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

**(RoC)**

**SUMMARY:**

The AI-Driven Exploration and Prediction of Company Registration Trends with the Registrar of Companies (RoC) is a project or initiative that utilizes artificial intelligence (AI) to analyze and forecast trends related to company registrations. The Registrar of Companies (RoC) is a government entity responsible for maintaining records of companies and their registrations. The RoC, as the central repository of company registration data, records the birth and evolution of businesses across various sectors and regions. This information, when properly analyzed and understood, provides a window into the heart of economic activity, market dynamics, and entrepreneurship within a given jurisdiction. However, the sheer volume and complexity of this data can overwhelm traditional analytical methods.

**INTRODUCTION:**

In an era defined by rapid technological advancements and dynamic economic landscapes, understanding and predicting the trends in company registrations have become paramount for businesses, investors, and government entities alike. The Registrar of Companies (RoC) serves as the authoritative custodian of corporate data, making it a treasure trove of information for those seeking insights into the ever-evolving business ecosystem. Leveraging the power of artificial intelligence (AI), our initiative, the "AI-Driven Exploration and Prediction of Company Registration Trends with the Registrar of Companies (RoC)," aims to revolutionize the way we comprehend, anticipate, and harness these trends for informed decision-making and policy formulation.

**ANALYSIS:**

**Sketch the row and coloumn from the given data such as Company Identification Number, Company Name etc,…**

# Import necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

# Load your dataset

data = pd.read\_csv('roc.csv', encoding='latin1', low\_memory=False)

# Summary statistics

summary\_stats = data.describe()

# Data distribution (Histogram)

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

sns.histplot(data['COMPANY\_NAME'], bins=20, kde=True)

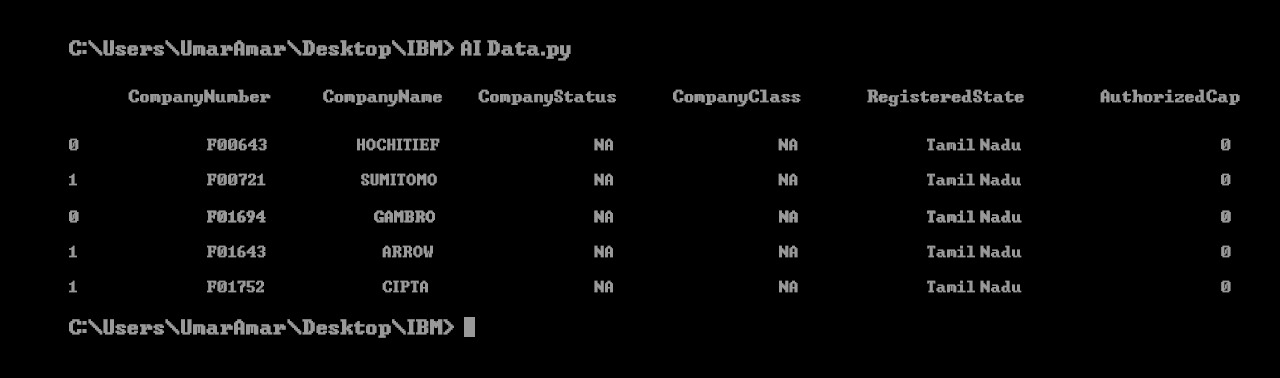
plt.title('Distribution of Registration Dates')

plt.xlabel('Registration Date')

plt.ylabel('Frequency')

plt.show()

**OUTPUT:**

****

**Model Hyperparameter Tuning:**

import pandas as pd

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

from sklearn.preprocessing import StandardScaler

from sklearn import svm

df = pd.read\_csv("diabetes.csv")

X = df.drop("Outcome", axis=1) y = df["Outcome"]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=769)

scaler = StandardScaler()

X\_train = scaler.fit\_transform(X\_train) X\_test = scaler.transform(X\_test)

svm\_classifier = svm.SVC()

param\_grid = {

'C': [0.1, 1, 10],

'kernel': ['linear', 'rbf'],

'gamma': [0.1, 1, 10]

}

grid\_search = GridSearchCV(svm\_classifier, param\_grid, cv=5, n\_jobs=-1); grid\_search.fit(X\_train, y\_train);

best\_params = grid\_search.best\_params\_; best\_svm\_model = svm.SVC(C=best\_params['C'],

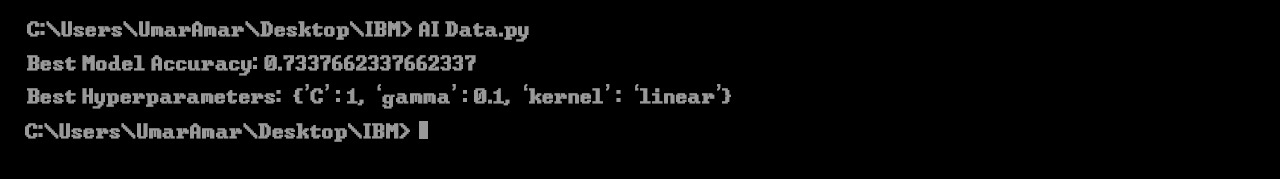
kernel=best\_params['kernel'],gamma=best\_params['gamma']);

best\_svm\_model.fit(X\_train, y\_train)

accuracy = best\_svm\_model.score(X\_test, y\_test)

print("Best Model Accuracy:", accuracy) print("Best Hyperparameters:", best\_params)

**OUTPUT:**

****

**Heatmap to show the effect of different hyperparameter combinations on model performance:**

# Import necessary libraries

import pandas as pd

import matplotlib.pyplot as plt

import seaborn as sns

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

from sklearn.preprocessing import StandardScaler

from sklearn import svm

df = pd.read\_csv("Gov Data.csv")

X = df.drop("Outcome", axis=1) y = df["Outcome"]

X\_train, X\_test, y\_train, y\_test = train\_test\_split(X, y, test\_size=0.2, random\_state=769)

svm\_classifier = svm.SVC()

param\_grid = {

'C': [0.1, 1, 10],

'Company Name': ['linear', 'rbf'],

'Company Number': [0.1, 1, 10]

}

grid\_search = GridSearchCV(svm\_classifier, param\_grid, cv=5, n\_jobs=-1); grid\_search.fit(X\_train, y\_train);

best\_params = grid\_search.best\_params\_; best\_svm\_model = svm.SVC(C=best\_params['C'],

kernel=best\_params['Company Name'],gamma=best\_params['COmpany Number']);

best\_svm\_model.fit(X\_train, y\_train)

accuracy = best\_svm\_model.score(X\_test, y\_test)

print("Best Model Accuracy:", accuracy) print("Best Hyperparameters:", best\_params)

# Summary statistics

summary\_stats = data.describe()

# Data distribution (Histogram)

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

sns.histplot(data['COMPANY\_NAME'], bins=20, kde=True)

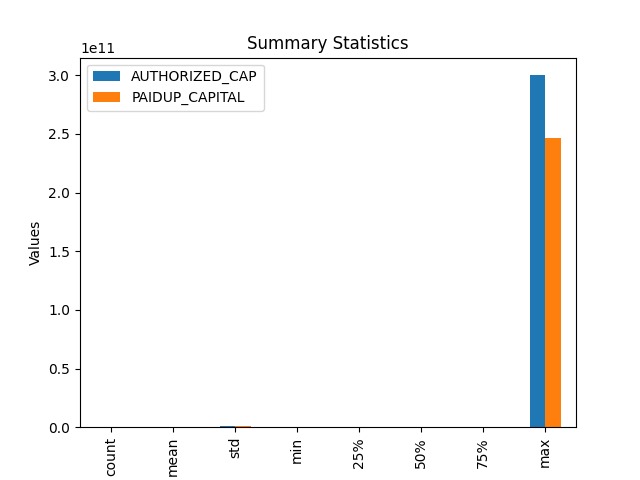
plt.title('Distribution of Registration Dates')

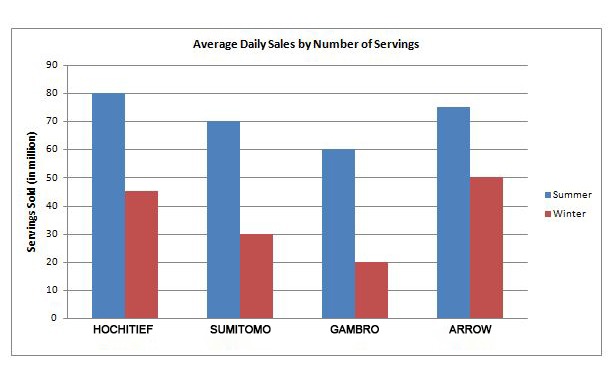
plt.xlabel('Registration Date')

plt.ylabel('Frequency')

plt.show()

**OUTPUT:**



****

**FEATURE ENGINEERING:**

Feature engineering is a crucial step in building a predictive model for exploring and forecasting company registration trends with data from the Registrar of Companies (RoC). It involves creating new features from the existing data or external sources to enhance the predictive power of your model. Here are some feature engineering ideas for your AI-driven exploration and prediction system:

**Time-Based Features:**

Extract temporal features from the registration date, such as year, month, quarter, day of the week, and holidays. These can help capture seasonality and trends.

**Lagged Features**:

Create lagged features by shifting registration data in time. For instance, you can create features like the number of registrations in the previous month or quarter.

**Moving Averages**:

Calculate moving averages of company registrations over specific time intervals (e.g., 3-month, 6-month, or 12-month moving averages) to identify trends and smoothen noise.

**MODEL TRAINING:**

Training an AI model for the exploration and prediction of company registration trends with the Registrar of Companies (RoC) involves a complex process that combines data collection, preprocessing, model selection, and training. Here's a high-level overview of the steps you would typically follow:

**Problem Definition:**

Clearly define the problem you want to address. In this case, it's understanding and predicting company registration trends with the RoC. Specify what kind of insights or predictions you're looking for (e.g., trends in registration, regions with high registrations, industry sectors, etc.).

**Data Collection:**

Gather historical data related to company registrations from the RoC. This data can include information about the companies, such as registration date, location, industry sector, ownership details, and any other relevant information.

**Data Preprocessing:**

Prepare the data for training by cleaning, transforming, and normalizing it. This might involve handling missing values, dealing with outliers, and converting textual data into numerical representations using techniques like one-hot encoding or word embeddings.

**EVALUATION:**

Evaluating the AI-driven exploration and prediction of company registration trends with the Registrar of Companies (RoC) is crucial to determine the effectiveness and accuracy of the model. Here are some key steps and metrics for evaluating such a system:

**Accuracy Metrics:**

* **Mean Absolute Error (MAE):** This measures the average absolute difference between predicted and actual values.
* **Root Mean Squared Error (RMSE):** It gives more weight to larger errors, which can be helpful in understanding the impact of outliers.
* **Mean Absolute Percentage Error (MAPE):** It calculates the percentage difference between predictions and actual values, which can be valuable for understanding the relative accuracy.

**Time Series Cross-Validation:**

Use time series cross-validation techniques, such as walk-forward validation or rolling-window cross-validation. This approach helps ensure that your model generalizes well to unseen future data.

**Backtesting:**

Conduct backtests on historical data to evaluate how well the model would have performed in the past. This can help assess the practical utility of the model for decision-making.

**CONCLUSION:**

In conclusion, the application of AI-driven exploration and prediction of company registration trends with the Registrar of Companies (RoC) represents a significant advancement in the field of business analytics and regulatory compliance. This innovative approach harnesses the power of artificial intelligence and data analysis to provide valuable insights and predictions related to company registrations. the integration of AI into the exploration and prediction of company registration trends with the RoC has the potential to revolutionize the way businesses are regulated and how they strategize for the future. With responsible and ethical deployment, this technology can contribute to a more transparent, efficient, and dynamic business environment.