

TECHFEST 2020-21

AI Classifier

Mr. Andy Cooper, COO of a multinational company, was back attending an AI focused seminar and evaluating opportunities in his own business. When he looked at global operations, he realized that every business unit and geography has its own process and data variation. It often leads to issues in gathering data uniformly and gaining economies of scale. Streamlining all the processes and building a uniform solution is a mammoth task.

He then zeroed down his focus on emails – which accounts for 70% of overall customer interactions. Emails are already in digital format and follow a set structure globally. He strongly believes that a retainable packaged ML/NLP solution can be designed - to be used globally.

TASK

Participants are supposed to design an AI model to classify email complaints of customers into pre-decided categories so that the company could work more efficiently.

PROBLEM STATEMENT

Participants are supposed to model an AI classifier that could take an input as bulk emails as training dataset and also it should be able to take bulk emails as input as testing dataset.

1. **Training data creation module** – It should be easy for business to provide data to train the model. Business only needs to keep emails into tagged folders and training data gets generated automatically.
2. **Classifier Module** – re-trainable ML module to categorize emails into business specific categories
3. **Run-time component** – Solution should be a plug and play component for any RPA/ Automation platform to use as a library or an API.

Participants will be provided with a training set of 1000 emails, on which they have to train the model. The model should be re-trainable and would be tested with a different training and testing dataset on the day of competition.

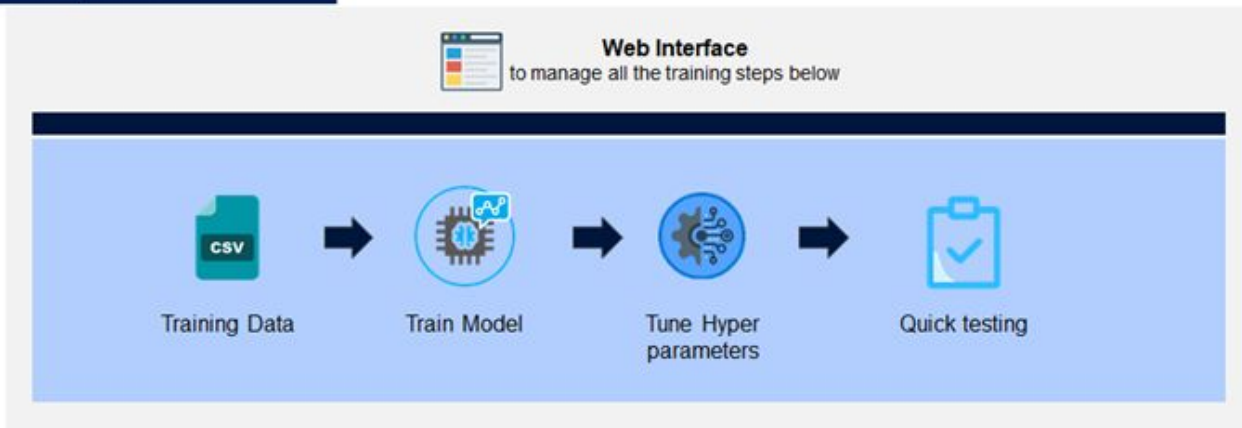
Participants are also expected to create a minimalist(design-wise) website which could take the input for training and testing data sets from local storage or online, and give the output in form of categorized emails.

Here is a flow diagram of what participants are supposed to do.

Training Data Creation



Training Module



Run Time Module



STRUCTURE

Stage 1:

Abstract Submission

Participants are supposed write the following details in a pdf format

1. Which classifier algorithm did they shortlist and why?
2. Front-end mockups - Web interface to load training data, what are the hyperparameters that participants would like to tune for the training set, and quick testing results.
3. Solution flow - How will your solution work from the user's perspective, from backend's perspective, and how will the algorithm work exactly.

Participants are supposed to submit abstracts on aiclassifier@techfest.org

Stage 2:

Model making

Selected participants will be provided with a training set of 1000 categorised emails. Participants are then supposed to design a re-trainable model on the basis of those mails.

EVALUATION

1. Selection of ML Algorithm – **20%**
2. Accuracy of classifier to classify 50 new email samples (Min expected accuracy – 90%) – **40%**
3. Re-train the model using 500 new mails (will be provided at the time of Demo) – **20%**
4. Accuracy of newly trained model across 20 new emails (Min expected accuracy – 75%)– **20%**

ELIGIBILITY

- Individuals or teams from the following categories are allowed:
 1. Students/research scholars of authorized institutions (students have to show their Valid College/School ID)
 2. Upto 3 years old college pass-outs.
- A team is allowed to have a maximum of 4 members..

REGISTRATION AND SUBMISSION

The Participants have to register on the official Techfest Website and fill all the necessary details. www.techfest.org ->(Hover on) Competitions-> Competitions -> AI Classifier -> Explore More -> Register -> Fill all your details - > Now you must create/Join a team

Participants are supposed to submit an abstract in with team_ID as the name of pdf format to aiclassifier@techfest.org.

The final submissions have to be in the form of website link, and output in the form of categorized emails to aiclassifier@techfest.org.

TIMELINE

Last date of Registration	10 November
Abstract Submission	10 November
Result Announcement	15 November
Final Rounds	18/19 December

GENERAL RULES

1. Every team has to register online on our website for the competition. A Team ID will be allocated to the team on registration which shall be used for future references.
2. A team can register at any point of time before 10th November 2020 and submit the final abstract and video (as mentioned in the structure).
3. The decision of the organizers or judges shall be treated as final and binding on all. Techfest has all the rights to verify the identity and accuracy of the details provided by the participants.
4. No responsibility will be held by Techfest, IIT Bombay for any late, lost or misdirected entries.
5. The idea presented by the teams should be original (not protected by means of patent/copyright/technical publication by anyone else).
6. Note that at any point of time the latest information will be that which is on the website. However, registered participants will be informed through mail about any changes on the website.
7. All modes of official communication will be through the Techfest e-mail.

Glossary of terms used:

1. Hyper-parameter:

A value that is set before the ML Training process begins. These parameters are tunable and can directly affect how well a model trains. Some examples of hyperparameters in machine learning:

- Learning Rate
- Number of Epoch
- Number of layers in a neural network
- Regularization constant
- Number of branches in a decision tree.

2. Training Data

The data set using which algorithm is trained. In supervised learning problems, each training data point consists of an output variable (Category) and one or more input variables.

3. Deep Learning:

Deep learning is part of a broader family of machine learning methods based on artificial neural networks with representation learning.

Ref: <https://machinelearningmastery.com/what-is-deep-learning/>

FAQs:

1. What is the role of web interface in the solution

Web interface is required to simplify the re-training and demonstrating the solution. It should allow user to browse files while loading training data or input emails. It should have functionality to call the classifier module to generate and display the results.

2. What is the Hyper parameter tuning

Hyper-parameter tuning is required to fine-tune the training process to archive optimized results. Web portals should have a simple table displaying hyperparameters used and also allow users to change it for re-training.

Hyper-parameters is the only thing you can change while retraining with the new set of emails.

3. What is the a Quick Test functionality

While training a model, you should keep some data aside for testing (Test data set). This functionality will allow you to test one or set of emails quickly while building or demonstrating the solution

Tip: Quick test and runtime component can be technically same

4. Front end Mock-ups

We are expecting a clear outline of the webpage which will allow users to easily train, test and demo the complete solution. Functionality is an important evaluation criteria than the look & feel.

5. **IMPORTANT: Re-trainability**

It is extremely critical for a solution to get retrained and perform with a completely new set of email. Solution should demonstrate smooth functionality of loading new emails, training and tuning the model and classify the new test data.

6. Will emails have attachments

Email set provided will be a mixed bag of emails. Some of the emails may have attachments and type of attachment may vary (PDF, excel, word, images etc.). Use of attachment content for training classifier models is optional.

Subject line, Content and Attachment Names will be free text with no predefined structure.

CERTIFICATE POLICY

Only those teams that are shortlisted for the finals and also give a final presentation about their work during Techfest 2020-21 would be awarded an e-Certificate of Participation. The top 5 entries from this event would be provided with a Certificate of Excellence.

PRIZES

The prize money will be awarded to top x winners via NEFT and will be processed within 30 working days after receiving the prize money from sponsors. Winners have to mail the following information (immediately after the announcement of results) to shubhamgautam@techfest.org

FORMAT OF MAIL :

Subject: AI Classifier, Team_ID - Position - (example- AI Classifier, AC191003 - 3 rd Position)

Body of mail:

1. Account Holder's Name
2. Account Number
3. Bank name and Branch name.
4. IFSC Code
5. Photograph of Bank Passbook as a proof