



ASSESSMENT FOR DATA SCIENCE – TRAINEE
DURATION: 6 MONTHS

Step 1

Scrap YouTube (or similar websites like Dailymotion) to collect data for classifying videos (based on description) in following 6 categories –

- Travel Blogs
- Science and Technology
- Food
- Manufacturing
- History
- Art and Music

For each of the aforementioned categories, scrape the video descriptions and prepare csv file which looks like the following:

Video id	Title	Description	Category
RcmrbNRK-jY	200 Days - A Trip Around the World Travel Film	My wife and I traveled to 17 countries in 200 days. This film is the story of our incredible trip! Enjoy! We used a GoPro and a Nikon D7000 for all of the filming. For Business Inquiries please e.mail me at 40northdesigns@gmail.com .	travel

Video id: It is the video URL which is unique for any video on YouTube. For example, in <https://www.youtube.com/watch?v=RcmrbNRK-jY>

RcmrbNRK-jY is the video ID.

Category: In general, you should first search for videos by category and then scrape them as it will save you a lot of time.

Please note that collecting at-least 1700 samples per category is mandatory.

Step 2

Choose and use one model from each of the following categories to classify your dataset.

#Category	Model types
1	Linear classifiers, Naive-Bayes classifiers or SVMs
2	Bagging models, boosting models or shallow NNs
3	CNN, LSTM, GRU, Bidirectional RNNs, or RCNNs

Step 3

Submit a report of the above 2 steps. In the report you must include -

1. The Data you scrapped in an excel sheet format.
2. A Brief summary of how you scraped and pre-processed data from YouTube. Please note that candidates with original ideas and code written completely from scratch will be given more points.
3. The report MUST include a reason why you chose that particular model from each category for the classification.
4. For every technique you used, you MUST report precision, recall, and F1 Score. This should be in a graphical representation.
5. Finally, you will have to explain why you got the results and what has made that difference.

Final submission to be done at ashish@precily.com only.

NOTE:

1. The completed task must be submitted within 2 days (48 hours) of receiving it.
2. Every step in the task is self-explanatory to the best of our knowledge. If any part is unclear, use your best judgment and mention in your report.
3. Your project will not be used for the benefit of the company in any manner. The intention of this task is ONLY to evaluate your skills.
4. We wish you all the best!