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

TEAM LEADER

510521104305:kabilan R

PHASE-3:DEVLOPEMENT PART-I

OBJECTIVIE:

The problem is to perform an AI-driven exploration and predictive analysis on the master details of companies registered with the Registrar of Companies (RoC). The objective is to uncover hidden patterns, gain insights into the company landscape, and forecast future registration trends.

PHASE 3: Devlopement Part-1

In this phase you will begin building your project. Please refer below the requirements technology wise:

Dataset Link: https://tn.data.gov.in/resource/company-master-data-tamil-nadu-upto-28th-february-2019

TECHNOLOGY-1

AI PROJECT

1. Dataset Loading and Preprocessing

- Load the dataset into your preferred environment (e.g., Python, R).
- Handle missing values, duplicates, and outliers.
- Perform data cleaning and transformation to make it suitable for analysis.

2. Exploratory Data Analysis (EDA)

- Generate summary statistics to understand the dataset's basic characteristics.
- Create data visualizations (e.g., histograms, scatter plots) to explore relationships between variables.
- Identify patterns, trends, and potential insights in the data.

3. Feature Engineering

- Select relevant features for modeling.
- Create new features or transform existing ones as needed.
- Explain the rationale behind feature selection and engineering.

4. Model Building and Training

- Select an appropriate machine learning or AI model for the problem (e.g., regression, classification, clustering).
- Split the dataset into training and testing sets.
- Train the model and tune hyperparameters if necessary.

5. Model Evaluation

- Evaluate the model's performance using suitable metrics (e.g., accuracy, F1-score, RMSE).
- Provide insights into model strengths and weaknesses.

6. Documentation

- **Create a document that includes:**
  - Introduction: Briefly explain the project's objective.
  - Data Loading and Preprocessing: Describe how the dataset was prepared.
  - EDA: Present findings from the exploratory data analysis.
  - Feature Engineering: Explain feature selection and engineering.
  - Model Building and Training: Describe the chosen model and its performance.
  - Model Evaluation: Discuss the evaluation results.
  - Conclusion: Summarize key takeaways and potential areas for improvement.

7. Sharing for Assessment

- Share the document with relevant stakeholders for assessment and feedback.

ADS PROJECT

1. Data Synthesis

- If you are using Artificial Data Synthesis (ADS), describe the process of generating synthetic data.
- Explain the methods and algorithms used for data synthesis.

2. Data Preprocessing for Synthesized Data

- Preprocess the synthesized data as you would for real data, including handling missing values and outliers.

3. EDA for Synthesized Data

- Apply EDA techniques to the synthesized data to understand its characteristics.
- Analyze and visualize the synthetic dataset to check for patterns and anomalies.

4. Model Building Using Synthetic Data

- Use the synthesized data to build machine learning models.
- Explain the choice of models and objectives for using synthetic data.

5. Model Evaluation Using Real and Synthesized Data

- Compare model performance when using both real and synthesized data.
- Evaluate the effectiveness of synthetic data in improving model performance.

6. Documentation

- Create a document with the following sections:
  - Introduction: Explain the purpose of the ADS project.
  - Data Synthesis: Describe the data synthesis process.
  - Data Preprocessing: Explain how synthesized data was prepared for analysis.
  - EDA for Synthesized Data: Present findings from the synthetic data.
  - Model Building and Evaluation: Discuss model performance using real and synthetic data.
  - Conclusion: Summarize key insights and the impact of using synthetic data.

7. Sharing for Assessment

- Share the document with relevant stakeholders for assessment and feedback.

Both projects should be well-documented, including code, methodologies, and findings. The document should serve as a comprehensive report of your work, explaining your analysis and outcomes clearly for assessment.

TECHNOLOGY-2

DATA ANALYSIS AND VISUALIZATION USING IBM COGNOS (DAC)

1. Data Loading and Preprocessing

- Import the dataset into IBM Cognos.
- Clean and preprocess the data as needed, addressing missing values, duplicates, and outliers.
- Describe the data cleaning and preprocessing steps you performed.

2. Data Exploration and Analysis in IBM Cognos

- Use IBM Cognos tools to perform various data analysis tasks, including but not limited to:
    - Descriptive statistics: Summary statistics for numeric variables.
    - Data profiling: Understanding data distributions and patterns.
    - Data quality assessment: Identifying data anomalies and issues.
    - Data segmentation: Grouping data based on relevant attributes.

3. Data Visualization in IBM Cognos

- Create visualizations to represent data insights:
    - Bar charts, line charts, and pie charts for categorical data.
    - Histograms and box plots for numeric data.
    - Scatter plots for understanding relationships.
    - Geographic maps if location data is available.
- Utilize appropriate visualization types to convey information effectively.

4. Advanced Analysis and Reporting

- Use IBM Cognos capabilities for advanced analysis, such as:
    - Time series analysis for trends and seasonality.
    - Predictive modeling if applicable, using IBM Cognos' predictive analytics features.
    - Creating dashboards and reports for presenting insights.

5. Document Creation

- Create a detailed document that includes the following sections:
    - Introduction: Explain the project's objective and the dataset used.
    - Data Loading and Preprocessing: Describe how the data was loaded and prepared.
    - Data Analysis: Present findings from data exploration and analysis.
    - Data Visualization: Showcase the visualizations created.
    - Advanced Analysis and Reporting: Discuss any advanced analysis and reports.
    - Conclusion: Summarize key insights and the value of using IBM Cognos for data analysis.

6. Sharing for Assessment

- Share the document and any relevant IBM Cognos reports or dashboards with your intended audience for assessment.

DOCUMENT STRUCTURE (SAMPLE)

1. Introduction

- Provide an overview of the project's objectives and the dataset used.

2. Data Loading and Preprocessing

- Explain the process of loading and preprocessing the dataset using IBM Cognos.

3. Data Analysis

- Describe the insights obtained through data exploration and basic analysis.

4. Data Visualization

- Showcase the visualizations created in IBM Cognos to convey data insights.

5. Advanced Analysis and Reporting

- Discuss any advanced analysis performed using IBM Cognos and the resulting reports or dashboards.

6. Conclusion

- Summarize key insights and the benefits of using IBM Cognos for data analysis.

By following this structure, you can create a comprehensive document that highlights the steps you've taken and the insights you've gained from your data analysis and visualization using IBM Cognos. Sharing this document with relevant stakeholders will facilitate assessment and feedback for your DAC project.

Creating an IoT (Internet of Things) project involves deploying IoT devices and developing Python scripts to meet specific project requirements. Here's an outline of the key steps and a structure for the project document for your IoT project: