**Doktar ML Engineer Case Study**

**Q-1:** The attached Excel file, 'Doktar\_Topology\_Sample\_Data\_20250627.xlsx,' contains two key variables: 'Measured VMC' and 'Normalized\_Values.' 'Measured VMC' represents the soil moisture percentage displayed to the farmer, while 'Normalized\_Values' are the corresponding normalized sensor readings. Please analyze the relationship between these two variables and develop a model to describe it.

1. **Exploratory Analysis & Transformations**
   * 1. **Relationship discovery** –To better understand the connection between 'Measured VMC' and 'Normalized\_Values,' please conduct an exploratory analysis.

1. **Model Design, Validation & Selection**
   * 1. **Candidate models & trade-offs** – Compare at least two viable approaches (e.g., linear/polynomial regression, isotonic regression, monotonic GBMs) and justify your final choice in terms of interpretability, complexity, and expected accuracy.
     2. **Validation strategy & metrics** – Specify your data-splitting or cross-validation scheme and the error metrics you will report, including why those choices suit the problem.

1. **Robustness & Drift Monitoring (Optional)**

*A brief written explanation is sufficient.*

* + 1. **Out-of-range handling** – What should the model do when future Normalized readings fall outside the training span?
    2. **Drift detection** – Outline a lightweight test or control-chart approach to flag shifts in the Measured–Normalized relationship over time.
    3. **Update policy** – Static, scheduled retrain, or continuous learning? State your choice and why.

1. **Regression Diagnostics (Optional)**
   * 1. Execute the residual-based diagnostics or statistical tests (non-linearity, heteroscedasticity, influential points) you’d run after fitting and how the findings might change your model.

**Q-2**: The Q-2 folder holds two kinds of material. First, you will find full frame PestTrap photographs that show a variety of insects. Second, you will see cropped reference images for two species: the insect we normally monitor and the black fly. Your task is concerned only with black flies. Mark every black fly that appears in the full frame photographs and leave all other insects, including the monitored species, unmarked.

Choose any detection workflow you prefer and feel free to use the reference crops if they help. When you submit your solution, include the annotated photographs together with a short, clear description of how you produced them. We will judge the outcome by visual inspection, so numerical accuracy measures are not required.

**Q-3**: Imagine you are tasked with deploying the two models you've developed as a live, production-ready API. Could you outline your proposed system architecture? Please describe the key components, the tools and technologies you would select (e.g., cloud services, frameworks), and how these components would interact to serve predictions.

**Q-4**: For this case study, we place a strong emphasis on code quality, structure, and ease of deployment. Please submit your solution as a well-organized Git repository that includes:

* Production-quality code with clear modular structure and comments.
* A Dockerfile that fully containerizes the application.
* A report (in PDF or Markdown format) containing your answers to Questions 1, 2, and 3. You may include code excerpts or visualizations where appropriate.