DATA SCIENCE INTERNSHIP PROGRAM

level I

SkillNova

www.skillnovatech.in

WELCOME TO OUR INTERNSHIP PROGRAM

- Kickstart Your Data Science Career
- Gain hands-on experience & build real-world projects
- Earn a valuable internship certificate

ABOUT US

SKILLNOVA

is a vibrant and diverse community that brings together individuals with similar objectives and ultimate goals. Our main focus is on creating opportunities that span various areas, including leadership development, learning, student engagement, and fostering shared interests.

We believe in the power of leadership and its ability to drive positive change. That's why we provide platforms and resources for our community members to develop their leadership skills. Through mentorship programs, workshops, and collaborative projects, we empower individuals to take on leadership roles and make a difference in their respective fields.

INSTRUCTIONS

UPDATE YOUR LINKEDIN PROFILES
FOR A WEB DEVELOPMENT INTERNSHIP, YOU WILL NEED TO COMPLETE ANY THREE PROJECTS AT YOUR CONVENIENCE FOR SUCCESSFUL COMPLETION OF THE INTERNSHIP.

MAINTAIN A SEPARATE GITHUB REPOSITORY (NAME SKILLNOVA FOR ALL THE TASKS AND SHARE THE LINK OF THE GITHUB REPO IN THE TASK SUBMISSION FORM . (IT WILL BE GIVEN LATER THROUGH EMAIL).

YOU CAN REFER TO ONLINE RESOURCES SUCH AS GOOGLE SEARCH AND READ TUTORIALS.

SUBMISSION

A TASK SUBMISSION FORM will be shared later through email.
Till then please continue your task.
A video need to be created to showcase your work, demo of your effort.

The video can be hosted on LinkedIn for proof of your work and build credibility among your peers.
You can tag SkillNova in such posts.

Please add #SkillNova in each of your task video postings on LinkedIn, Additionally, you can also add hashtags such as #internship #datascience. for more reach and visibility.

INTERNSHIP STRUCTURE

- Duration: 4 Weeks
- Mode: Online, Self-Paced
- Components: Weekly MCQs + Final Project
- Assessment: Complete all assignments & submit the projects

WEEKLY ASSIGNMENTS SCHEDULE

- Week 1: Python Basics, Data Cleaning, Pandas.
- Week 2: Data Visualization (Matplotlib, Seaborn).
- Week 3: Machine Learning Basics (Regression, Classification).
- Week 4: Advanced ML (Clustering, Model Evaluation).

WEEKLY PROJECTS SCHEDULE

- Week 1: Exploratory Data Analysis (EDA) on a dataset.
- Week 2: Data Visualization Dashboard.
- Week 3: Predictive Model for House Prices.
- Week 4: Customer Segmentation using Clustering.

FINAL PROJECT SUBMISSION

- Choose three project from the given options
- Implement concepts learned during the internship
- Submit the project before the deadline

PROJECT 1 - EDA ON A DATASET

Dataset:

- Titanic Dataset
  - Description: Contains information about passengers on the Titanic, including survival status, age, gender, class, etc.
  - Download Link: Titanic Dataset on Kaggle

Tools:

- Python Libraries: Pandas, NumPy, Matplotlib, Seaborn.
- IDE: Jupyter Notebook, Google Colab.

Materials:

1. Tutorials:
   - Pandas Tutorial for Beginners
   - Matplotlib and Seaborn Tutorial
2. Tasks:
   - Load the dataset using Pandas.
   - Perform data cleaning (handle missing values, remove duplicates).
   - Generate summary statistics and visualizations (e.g., survival rate by gender, age distribution).

PROJECT 2: DATA VISUALIZATION DASHBOARD

Dataset:

- COVID-19 Dataset
  - Description: Contains daily COVID-19 cases, deaths, and recoveries by country.
  - Download Link: COVID-19 Dataset on Kaggle

Tools:

- Python Libraries: Pandas, Matplotlib, Seaborn, Plotly.
- Dashboard Tool: Streamlit.

Materials:

1. Tutorials:
   - Plotly Tutorial
   - Streamlit Tutorial
2. Tasks:
   - Create interactive visualizations (e.g., line charts, bar charts, maps).
   - Build a dashboard using Streamlit to display global COVID-19 trends.

PROJECT 3: PREDICTIVE MODEL FOR HOUSE PRICES

Dataset:

- House Prices Dataset
  - Description: Contains features of houses (e.g., size, location, number of rooms) and their sale prices.
  - Download Link: House Prices Dataset on Kaggle

Tools:

- Python Libraries: Pandas, Scikit-learn, Matplotlib, Seaborn.
- IDE: Jupyter Notebook, Google Colab.

Materials:

1. Tutorials:
   - Scikit-learn Regression Tutorial
   - Feature Engineering for House Prices
2. Tasks:
   - Perform feature engineering (e.g., handle missing values, encode categorical variables).
   - Train a regression model (e.g., Linear Regression, Random Forest).
   - Evaluate the model using metrics like RMSE and R-squared.

PROJECT 4: CUSTOMER SEGMENTATION USING CLUSTERING

Dataset:

- Mall Customer Segmentation Dataset
  - Description: Contains customer information (e.g., age, income, spending score) for segmentation.
  - Download Link: Mall Customer Dataset on Kaggle

Tools:

- Python Libraries: Pandas, Scikit-learn, Matplotlib, Seaborn.
- IDE: Jupyter Notebook, Google Colab.

Materials:

1. Tutorials:
   - K-Means Clustering Tutorial
   - Customer Segmentation Guide
2. Tasks:
   - Perform clustering using K-Means.
   - Visualize clusters (e.g., scatter plot of income vs. spending score).
   - Interpret the results to define customer segments.

CERTIFICATE & COMPLETION

- Complete all assignments & submit your final project.
- Receive your internship certificate.
- Enhance your resume & boost your career.

CONTACT & SUPPORT

- Email: contact.skillnova@gmail.com
- Website: skillnovatech.in
- Follow us on Social Media