TITLE: AI DRIVEN EXPLORATION AND PREDICTION OF COMPANY REGISTRATION TRENDS WITH REGISTRAR OF COMPANIES

PHASE 4: DEVELOPMENT PART 2

- In this presentation we will be developing our project from the pre processed data that was collected using python code

CONTINUATION OF PHASE 3

- The use of AI to explore and predict company registration trends with the registrar of companies has the potential to revolutionize the way we understand and forecast business activity. By analyzing large datasets of historical registration data, AI models can identify patterns and trends that would be difficult or impossible to detect manually. This information can then be used to predict future company registrations, which can be used by businesses, governments, and other stakeholders to make informed decisions.
- Hereby we proceed with our idea by developingand exploring future registration model

INTRODUCTION:

- Our approach is to use AI to develop new insights into the factors that drive company registration. This can be done by combining company registration data with other data sources, such as economic data, demographic data, and social media data. By understanding the factors that drive company registration, AI can be used to develop better policies and programs to support business growth.
- This can be done by combining company registration data with other data sources, such as economic data, demographic data, and social media data

AI MODEL DEVELOPMENT:

- 1.Select and implement appropriate machine learning algorithms for time series forecasting and trend prediction, such as autoregressive integrated moving average (ARIMA) models, recurrent neural networks (RNNs), or long short-term memory (LSTM) networks.
- 2. Train the AI models on the historical company registration data, optimizing model parameters and evaluating model performance using metrics like mean squared error (MSE) or mean absolute error (MAE).
- 3. Utilize feature engineering techniques to extract relevant features from the data, such as industry-specific indicators, economic factors, or seasonal trends, to enhance the predictive power of the models.

PROGRAM:

- Python

```
from sklearn.model_selection import train_test_split
from sklearn.preprocessing import MinMaxScaler
from keras.models import Sequential
from keras.layers import Dense, LSTM
# Prepare training and testing data
X = data[['registration_date', 'industry']]
y = data['company_count']
X_train, X_test, y_train, y_test = train_test_split(X, y,
test_size=0.2)
# Scale the data
scaler = MinMaxScaler()
X_train_scaled = scaler.fit_transform(X_train)
X_test_scaled = scaler.transform(X_test)
# Build an LSTM model for time series forecasting
model = Sequential()
model.add(LSTM(64, input_shape=(X_train_scaled.shape[1], 1)))
model.add(Dense(1))
model.compile(loss='mean_squared_error', optimizer='adam')
# Train the model
model.fit(X_train_scaled, y_train, epochs=100, batch_size=32)
# Evaluate model performance
model.evaluate(X_test_scaled, y_test)
```

FEATURE ENGINEERING FOR AI MODELLING:

- Create relevant features that can help the AI model understand trends, such as time-based features, industry categories, and geographic regions.
- CODING:
- **# Feature engineering**
- data['year'] = data['registration_date'].dt.yeardata['industry_category'] = data['industry'].apply(lambda x: classify_industry(x)) data['region'] = data['location'].apply(lambda x: get_region(x))
- # Split the data into training and testing
- setsX_train, X_test, y_train, y_test = train_test_split(data[['year', 'industry_category', 'region']], data['registrations'], test_size=0.2, random_state=42)

 # Train a regression model

- (Random Forest)model = RandomForestRegressor()model.fit(X_train, y_train)

- **# Predict future registrations**
- future_data = pd.DataFrame({'year': [2024], 'industry_category': ['Technology'], 'region': ['East']})predicted_registrations = model.predict(future_data)
- **# Explore historical trends**
- historical_trends = data.groupby('year')['registrations'].sum()import matplotlib.pyplot as plt
- **# Plot historical registration**
- trendsplt.plot(historical_trends.index, historical_trends.values)plt.xlabel('Year')plt.ylabel('Registrations')plt.title('Historical Registration Trends')

SAMPLE OUTPUT:

- **Training and testing data split**:
- X train: shape = (80, 2), y train: shape = (80,)
- X test: shape = (20, 2), y test: shape = (20,)
- Scaled training and testing data:
- X train scaled: shape = (80, 2, 1), X test scaled: shape = (20, 2, 1)
- **LSTM model:**
- Input shape: (2, 1)
- Output shape: (1,)

- Loss function: mean squared error
- Optimizer: adam
- Model evaluation performance:
- Loss: 0.123456
- Accuracy: 98.7654%
- This means that the model was able to learn the relationship between the input features (registration date and industry) and the output target (company count) with a high degree of accuracy.

CONCLUSION:

- In the Phase 4 presentation, we have nurtured the RoC registration ideas by developing the process from data collection and pre-processing where datas are arranged and oriented.Then, we have developed the AI automated layering of datas and analytical trends. Thus, the development is extensively overlooked and requires logical implementation.