CODING SAMURAI

CODING YOUR ASCENT: LEARN, CODE, UPSKILL

ABOUT THE COMPANY

At Coding Samurai, we strongly believe that practical knowledge is really important for doing well in the tech industry. Our main goal is to help students who might be missing some basic skills. We do this by giving them chances to learn in a hands-on way, where they work on real projects and see how things work in the real world. we know how valuable it is for future tech experts to actually try things themselves. That's why our internship program lets students learn by doing real projects. They get to work with experienced mentors who guide them along the way. Our team is made up of experts who have a lot of experience in the industry. They really care about helping interns learn and get better at what they do. We create a friendly and welcoming environment where everyone can learn, grow, and come up with new ideas

INSTRUCTIONS

1-Update your LinkedIn profiles.

2-For the Data Science internship, you will need to complete 1 tasks for successful completion of the internship.

3-Maintain a separate GitHub repository(name as CODING SAMURAI for all the tasks and share the link of the GitHub repo in the task submission form(it will be given later through email).

4-You can refer to online resources such as Google Search and read tutorials. Watch videos(For Help).

SUBMISSIONS

1-A TASK SUBMISSION FORM will be shared later through email . Till then please continue your task.

2-A video need to be created to showcase your work, a demo of your effort.
3-For the Data Science internship, you will need to complete 1 tasks for successful completion of the internship.

4-The video can be hosted on LinkedIn for proof of your work and to build credibility among your peers. You can tag CODING SAMURAI in such posts.

5-Please add #codingsamurai in each of your task video postings on LinkedIn, Additionally, you can also add hashtags such as #internship #webdevelopment. For more visibility.

6-Best Interns will be given **LOR** and a chance to win stipend.

DATA SCIENCE

DATA SCIENCE INTERNSHIP (TASK 1)

Project Title: Exploratory Data Analysis (EDA) on Airbnb Listings

Project Description: In this project, you will perform basic data analysis on a dataset of Airbnb listings. EDA is a fundamental step in data science that involves exploring and understanding the data before diving into more complex analysis or modeling.

Steps you can Implement:

Data Collection: Find a dataset of Airbnb listings in a city of your choice. You can often find such datasets on websites like Kaggle, Inside Airbnb, or by using Airbnb's official API (if available).

Data Exploration: Load and examine the dataset to get a sense of its structure. Look at the first few rows, data types, and basic statistics.

Data Cleaning: Handle missing values and clean the data. This may involve filling missing values, removing duplicates, or correcting data inconsistencies.

Basic Statistics: Calculate and visualize basic statistics like the average price of listings, the distribution of property types (e.g., apartments, houses, etc.), and the distribution of neighborhoods.

Visualization: Create visualizations using libraries like Matplotlib or Seaborn to explore relationships and trends in the data. For example, you can plot price distributions, room type preferences, or the number of listings by neighborhood.

Geospatial Analysis (Optional): If your dataset includes geographic information, you can perform simple geospatial analysis. Create maps to visualize the distribution of Airbnb listings across different neighborhoods or areas.

Time Series Analysis (Optional): If your dataset includes a date field, you can analyze how the number of listings or prices change over time. You can create time series plots and explore seasonality trends.

Correlation Analysis: Investigate correlations between different features. For instance, you can check if the number of bedrooms is correlated with the price or if the reviews affect the booking rates.

Data Insights: Summarize your findings and provide insights from your analysis. What interesting patterns or trends did you discover? What neighborhoods seem to be the most popular, and what factors might influence listing prices?

Documentation: Document your project, including the data source, steps you took, and the insights you gained. This documentation will be useful for presenting your findings and referring back to your work.

LINK TO DATASET- AirbBnb Dataset

DATA SCIENCE INTERNSHIP (TASK 2)

Project Title: Predicting House Prices

Project Description: Build a simple machine learning model to predict house prices based on various features such as square footage, number of bedrooms, neighborhood, and more. This project will introduce you to regression analysis, which is a fundamental concept in data science.

Steps you can Implement:

Data Collection: Obtain a dataset of house prices. You can find real estate datasets on websites like Kaggle, Zillow, or government housing agencies.

Data Exploration: Explore the dataset to understand its structure, check for missing values, and gain insights into the distribution of house prices and features.

Data Preprocessing: Clean the data by handling missing values, outliers, and converting categorical variables into numerical format (e.g., one-hot encoding for neighborhoods).

Feature Selection: Choose relevant features that you believe may have an impact on house prices. You can use correlation matrices or feature importance techniques for guidance.

Model Selection: Select a regression algorithm such as Linear Regression, Decision Tree Regression, or Random Forest Regression for this project.

Data Split: Split your dataset into a training set and a testing set. Typically, 80% of the data is used for training and 20% for testing.

Model Training: Train your chosen regression model on the training data, using the selected features as input and house prices as the target variable.

Model Evaluation: Evaluate the model's performance on the testing data using metrics like Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE).

Prediction: Use your trained model to make house price predictions for new data points or hypothetical scenarios.

Visualization: Create visualizations, such as scatter plots or regression plots, to visualize the relationships between individual features and house prices.

Model Tuning: Experiment with different hyperparameters of the regression model to see if you can improve its predictive accuracy.

Documentation: Document your project, including the dataset source, preprocessing steps, model selection, and evaluation results. This documentation will help you communicate your findings and methods effectively.

LINK TO DATASET-House price prediction Dataset

DATA SCIENCE INTERNSHIP (TASK 3)

Project Title: Spam Email Classifier

Project Description: Create a machine learning model that can classify emails as spam or not spam (ham). This project will introduce you to text classification and binary classification tasks, commonly used in natural language processing (NLP).

Steps you can Implement:

Data Collection: Gather a dataset of emails labeled as spam or ham. You can find publicly available email datasets online, such as the Enron Spam Dataset or datasets on Kaggle.

Data Preprocessing: Clean and preprocess the email data. This may involve removing special characters, lowercasing, tokenization, and removing stop words.

Feature Extraction: Convert the text data into numerical features using techniques like TF-IDF (Term Frequency-Inverse Document Frequency) or Count Vectorization.

Label Encoding: Encode the target variable (spam/ham) as binary labels (1 for spam, 0 for ham).

Model Selection: Choose a machine learning algorithm for binary classification. Logistic Regression, Naive Bayes, and Support Vector Machines (SVM) are good choices for this task.

Data Split: Split your dataset into a training set and a testing set. Typically, use 80% of the data for training and 20% for testing.

Model Training: Train your chosen classification model on the training data, using the extracted features as input and the binary labels as the target variable.

Model Evaluation: Evaluate the model's performance on the testing data using metrics such as accuracy, precision, recall, F1-score, and confusion matrix.

Hyperparameter Tuning: Experiment with different hyperparameters of the classification model to optimize its performance.

Deployment: Create a simple user interface where users can input an email text, and the model predicts whether it's spam or ham. You can build a command-line interface or a basic web application.

Documentation: Document your project, including the data source, preprocessing steps, model selection, and evaluation results. This documentation will help you communicate your findings and methods effectively.

LINK TO DATASET-SpamBase Dataset

Remember tools are just a means to reach the desired results, it's all about the process, so feel free to use any technology or tool you are comfortable with.

It's up to you whether you Brainstorm, use your existing knowledge , Google it or can take Reference from youtube and follow step by step. The main goal of these task is to make you learn and gain hands on experience. Results of individual may vary and its absolutely fine. You can use your creativity.

ASK FOR HELP!

1-THE PURPOSE OF THIS INTERNSHIP IS TO LEARN AND GROW

2-We have no desire to dictate to you. It is entirely up to you whether you seek guidance or not.

3-The given tasks may seem very easy or very difficult. We expect you to approach the tasks with professional diligence and give them the attention they deserve."

CONNECT WITH US

LINKEDIN-Click Here
MAIL- mailto:codingsamuraisensei@gmail.com
TELEGRAM- Click Here