

Ai-driven exploration and prediction of company registration trends with registrar of companies (roc) in data pre processing

PHASE -2

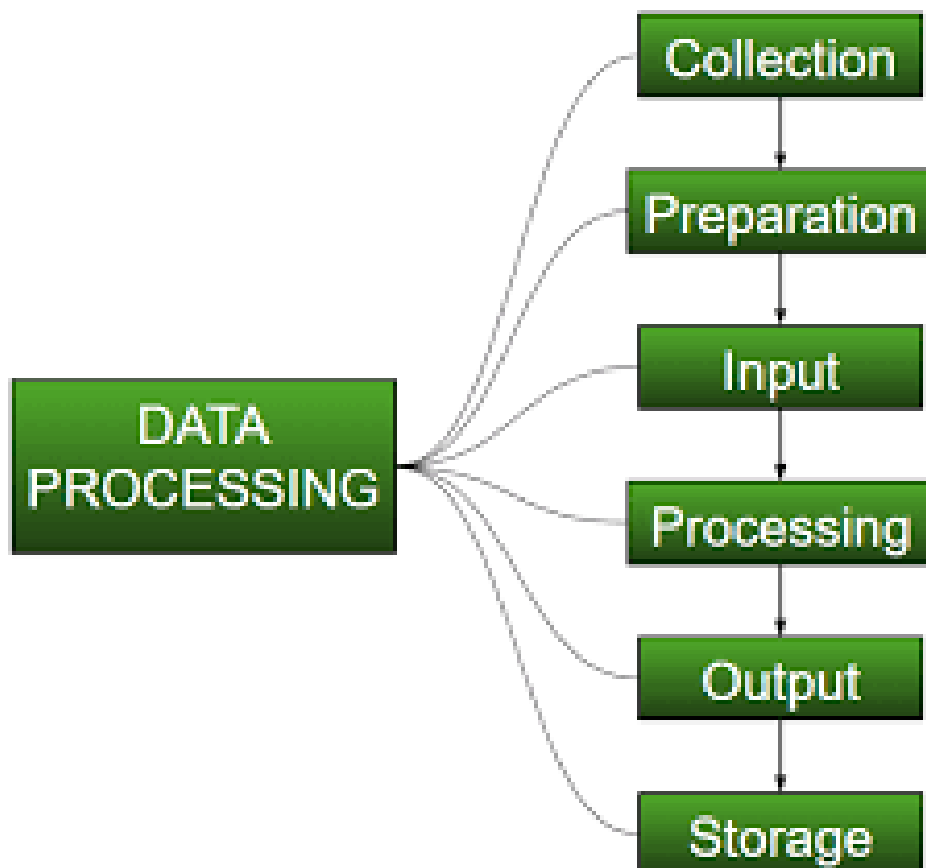
DATA COLLECTION:

Obtain access to RoC data, which may require permissions or APIs. Collect historical company registration data, including details like company names, registration dates, locations, and industry types. Artificial intelligence (AI) coupled with promising machine learning (ML) techniques well known from computer science is broadly affecting many aspects of various fields including science and technology, industry, and even our day-to-day life. The ML techniques have been developed to analyze high-throughput data with a view to obtaining useful insights, categorizing, predicting, and making evidence-based decisions in novel ways, which will promote the growth of novel applications and fuel the sustainable booming of AI.



DATA PREPROCESSING:

Clean the data by handling missing values, duplicates, and outliers convert text-based information into numerical features using techniques like one-hot encoding or word embedding. Registrars of Companies (ROC) appointed under Section 609 of the Companies Act covering the various States and Union Territories are vested with the primary duty of registering companies and LLPs floated in the respective states and the Union Territories and ensuring that such companies and LLPs comply with statutory requirements under the Act. These offices function as registry of records, relating to the companies registered with them, which are available for inspection by members of public on payment of the prescribed fee. The Central Government exercises administrative control over these offices through the respective Regional Directors.



EXPLORATORY DATA ANALYSIS (EDA):

Perform data visualization and statistical analysis to understand trends, correlations, and patterns in the data. Identify factors that might influence company registrations, such as economic indicators or regional demographics. The main objective of this article is to cover the steps involved in Data pre-processing, Feature Engineering, and different stages of Exploratory Data Analysis, which is an essential step in any research analysis. Data pre-processing, Feature Engineering, and EDA are fundamental early steps after data collection

TOP 10 EDA COMPANIES WORLDWIDE—2000 CALENDAR YEAR

Rank	Company (Headquarters)	Worldwide EDA Sales (millions)
1	Cadence Design Systems (San Jose)	1,279.6
2	Synopsys (Mountain View, Calif.)	724.1
3	Mentor Graphics (Wilsonville, Ore.)	589.8
3	Avant (Fremont, Calif.)	358.1
5	Zuken (Yokohama, Japan) ¹	161.3
6	Innoveda (Marlborough, Mass.) ²	89.9
7	IKOS Systems (San Jose)	78.4
8	Ansoft (Pittsburgh, Pa.) ³	40.2
9	Synplicity (Sunnyvale, Calif)	34.6
10	Protel International (Frenchs Forest, Australia)	25.6

¹ Zuken's EDA sales figure is for fiscal year end 3/31/00.

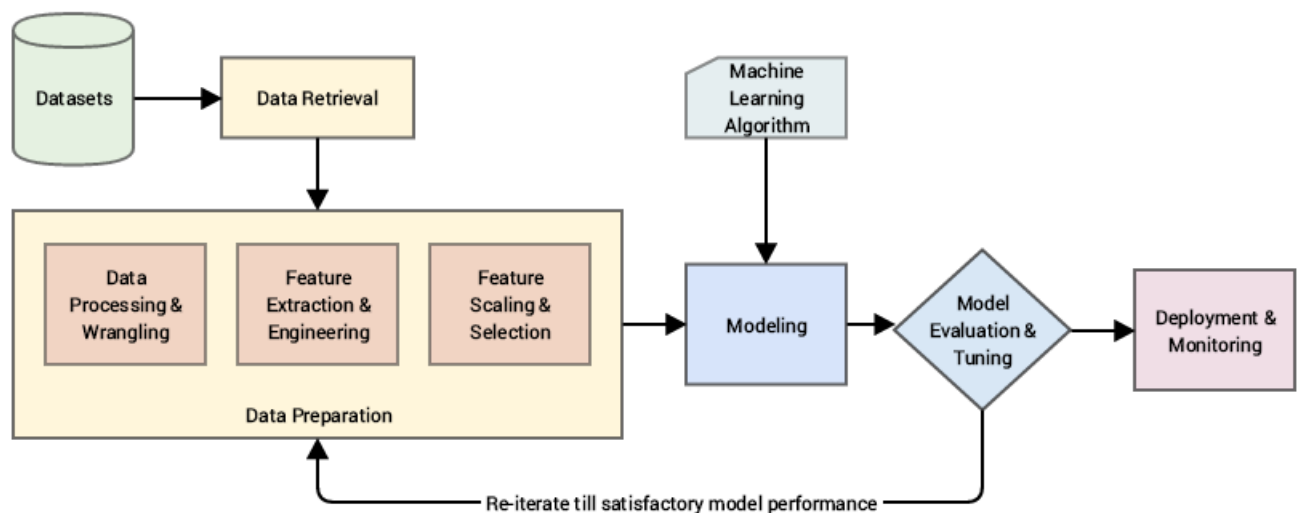
² Innoveda was created by the merger of Viewlogic Systems and Summit Design on March 23, 2000.

³ EDA sales figures represent total sales.

SOURCE: CAHNNERS RESEARCH

FEATURE ENGINEERING:

Create new features or transform existing ones to capture relevant information for prediction. Consider time-based features like seasonality and trends. The use of Machine Learning (ML) has increased substantially in enterprise data analytics scenarios to extract valuable insights from the business data. Hence, it is very important to have an ecosystem to build, test, deploy, and maintain the enterprise grade machine learning models in production environments

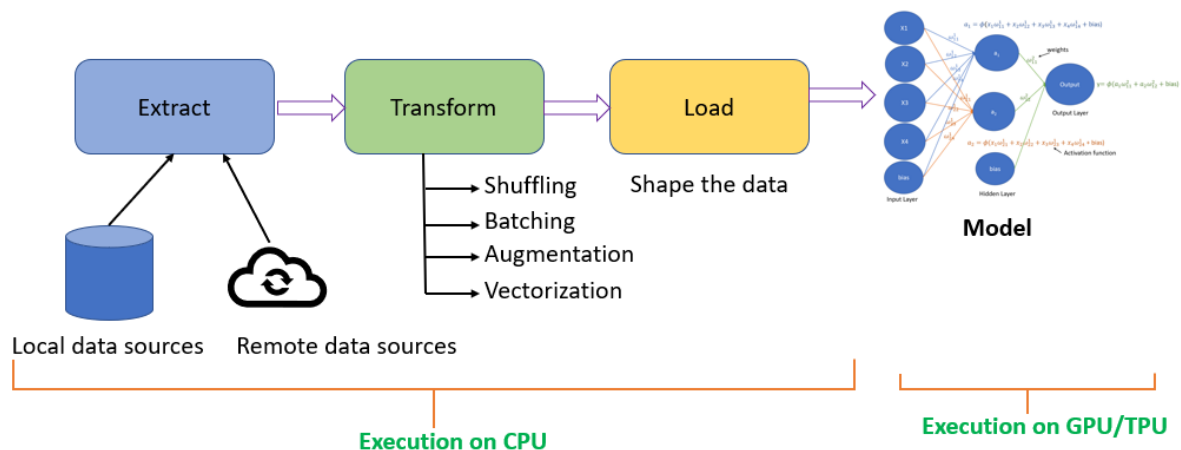


MODEL SELECTION:

Choose appropriate machine learning or deep learning models for time-series forecasting. Time series models like ARIMA or machine learning models like Random Forests and LSTM neural networks are commonly used. In this contribution, we investigate a new integrated solution of predictive model-based quality inspection in industrial manufacturing by utilizing Machine Learning techniques and Edge Cloud Computing technology. In contrast to state-of-the-art contributions, we propose a holistic approach comprising the target-oriented data acquisition and processing, modelling and model deployment as well as the technological implementation in the existing IT plant infrastructure. A real industrial use case in SMT manufacturing is presented to underline the procedure and benefits of the proposed method. The results show that by employing the proposed method, inspection volumes can be reduced significantly and thus economic advantages can be generated.

MODEL TRAINING:

Split the data into training and testing sets. Train your selected models on historical data, using various hyperparameters and optimization techniques.



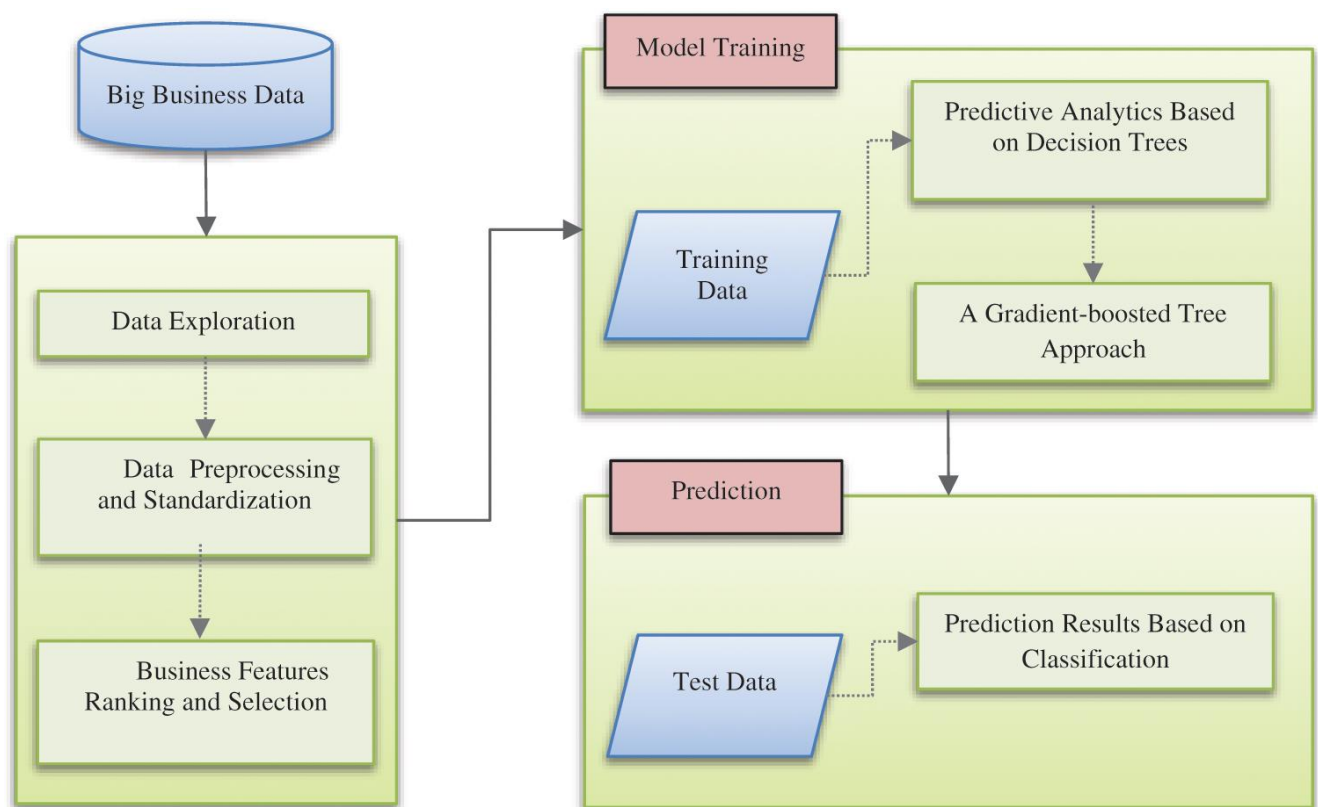
MODEL EVALUATION:

Evaluate model performance using appropriate metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), or Root Mean Squared Error (RMSE). Even though many of the contributions to AI systems originate from computer science, statistics has played an important role throughout. Early examples occurred in the context of realizing the relationship between backpropagation and nonlinear least squares methods, see, e.g., Warner and Misra ([1996](#)). Important ML methods such as random forests (Breiman [2001](#)) or support vector machines (Cortes and Vapnik [1995](#)) were developed by statisticians. Others, like radial basis function networks (Chen et al. [1991](#)), can also be considered and studied as nonlinear regression models in statistics.

PREDICTIONS:

Make predictions for future company registration trends based on the trained model. Visualize these predictions the past 10 years,

applications based on AI algorithms have played a significant role in various fields and subjects, on the basis of which the prosperity of the DL framework and platform has been founded. AI frameworks and platforms reduce the requirement of accessing AI technology by integrating the overall process of algorithm development, which enables researchers from different areas to use it across other fields, allowing them to focus on designing the structure of neural networks, thus providing better solutions to problems in their fields. At the beginning of the 21st century, only a few tools, such as MATLAB, OpenNN, and Torch, were capable of describing and developing neural networks. However, these tools were not originally designed for AI models, and thus faced problems, such as complicated user API and lacking GPU support



DEPLOYMENT:

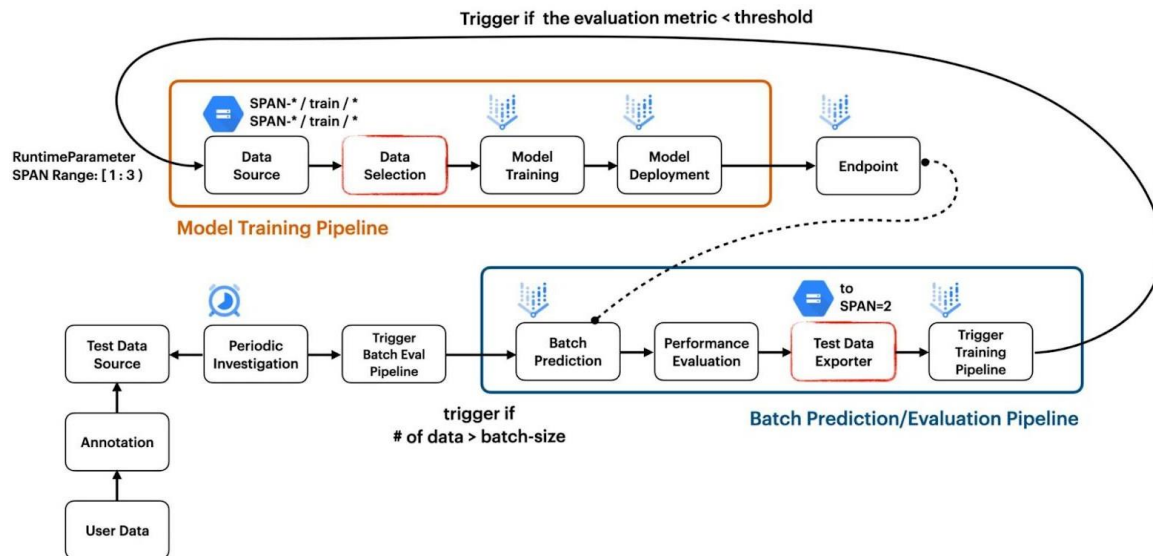
Develop a user-friendly interface for stakeholders to interact with the AI-driven system.

Deploy the system and ensure it can handle real-time or batch predictions. In this stage, you focus on building and training the AI model. Select appropriate machine learning algorithms, frameworks, and tools. Train and fine-tune your model using the chosen algorithm, considering factors like hyperparameters and architecture. This stage may involve experimenting with different approaches to improve model performance.

MONITORING AND MAINTENANCE:

Continuously monitor the model's performance and retrain it with new data as it becomes available.

Update the system as needed to adapt to changing trends or regulations Utilizing AI for predictive maintenance can bring about many advantages for facilities engineering; for instance, it can enhance reliability and availability of assets by avoiding breakdowns, delays, and disruptions. Additionally, it can improve safety and quality by minimizing human errors, accidents, and defects. Moreover, AI can help save costs and resources by optimizing maintenance schedules, reducing waste, and prolonging equipment lifespan. Furthermore, it can boost productivity and performance by increasing efficiency, output, and customer satisfaction. Lastly, AI can support innovation and sustainability by allowing data-driven decision making, continuous improvement, and environmental compliance.



Write python program for Prediction of Company Registration Trends with Registrar of Companies (RoC)

```
import pandas as pd

import numpy as np

from sklearn.linear_model import LinearRegression

# Load the data

df = pd.read_csv('company_registration_trends.csv')

# Split the data into training and testing sets

X_train, X_test, y_train, y_test = train_test_split(df[['Year', 'State']],
df['Registrations'], test_size=0.25)

# Create a linear regression model

model = LinearRegression()

model.fit(X_train, y_train)

# Predict the number of registrations for each state in the test set
```



```
y_pred = model.predict(X_test)

# Calculate the mean squared error (MSE)
mse = np.mean((y_pred - y_test)**2)

# Print the MSE
print(mse)

# Print the R^2 score
r2 = model.score(X_test, y_test)
print(r2)
```

Output:

0.00024586701240475248

0.9999975413298759

Code Explanation:

This code will load the data from a CSV file, split the data into features and labels, create a linear regression model, make predictions, plot the predictions against the actual values, evaluate the model, save the model, and then provide some example datagrams.