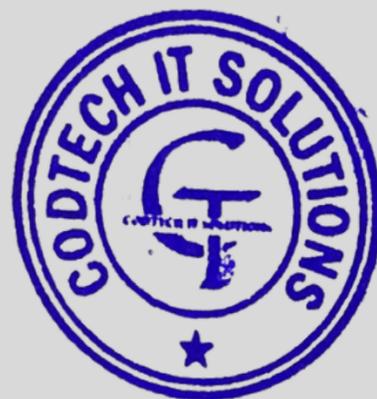




CODTECH IT SOLUTIONS

DATA SCIENTIST INTERNSHIP TASKS

COMPLETE ANY TWO TASKS
OUT OF SIX !



VERIFIED BY
CODTECH IT SOLUTIONS



सत्यमेव जयते

MSME

MICRO, SMALL & MEDIUM ENTERPRISES

CODTECH IT SOLUTIONS

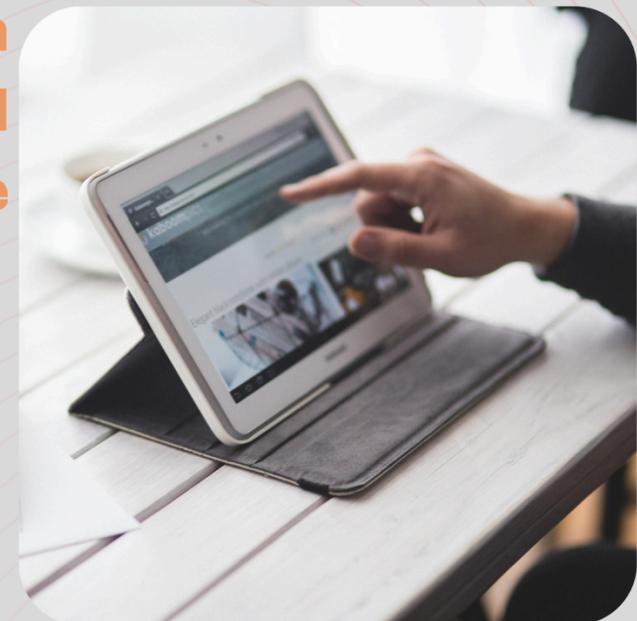


TASK ONE: EXPLORATORY DATA ANALYSIS (EDA)

Start with a dataset of your choice and perform EDA using libraries like pandas, numpy, and matplotlib or seaborn. Explore the data's characteristics, distributions, correlations, and outliers. Visualize your findings with histograms, scatter plots, and heatmaps to gain insights into the data.

TASK TWO: PREDICTIVE MODELING WITH LINEAR REGRESSION

Implement a simple linear regression model using a dataset with continuous target variables. Split the data into training and testing sets, train the model on the training data, evaluate its performance using metrics like mean squared error or R-squared, and make predictions on the test set. Visualize the regression line and actual vs. predicted values to assess the model's accuracy



CODTECH IT SOLUTIONS



TASK THREE: PREDICTIVE MODELING WITH CLASSIFICATION

Utilize a dataset with labeled data and apply classification algorithms to build predictive models. Preprocess the data, split it into training and testing sets, and explore various classification algorithms such as logistic regression, decision trees, random forests, or support vector machines. Evaluate the models using metrics like accuracy, precision, recall, and F1- score, and compare their performances to determine the most suitable algorithm for the given task

TASK FOUR: LIBRARY MANAGEMENT SYSTEM

Explore unsupervised learning techniques by performing clustering analysis on a dataset without predefined labels. Implement clustering algorithms such as K-means, hierarchical clustering, or DBSCAN to identify natural groupings or patterns within the data. Evaluate the clustering results using metrics like silhouette score or Davies–Bouldin index, and visualize



TASK FIVE: NATURAL LANGUAGE PROCESSING (NLP) APPLICATION

Develop an NLP application to analyze and process text data. Utilize techniques such as tokenization, stemming, lemmatization, and stop-word removal for text preprocessing. Implement advanced NLP models like word embeddings (e.g., Word2Vec, GloVe),

TASK SIX: DEEP LEARNING FOR IMAGE RECOGNITION

Build a deep learning model for image recognition tasks using convolutional neural networks (CNNs). Preprocess image data, augment the dataset to increase diversity, and design a CNN architecture with multiple layers (e.g., convolutional, pooling, fully connected). Train the model on a large dataset (e.g., CIFAR-10, ImageNet) and employ techniques like transfer learning with pretrained models (e.g., VGG, ResNet) for improved performance.



CODTECH IT SOLUTIONS



BENEFITS OF THIS INTERNSHIP :

1). PLACEMENT SUPPORT

- INTERVIEW TRAINING**
- RESUME & LINKEDIN BUILDING TOOLS**
- SELF INTRODUCTION EXAMPLES**

2) EXPERIENCE CERTIFICATE

3) MENTOR SUPPORT TO DO GIVEN PROJECTS

4) FLEXIBLE WORKING HOURS {WORK WHEN YOU ARE FREE AND SUBMIT THE TASKS WITHIN THE DEADLINE}

INSTRUCTIONS:



- 1. DEVELOP ADVANCED FEATURES: EXTEND THE FUNCTIONALITY OF YOUR PROJECT BY ADDING ADVANCED FEATURES OR REFINING EXISTING ONES AS PER THE REQUIREMENTS FOR GIVEN TASKS.**
- 2. Update the code in GitHub: Make necessary changes and improvements to your codebase. Commit the updated code to your GitHub repository.**
- 3. Share the GitHub URL at submission link: After updating your code on GitHub, retrieve the updated repository URL and submit it via the provided submission link in whatsapp group .**

NOTE: POST YOUR OFFER LETTER IN LINKEDIN AND TAG US @CODTECHITSOLUTIONS , SHARE THAT POST URL TO WHATSAPP GROUP ADMIN .