



Ashutosh <ashutoshje223@gmail.com>

Your application for Data Scientist Intern

1 message

Soulpage IT Solutions <hello@soulpageit.com>

Sun, Jun 15, 2025 at 8:04 PM

Reply-To: do-not-reply@soulpageit.com

To: ashutoshje223@gmail.com

Assignment for Data Scientist Intern Position

Hello

Greetings of the day.

Thank you for applying for the Data Scientist Intern position at Soulpage IT Solutions Pvt. Ltd.

As the first step in our interview process, we would like to assess your practical knowledge through an assignment. Please find the details of the assignment below. You need to follow the mentioned conditions, write your code, and revert to us within 3 days from the date you receive this test.

Data Scientist Assignment file: <https://drive.google.com/drive/folders/1ThHnUQjkCNTOKXnvySVfpHZxZbYsFfMQ>

Problem Statement: License Plate Recognition

Detecting and recognizing vehicle license plates represents a widely recognized challenge that has garnered significant attention. This challenge revolves around the utilization of two distinct datasets:

1. A collection of vehicle images (totaling 900 images) sourced from the internet and meticulously annotated. These annotations encompass the precise coordinates of bounding boxes encapsulating the license plates within each image.
2. An assemblage of license plate images (also amounting to 900 images) where the annotations take the form of the alphanumeric text inscribed on each license plate.

The primary aim of this task is twofold: first, to identify and locate the license plates affixed to the vehicles in the images, and second, to perform character recognition on these license plates, deciphering the alphanumeric text they contain.

The dataset is organized into three distinct sets:

Training Set 1: This set comprises 900 images, with each image featuring a single car along with its corresponding license plate. The provided annotations include the coordinates of the bounding box (ymin, xmin, ymax, xmax) that outlines the license plate in each image.

Training Set 2: Consisting of another 900 images, this set focuses solely on license plates. Each image in this set contains a license plate, and the provided annotations contain the characters present on each license plate.

Test Set: This set consists of 201 images and is structured similarly to the first training set. In this set, your task is twofold: you need to detect the license plates within the images and recognize the characters on those plates.

Evaluation Criteria:

1. Pre-processing, data analysis, and understanding
2. Data exploration
3. Model building
4. Accuracy of the character recognition from the license plate

After completing the test, please share a Jupyter Notebook with proper comments about your code hr@soulpageit.com. Failing to send the task within the given timeline may lead to disqualification from the selection process.

In case of any queries, please reach us at hr@soulpageit.com so that we can assist you promptly.

Best of luck!

Team Soulpape IT Solutions

<https://soulpageit.com>

Please do not reply to this system generated message.