AI-DrivenExplorationandPredictionofCompanyRegistrationTrendswith RegistrarofCompanies (RoC)

TEAMMEMBERS510521104009:R.EZHILARASAN

PHASE 3:DEVLOPEMENTPART-I

OBJECTIVIE:

The problem is to perform an AI-driven exploration and predictiveanalysis on the master details of companies registered with the Registrarof Companies (RoC). The objective is to uncover hidden patterns, gaininsights into the company landscape, and forecast future registrationtrends.

PHASE3:DevlopementPart-1

In thisphaseyou willbegin buildingyourproject.Pleasereferbelowtherequirementstechnologywise:

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

TECHNOLOGY-1

AI PROJECT

1. DatasetLoadingandPreprocessing

- Loadthedatasetintoyourpreferredenvironment(e.g.,Python,R).
- Handlemissingvalues,duplicates,and outliers.
- Performdatacleaningand transformationtomakeit suitableforanalysis.

2. ExploratoryDataAnalysis(EDA)

- Generatesummarystatisticstounderstandthedataset'sbasicchar acteristics.
- Createdatavisualizations(e.g.,histograms,scatterplots)toexpl orerelationshipsbetweenvariables.
- Identifypatterns, trends,andpotentialinsightsinthedata.

3. FeatureEngineering

- Selectrelevantfeaturesformodeling.
- Createnewfeaturesortransformexistingonesasneeded.
- Explaintherationalebehind featureselectionandengineering.

4. ModelBuildingandTraining

- Select anappropriatemachinelearningorAImodelfortheproblem(e. g., regression, classification,clustering).
- Splitthedatasetintotrainingandtestingsets.
- Trainthemodelandtunehyperparametersifnecessary.

5. ModelEvaluation

- Evaluatethemodel'sperformanceusingsuitablemetrics(e.g.,accuracy,F1-score,RMSE).
- Provideinsightsintomodelstrengthsandweaknesses.

6. Documentation

- **Createa documentthatincludes:**
  - Introduction:Brieflyexplaintheproject'sobjective.
  - DataLoadingandPreprocessing:Describehowthedatasetwas prepared.
  - EDA:Present findingsfromtheexploratorydataanalysis.
  - FeatureEngineering:Explainfeatureselectionandengineering.
  - ModelBuildingandTraining:Describethechosenmodelanditsperformance.
  - ModelEvaluation:Discusstheevaluationresults.
  - Conclusion:Summarizekeytakeawaysandpotentialareasfori mprovement.

7. SharingforAssessment

- Share the document with relevant stakeholders for assessment andfeedback.

ADS PROJECT

1. DataSynthesis

- If you areusingArtificialDataSynthesis(ADS),describetheprocessofgeneratingsyntheticdata.

- Explain themethodsandalgorithmsusedfordatasynthesis.

2. DataPreprocessingforSynthesizedData

- Preprocessthesynthesizeddataasyou would forrealdata,includinghandlingmissingvaluesandoutliers.

3. EDAfor SynthesizedData

- ApplyEDAtechniquestothesynthesizeddatatounderstanditschara cteristics.
- Analyze and visualize the synthetic dataset to check for patternsandanomalies.

4. ModelBuildingUsingSyntheticData

- Usethesynthesizeddatatobuild machinelearningmodels.
- Explainthechoiceofmodelsand objectivesforusingsyntheticdata.

5. ModelEvaluationUsingRealandSynthesizedData

- Comparemodelperformancewhenusingboth realandsynthesizeddata.
- Evaluatetheeffectivenessofsynthetic datainimprovingmodelperformance.

6. Documentation

- Createadocumentwiththefollowingsections:
  - Introduction:ExplainthepurposeoftheADSproject.
  - DataSynthesis:Describethedatasynthesisprocess.

DataPreprocessing:Explainhowsynthesizeddatawaspre
paredforanalysis.

-
- EDAforSynthesizedData:Present findingsfromthesyntheticdata.
  Model Building and Evaluation: Discuss model performanceusingrealandsyntheticdata.
- Conclusion:Summarizekeyinsightsandtheimpactofusingsynthe ticdata.

7. SharingforAssessment

- Sharethedocumentwithrelevantstakeholdersforassessmentandfeedback.

Both projects should be well-documented, including code,methodologies, and findings. The document should serve as acomprehensivereportofyourwork,explainingyouranalysisandoutcomes c learlyforassessment.

TECHNOLOGY-2
DATAANALYSISANDVISUALIZATIONUSING IBMCOGNOS(DAC)

1. DataLoadingandPreprocessing

Import thedatasetintoIBM Cognos.

- Clean andpreprocessthedataasneeded, addressingmissingvalues,duplicates,andoutliers.
- Describethedatacleaningandpreprocessingstepsyou performed.

-
-

2. DataExplorationandAnalysisinIBMCognos

- Use IBM Cognos tools to perform various data analysis tasks,includingbutnotlimitedto:
  - Descriptivestatistics:Summarystatisticsfornumericvariables.
  - Dataprofiling:Understandingdatadistributionsandpatterns.
  - Dataqualityassessment:Identifyingdataanomaliesandissues.
  - Datasegmentation:Groupingdatabasedonrelevantattributes.

3. DataVisualizationinIBMCognos

- Createvisualizationstorepresent datainsights:
  - Barcharts,linecharts,andpiechartsforcategoricaldata.
  - Histogramsandboxplotsfornumericdata.
  - Scatterplotsforunderstandingrelationships.
  - Geographicmapsiflocationdataisavailable.
- Utilizeappropriatevisualizationtypestoconveyinformationeffectively.

4. AdvancedAnalysisandReporting

- UseIBMCognoscapabilitiesforadvancedanalysis,suchas:
  - Timeseriesanalysisfortrendsandseasonality.
  - Predictivemodelingifapplicable,using IBMCognos'predictiveanalytics features.
  - Creatingdashboardsand reportsforpresentinginsights.

-
-

5. DocumentCreation - Createadetailed

documentthatincludesthefollowingsections:

- Introduction:Explaintheproject'sobjectiveandthedatasetused.
- DataLoadingandPreprocessing:Describehowthedata wasloadedandprepared.
  DataAnalysis:Presentfindingsfromdataexplorationandanalysis.
  DataVisualization:Showcasethevisualizationscreated.
- AdvancedAnalysisandReporting:Discussanyadvancedanalysisandreports.
- Conclusion:SummarizekeyinsightsandthevalueofusingIBM Cognos fordataanalysis.

6. SharingforAssessment

- Sharethedocument and anyrelevantIBMCognosreportsordashboardswith yourintendedaudienceforassessment.

DOCUMENTSTRUCTURE(SAMPLE)

1. Introduction

- Provideanoverviewoftheproject'sobjectivesandthedatasetused.

-
-

2. DataLoadingandPreprocessing

- Explain theprocessofloadingandpreprocessingthedatasetusingIBMCognos.

3. DataAnalysis

- Describe the insights obtained through data exploration and basicanalysis.

4. DataVisualization

- ShowcasethevisualizationscreatedinIBM Cognostoconveydatainsights.

5. AdvancedAnalysisandReporting

- Discuss any advanced analysis performed using IBM Cognos andtheresultingreports ordashboards.

6. Conclusion

- SummarizekeyinsightsandthebenefitsofusingIBMCognosfordataanalysis.

By following this structure, you can create a comprehensive documentthat highlights the steps you've taken and the insights you've gainedfrom your data analysis and visualization using IBM Cognos. Sharingthis document with relevant stakeholders will facilitate assessment andfeedbackforyourDACproject.

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