Deepfake ML Model Hackathon
Synergy'25

Problem Statement: Predicting the Distribution
Overview

Participants in this challenge are tasked with developing a solution to predict deepfake detection outputs for a new set of test images. The provided dataset consists of:

1. **Training Images**: A labeled collection of images used to train your predictive model.

2. **JSON File**: A structured file containing deepfake detection outputs generated by a proprietary model for the corresponding training images.

The goal is to analyze the relationships between the features in the images and the outputs in the training JSON file to create a robust predictive model. Your model will then be applied to the test images to produce outputs similar to those generated by the proprietary model.

Objective

The primary objective of the challenge is to:

Develop a model that accurately predicts deepfake detection outputs for the test images based on the provided training data and the JSON file. The model outputs will be compared to ones generated by the proprietary model and be graded accordingly

Participants must demonstrate their ability to:

1. Extract meaningful insights from the training images and the JSON data.

2. Build a predictive pipeline that generalizes effectively to unseen data.

3. Deliver outputs in a format that matches the structure of the test JSON file.

Evaluation Criteria

Submissions will be evaluated based on the following criteria:

1. **Accuracy**:

   • The predicted outputs in the test JSON file will be compared against the ground truth JSON file provided during the evaluation phase.

   • Scoring will be based on how closely the predicted outputs match the actual outputs.

2. **Robustness**:

   • The ability of the solution to generalize across diverse and unseen test data will be assessed.

•Models that perform well across variations in the test dataset will receive higher scores.

3.**Scalability**:

•Submissions that are computationally efficient and scalable will be given preference.

Dataset Description

The dataset provided for this challenge includes:

1.**Training Images**:

•A diverse set of images, labeled appropriately.

•These images will serve as the primary input data for training your predictive model.

2.**Training JSON File**:

•A structured JSON file containing deepfake detection outputs for the training images.

•This file includes results produced by the proprietary deepfake detection model.

The dataset can be found at Deepfake Detection Hackathon.

Expected Output

Participants are required to submit the following:

1.**Predicted JSON Outputs**:

•A JSON file containing the predicted deepfake detection outputs for the test images. The file name should be "teamname_prediction.json"

•The format of this JSON file must match that of the training JSON file for consistency and accurate evaluation.

2.**Documentation**:

•A PowerPoint presentation explaining the methodology used in solving the problem. The name of the ppt should be "teamname_presentation"

•The report should include:

•Steps for preprocessing the data.

•Techniques for feature extraction and representation.

•Details about the model(s) used, including architecture, hyperparameters, and training strategy.

•Reasoning behind the chosen approach.

•Challenges faced during the process and how they were overcome.

Key Points to Consider

•**Feature Extraction**:

•Participants are encouraged to extract and analyze features from the training images and JSON file to gain a better understanding of the patterns in the data.

•**Model Design**:

•The model must effectively map the relationships between image features and the outputs in the training JSON file.

•Advanced techniques such as deep learning, transfer learning, or multimodal learning may be utilized.

•**Evaluation Approach**:

•Predictions should aim for high accuracy while maintaining robustness across variations in test data.

•Consider the scalability of the solution for larger datasets.

Submission Requirements

Submissions must include:
Upload everything in a google drive and upload the link on Unstop, once the submission round begins.

1.**Predicted Test JSON File**:

•A JSON file containing the predicted outputs for the test images.

2.**Documentation**:

•A comprehensive report outlining the methodology, model design, and justification for the approach.

3.**Code Repository** (optional but encouraged):

•A GitHub or similar repository containing the code used for the project.

•Ensure that the repository is well-structured and includes a README file with instructions for running the code.

Challenge Guidelines

•Participants must ensure that their solutions are original and do not violate any copyright or intellectual property laws.

•Any use of pre-trained models or external data must be disclosed in the documentation.

This challenge offers an opportunity to demonstrate expertise in predictive modeling, computer vision, and data-driven decision-making. Best of luck!