Data Science Projects Plan for Beginners

Project Overview

This document outlines the details of three beginner-friendly Data Science projects to be completed within 30 days, with each project taking 10 days to complete. The projects will help participants develop essential data science skills, including data analysis, machine learning, and natural language processing. The start date for these projects is 2nd September 2024. After completing these projects, participants will be ready to move on to more intermediate-level projects. Participants are encouraged to share their completed projects on LinkedIn and add them to their professional portfolios.

Project Timeline: 30 Days

- 2nd September - 11th September: Project 1 - Exploratory Data Analysis (EDA) on Titanic Dataset
- 12th September - 21st September: Project 2 - Predicting House Prices
- 22nd September - 1st October: Project 3 - Sentiment Analysis on Movie Reviews

Project 1: Exploratory Data Analysis (EDA) on Titanic Dataset

Objective:

Analyze the Titanic passenger data to uncover insights such as survival rates and the impact of various factors like gender, age, and class.

Steps

1. Download Dataset:
   - Dataset Link: (https://www.kaggle.com/c/titanic/data)

2. Data Loading:
   - Use Pandas to load the dataset into a DataFrame.

3. Data Cleaning:
   - Handle missing values and correct data types.
   - Identify and address any data inconsistencies.

4. Exploratory Data Analysis:
   - Analyze survival rates based on different features (e.g., gender, age, class).
   - Use Pandas and Seaborn to generate descriptive statistics and visualizations.

5. Visualization:
   - Create bar plots, histograms, and pie charts to visualize findings.
   - Summarize the key insights with appropriate visuals.

6. Documentation:
   - Document the findings in a Jupyter Notebook or report.

Project 2: Predicting House Prices

Objective:

Build a regression model to predict house prices based on various features like area, number of rooms, and location.

Steps:

1. Download Dataset:

- Dataset Link: (https://www.kaggle.com/c/house-prices-advanced-regression-techniques/data)

2. Data Loading:
   - Load the dataset using Pandas.

3. Data Preprocessing:
   - Clean and preprocess the data (handle missing values, normalize/scale features).
   - Perform feature engineering by creating new relevant features.

4. Model Building:
   - Split the data into training and testing sets.
   - Use Scikit-learn to build a linear regression model.
   - Train the model and make predictions on the test set.

5. Model Evaluation:
   - Evaluate the model's performance using metrics like RMSE, MAE, and R^2.
   - Visualize the predicted vs. actual house prices.

6. Documentation:
   - Compile the analysis, model, and results into a Jupyter Notebook or report.

Project 3: Sentiment Analysis on Movie Reviews

Objective:
Perform sentiment analysis on movie reviews to classify them as positive or negative.

Steps:
1. Download Dataset:

- Dataset Link: (https://www.kaggle.com/c/sentiment-analysis-on-movie-reviews/data)

2. Data Loading:
   - Load the dataset using Pandas.

3. Text Preprocessing:
   - Perform tokenization, remove stopwords, and apply stemming/lemmatization using NLTK.
   - Convert text data into numerical form using TF-IDF.

4. Model Building:
   - Split the data into training and testing sets.
   - Build a classification model using Logistic Regression or Naive Bayes.
   - Train the model and predict sentiment on the test set.

5. Model Evaluation:
   - Evaluate the model using metrics like accuracy, precision, recall, and F1-score.
   - Visualize the sentiment distribution.

6. Documentation:
   - Document the model, process, and findings in a Jupyter Notebook or report.

Instructions for Sharing and Adding Projects to Portfolios

Uploading Projects to LinkedIn:
1. Create a LinkedIn Post:
   - Write a brief summary of the project, including objectives, methods used, and key findings.
   - Attach relevant visuals (e.g., charts, model performance metrics).

2. Add to LinkedIn Profile:
   - Go to your LinkedIn profile and scroll to the "Featured" section.

- Click on “+” to add a new feature and select “Media” to upload your project documentation (e.g., a PDF or link to a GitHub repository).
- Add a description and include relevant keywords.

Adding Projects to Portfolios:

1. Create a GitHub Repository:
   - Create a public repository for each project on GitHub.
   - Upload the Jupyter Notebooks, datasets, and any additional files used in the project.
   - Include a detailed README file with instructions and a project summary.

2. Portfolio Website:
   - If you have a personal website, create a section for projects.
   - Include a brief description, key findings, and links to the GitHub repositories.

Conclusion

Completing these projects will enhance your understanding of essential data science techniques. Sharing your work on LinkedIn and adding it to your portfolio will help showcase your skills to potential employers and collaborators. Good luck!