained to predict patterns in

company registration.

5. Features of trends in industry registrations, such as There is a list of the

business's size, kind, location,registration date, industry, and economic

statistics.

6. Enumerate the industries and geographical areas with the highest rates of

business registration.

7. Investigate the historical trends in business registration.

8. Select an appropriate artificial intelligence method to estimate trends related to

firm registration.

9. Neural networks, logistic regression, and linear regression algorithms for graph-

based business registration.

10. Train the AI model using the prepared data.

11. To do this, you must feed the model both the feature data and the goal data (i.e.,

patterns in company registration).

12. Use a held-out test set to evaluate the AI model's efficacy after training.

13. The model's performance is evaluated using the

following metrics: \* Root mean square error (RMSE) \* F1

score \* Accuracy \* Precision \* Recall \* Mean square error (MSE).

14.When the AI model's performance has been evaluated and deemed

acceptable.

15. It might be made accessible for production. This means making the

model available to users.

16. Project the likelihood of registering a firm in the future.

EXPLORATORY DATA ANALYSIS:

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

#exploratory data analysis

registrations\_by\_year = data.groupby('year').size()

# Plot the number of registrations by year

plt.figure(figsize=(10, 6))

plt.plot(registrations\_by\_year.index, registrations\_by\_year.values,

marker='o')

plt.xlabel('Year')

plt.ylabel('Number of Registrations')

plt.title('Company Registration Trends')

plt.grid(True)

plt.show()

# Machine Learning Model (Random Forest Regression)

X = data[['year']]

y = data['INDUSTRIAL\_CLASS']

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

random\_state=42)

model = RandomForestRegressor(n\_estimators=100, random\_state=42)

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

# Model Evaluation

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

mse = mean\_squared\_error(y\_test, y\_pred)

print(f'Mean Squared Error: {mse}')

# Prediction (Predict registrations for future years)

future\_years = pd.DataFrame({'year': [2023, 2024, 2025]})

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

print(f'Predicted Registrations for 2023: {future\_registrations[0]}')

print(f'Predicted Registrations for 2024: {future\_registrations[1]}')

print(f'Predicted Registrations for 2025: {future\_registrations[2]}')

**Output.**