Theft Detection Algorithm Assignment

Company: Kudosware

Role: Data Scientist

This assignment is designed to assess your skills in building a theft detection algorithm using video data, specifically focusing on shoplifting scenarios.

Dataset: DCSASS Dataset https://www.kaggle.com/datasets/mateohervas/dcsass-dataset

The DCSASS dataset contains a variety of video clips categorized into different activities, including shoplifting. We will be focusing on videos labeled as "shoplifting" and "normal" to train our theft detection model.

- Just use videos under shoplifting category

Task:

1. **Data Preprocessing:** Download the DCSASS dataset and filter for videos labeled as "shoplifting" and "normal." Preprocess the videos by converting them into a suitable format for your chosen machine learning model.
2. **Feature Engineering:** Extract relevant features from the preprocessed videos that can be used to identify shoplifting activities.
3. **Model Building:** Develop a machine learning model to classify video frames/clips as containing shoplifting or not containing shoplifting.
4. **Evaluation:** Evaluate the performance of your model using relevant metrics like:
   - **Accuracy:** Overall percentage of correct classifications.
   - **Precision:** Ratio of true positives (correctly identified shoplifting) to total positive predictions.
   - **Recall:** Ratio of true positives to actual shoplifting instances in the dataset.
   - **F1 Score:** Harmonic mean of precision and recall, balancing both metrics.
5. **Report:** Prepare a report summarizing your approach, including:
   - Data preprocessing steps and rationale for chosen techniques.
   - Feature engineering methods and their effectiveness in identifying shoplifting.
   - Model selection with justification and hyperparameter tuning (if applicable).
   - Evaluation results, including confusion matrix and chosen metrics.
   - Discussion of strengths, weaknesses, and potential improvements to your approach.

Deliverables:

- Jupyter Notebook or Python script containing your code
- Report summarizing your methodology, results, and discussion

Evaluation Criteria:

- Clarity and efficiency of code
- Appropriateness of feature engineering techniques
- Model selection and justification
- Performance of the model on the shoplifting detection task
- Communication and insights presented in the report

Time Limit:

Submit it by 26 April. Upload the required documents to google drive and make it **public** and send it over on internshala.

Additional Notes:

- Feel free to explore libraries like OpenCV or TensorFlow for video processing and model building.
- You can make any reasonable assumptions about shoplifting scenarios based on the video descriptions or additional research.
- Be creative and showcase your problem-solving skills in building an effective shoplifting detection algorithm using the DCSASS dataset.