Tasks List

Graduate Rotational Internship Program

The Sparks Foundation

Data Science & Business Analytics Tasks

Prediction using Supervised ML

(Level - Beginner)

- Predict the percentage of an student based on the no. of study hours.
- This is a simple linear regression task as it involves just 2 variables.
- You can use R, Python, SAS Enterprise Miner or any other tool
- Data can be found at http://bit.ly/w-data
- What will be predicted score if a student studies for 9.25 hrs/ day?
- Sample Solution : https://bit.ly/2HxiGGJ
- Task submission:
  1. Host the code on GitHub Repository (public). Record the code and output in a video. Post the video on YouTube
  2. Share links of code (GitHub) and video (YouTube) as a post on **YOUR LinkedIn profile**, not TSF Network.
  3. Submit the LinkedIn link in Task Submission Form when shared.

Prediction using Unsupervised ML

(Level - Beginner)

- From the given 'Iris' dataset, predict the optimum number of clusters and represent it visually.
- Use R or Python or perform this task
- Dataset : https://bit.ly/3kXTdox
- Sample Solution : https://bit.ly/3cGyP8j
- Task submission:
  1. Host the code on GitHub Repository (public). Record the code and output in a video. Post the video on YouTube
  2. Share links of code (GitHub) and video (YouTube) as a post on YOUR LinkedIn profile
  3. Submit the LinkedIn link in Task Submission Form when shared.
  4. Please read FAQs on how to submit the tasks.

Exploratory Data Analysis – Retail
(Level – Beginner)

#3

- Perform ‘Exploratory Data Analysis’ on dataset ‘SampleSuperstore’
- As a business manager, try to find out the weak areas where you can work to make more profit.
- What all business problems you can derive by exploring the data?
- You can choose any of the tool of your choice (Python/R/Tableau/PowerBI/Excel/SAP/SAS)
- **Dataset:** https://bit.ly/3i4rbWl
- **Beginner Level** - Create dashboards. Screen-record along with your audio explaining the charts and interpretations.
- Task submission:
  1. Create the dashboards and/or storyboard and record it
  2. Upload the recording either on YouTube or LinkedIn
  3. Create a LinkedIn post as suggested in FAQs

Exploratory Data Analysis – Terrorism
(Level – Intermediate)

#4

- Perform ‘Exploratory Data Analysis’ on dataset ‘Global Terrorism’
- As a security/defense analyst, try to find out the hot zone of terrorism.
- What all security issues and insights you can derive by EDA?
- You can choose any of the tool of your choice (Python/R/Tableau/PowerBI/Excel/SAP/SAS)
- **Dataset:** https://bit.ly/2TK5Xn5
- **Intermediate Level** - Create storyboards. Screen-record along with your audio explaining the charts and interpretations. Use images.
- Task submission:
  1. Create the dashboards and/or storyboard and record it
  2. Upload the recording on Youtube, share the link on LinkedIn
  3. Submit LinkedIn post link in Task Submission Form when shared
  4. Please read FAQs on how to submit the tasks.

Exploratory Data Analysis – Sports
(Level – Advanced)

#5

- Perform ‘Exploratory Data Analysis’ on dataset ‘Indian Premier League’
- As a sports analysts, find out the most successful teams, players and factors contributing win or loss of a team.
- Suggest teams or players a company should endorse for its products.
- You can choose any of the tool of your choice (Python/R/Tableau/PowerBI/Excel/SAP/SAS)
- **Dataset:** https://bit.ly/34SRn3b
- **Advanced Level** - Create storyboards. Screen-record along with your audio explaining the charts and interpretations. Use annotations, animation and images.
- Task submission:
  1. Create the dashboards and/or storyboard and record it
  2. Upload the recording on Youtube, share the link on LinkedIn
  3. Submit LinkedIn post link in Task Submission Form when shared
  4. Please read FAQs on how to submit the tasks.

Prediction using Decision Tree Algorithm
(Level – Intermediate)

- Create the Decision Tree classifier and visualize it graphically.
- The purpose is if we feed any new data to this classifier, it would be able to predict the right class accordingly.
- Dataset : https://bit.ly/3kXTdox
- Sample Solution : https://bit.ly/2G6sYx9
- Task submission:
  1. Host the code on GitHub Repository (public). Record the code and output in a video. Post the video on YouTube
  2. Share links of code (GitHub) and video (YouTube) as a post on YOUR LinkedIn profile
  3. Submit the LinkedIn link in Task Submission Form when shared.
  4. Please read FAQs on how to submit the tasks.

#7

Stock Market Prediction using Numerical and Textual Analysis

(Level – Advanced)

- **Objective:** Create a hybrid model for stock price/performance prediction using numerical analysis of historical stock prices, and sentimental analysis of news headlines
- Stock to analyze and predict - SENSEX (S&P BSE SENSEX)
- Download historical stock prices from finance.yahoo.com
- Download textual (news) data from https://bit.ly/36fFPI6
- Use either R or Python, or both for separate analysis and then combine the findings to create a hybrid model
- You are free to select a **different** stock to analyze and news dataset as well while not changing the **objective** of the task.

Timeline Analysis : Covid-19
(Level – Advanced)

#8

- Create a storyboard showing spread of Covid-19 cases in your country or any region (Asia, Europe, BRICS etc) using **Tableau, Power BI or SAP**
- Use animation, timeline and annotations to create attractive and interactive dashboards and story
- Identify interesting patterns and possible reasons helping Covid-19 spread with basic as well as advanced charts
- Screen-record the completed storyboard along with your audio explaining the charts and giving recommendations.
- Dataset: Daily updated .csv file on https://bit.ly/30d2gdi
- Task submission:
  1. Create the dashboards and/or storyboard and record it
  2. Upload the recording on Youtube, share the link on LinkedIn
  3. Submit LinkedIn post link in Task Submission Form when shared
  4. Please read FAQs on how to submit the tasks.