

The PAN-BS AI-ML Challenge

Overview (Event Report)



Presented by the LogicLooM Organizers

BS Degree Programme

IIT Madras

August 11, 2024



<https://www.youtube.com/watch?v=MCmgfA7kQkg>



<https://mlig-iitmbs.github.io/ml-challenge>



Introduction

Track 1 : Fake News Detection in Dravidian Languages

- ▶ In an age of information overload, accurately categorizing fake news is crucial for fostering reliable communication. The task underscores the need to explore the effectiveness of NLP in understanding Dravidian languages, which are less widely spoken.

Track 2 : AI-Generated Text Detection in Articles

- ▶ With the rapid advancement of AI, distinguishing between human-written and AI-generated content is increasingly challenging. This challenge aims to explore capabilities of ML models to accurately identify the origin of textual content, contributing to the development of robust techniques for detecting AI-generated text.

- ▶ 183 teams..., 127 in Track-2 and 56 in Track-1!
- ▶ A total of 342 participants!
- ▶ 5 mentor sessions (where we started from the scratch - classification basics) hosted through out the challenge.
- ▶ Foundational students were encouraged to participate! Our goal wasn't to make them master the intricate aspects right away, but rather to help them gain a solid understanding of the basics and how ML works.
- ▶ Majority of the Foundational students who were new to ML were able to build models and submit their solutions successfully, two of them have even made it to this!
- ▶ 12 teams from Track-2 & 6 from Track-1 have qualified for this presentation round, based on the run submission mac-F1 & Acc. metric scores.

Teams: Before & After the Mentor Sessions :D



Aug 04: Release of Problem statement, Labelled Train data & milestones

- ▶ Aug 07: Release of Unlabelled test set The test set (without ground truth labels were provided to the teams for their developed models to generate predictions.
- ▶ **Aug 08-09: Submission of runs and code** Each team was allowed to submit a max. of 3 runs on our event portal. To help them optimise their models, the portal showed the scores for the 1st 2 runs. The best performing model score was considered.



Meet the Top Achievers!!




SL No.	Participants	Score
1	Kartik Agrawal	0.765,0.419
2	Keshari Nath Chaudhary, Pratiksha Naik	0.841,0.4
3	Shaikh Gufran Jabbar, Sukanya S	0.871,0.405
4	Nitish Rishi	0.803,0.364
5	Aniket Dash, Deva Vasista	0.886,0.336
6	Sanyam Miital, Nithish Kumar	0.879,0.318

Note



In descending order of the best macavg-F1 & Accuracy scores respectively


SL No.	Participants	Score
1	Sai Ruthvik, Shankha Subhra Saha	0.986,0.963
2	Parashmani Datta, Athish Sivakumaran	0.983,0.956
3	Darshan kumar	0.983,0.951
4	Saminathan C, Vanchit Visanth M S, Sahishnuram S	0.969,0.922
5	Siddharth Roy, Sarfraz Ahmed	0.954,0.892
6	Gaurav Singh, Sakshi	0.95,0.863
7	Ayaan Qureshi, Nikhil Maurya	0.938,0.861
8	Shiva Kumar	0.917,0.833
9	Rohit Satheesh, Mohit Kumar	0.888,0.79
10	Saravanan K, Stuti Bahuguna	0.886,0.783
11	Arka Dash, Nimish Shinde	0.82,0.713
12	Manaswita Mandal, Debapriyo Saha	0.805,0.697


Note: In descending order of the best macavg-F1 & Accuracy scores respectively


**IITM-PAN BS AI-ML
CHALLENGE**


AI-ML Challenge





 Dashboard


 Development Phase LIVE


 Submission Phase Inactive


 My Performance


 Discuss

 Logout

 Your team has registered for track Fake News Detection in Dravidian Languages.

 Development & Training Phase - Problem statement, milestones, train data release on 4th

 Testing Phase - Test dataset release on 7th

 Final Submission Phase - 8th and 9th

Useful Resources

Understanding basics of Classification

Understanding Binary classification (DeepLearningAI)

Simple ML model to predict new category for multi-class classification

Top ML algorithms for classification

Building first ML model in Python

Best Binary classification models in ML

Tutorial on Logistic regression (binary classification)

Python Pre-requisites

Getting Started with Python (English)




Getting Started with Python (Hindi)

G-Colab Tutorial

Pandas & NumPy libraries



SkLearn Basic Documentation




Data Analysis & Handling in Python




Our Dashboard


Somsubhra





**IITM-PAN BS AI-ML
CHALLENGE**




AI-ML Challenge


 **Submission Successful!**
-- / 132 test cases passed.

 Dashboard

 Development Phase LIVE


 Discuss



 Due at 2024-08-09 23:59:59
You can submit a **max of 3 runs**, the best submission will be considered for grading.

Instructions

- Please ensure that the run submission is in the format of **.csv** (comma-delimited) and you are uploading the **correct file only**.
- The file must contain 2 columns in this order: 1st and 2nd columns must be named **`ID`** and **`Predicted Type`** respectively.
- Please **don't modify the `ID`** column. The **`Predicted Type`** column should contain the **labels** predicted by your model.
- Cross-check that your predicted labels align to the submission template/format (i.e. **numeric labels** 0/1/2/3/4 for Track-1 and 0/1 for Track-2).

 **Test Run Summary**
Scores were compiled based on dummytest.csv (text/csv, 1495B)

Model Evaluation Summary

🔗 Check the My Performance tab for the scores.

Choose File

No file chosen

Testing & Submission Phase

“Without data, you’re just another person without an opinion”

W. Edwards Deming

Thank You!