Techfest 2024-25
ClassiFi

Objective:
ClassiFi aims to develop a self-bootstrapping document classification method using AI to categorise documents into more specific groups progressively without requiring extensive time for initial training data creation. By iteratively improving the model with each cycle, the system achieves increasing levels of classification detail over time.

To categorize documents into more specific groups without dedicating considerable time to training data development, implement a self-bootstrapping document classification approach. This method utilizes AI techniques to iteratively enhance the classification model over time. Below is a high-level overview of the processes needed to implement this solution.

Tasks:

1. Start with a Small Set of Labeled Documents
- Begin by manually labelling a small set of documents into broad categories. These high-level categories will act as the foundation for the classification process.

2. Train an Initial Classification Model
- Use the labelled document set to train an initial AI-based classification model. This baseline model will provide the foundation for subsequent classification tasks.

3. Classify Unlabeled Documents
- Apply the initial model to a larger set of unlabeled documents. The AI model will automatically categorize these documents into broad categories.

4. Review and Label Classified Documents
- Review the automatically classified documents. Manually label them into more specific subcategories within the broad categories to increase the model's granularity.

5. Incorporate Labeled Data into Training Set
- Add the newly labelled documents to the initial training set to create an expanded set of labelled data.

6. Retrain the Classification Model
- Use the expanded training set to retrain the classification model, improving its ability to classify documents into specific subcategories.

7. Repeat Iteratively
- Repeat steps 3 to 6, gradually increasing the level of classification detail in each iteration. As the process continues, the model's classification capabilities will improve.

Solution Components:

1. **Document Repository**: A storage system for documents requiring classification.
2. **Initial Training Set**: A small set of documents manually categorized into broad groups.
3. **Classification Model**: An AI-based document classification model that will progressively improve.
4. **Unlabeled Document Set**: A larger set of documents requiring classification.
5. **Review and Labeling Interface**: A user-friendly interface allowing human reviewers to label documents into more detailed subcategories.
6. **Expanded Training Set**: A dataset that combines the initially labelled documents with those manually labelled during the review process.
7. **Retraining Module**: A system to retrain the model using the expanded training set.
8. **Iteration Control**: A mechanism to manage the number of classification iterations and subcategory refinement.

Implementation Considerations:

1. **User-Friendly Interface**: Prioritize simplicity and usability in the design.
2. **Integration with Existing Systems**: Seamlessly integrate with internal knowledge repositories like GitHub, internal wikis, etc.
3. **Data Privacy and Security**: Implement strong authentication and encryption measures to protect sensitive organisational information.
4. **Third-Party Software and Data**: Participants are not allowed to use third-party software or data applications. All solutions must be built using the provided resources or tools developed in-house. Violation of this rule could lead to disqualification.
5. **Custom Sample Data**: Participants are free to use their own sample data for AI training, but it must remain relevant and aligned with the scope of the problem statement. Submissions with data significantly deviating from the problem scope may be rejected.

STRUCTURE:

There will be three stages to the competition -

Round 1: Abstract review meet

For abstract submission, registered teams must create a presentation with no more than 15 slides. The presentation should include a detailed plan of action, the underlying logic, implementation details, and the software and online platforms that will be utilized in the project. Shortlisted teams will showcase their work and improvements will be suggested by the experts

Round 2: Final review meet

The updated report will be showcased based on which **5 teams** will be shortlisted for the final presentation.

Round 3: Presentation and Working Demonstration
Working demonstration and presentation of the solution
Top 5 teams shortlisted from Stage 2 will present their solution at Techfest 2024-25

ABSTRACT SUBMISSION:

- Prepare a report and send it to classifi@techfest.org
- Subject of the mail: ClassiFi_TeamID_TeamLeaderName
- Mention all your team details clearly in the mail.

TEAM SPECIFICATIONS AND ELIGIBILITY

- All students with a valid identity card from their respective educational institutions are eligible to participate in the competition
- A team can have a maximum of 4 members.
- Students from different institutes can be a part of the same team.

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 19th October 2024 and submit the final abstract.
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 using patent /copyright /technical publication by anyone else).
6. Note that at any point in 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.

CERTIFICATE POLICY

1. The top three teams in the grand finale will be awarded a Certificate of Excellence.
2. E-Certificate of participation will be given to the teams scoring more than the critical marks which will be decided later.

Timeline

Last date of Registration	20th October	Participants need to register before this date
Last date for abstract submissions	20th October	Submission of First draft report
Round 1	27th October	Experts will review your abstract and guide you for improvements
Round 2	2nd December	Participants will showcase updated reports and 5 teams will be shortlisted
Round 3	17th-19th December	Final Live presentation in front of Judges

PRIZE MONEY

The Prize money will be awarded to the **top 3** Winners via NEFT and will be processed within **30** working days after receiving the Prize Money from Sponsors.

The Winners have to mail the following information (immediately after the announcement of the results) to tarun@techfest.org .

Subject: Competition name, team ID - your position (example- ClassiFi, ClassiFi-2309234 – 1st position)

Body of mail-

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