SAP Machine Learning Challenge

Please read the following instructions carefully

Which Novel Do I Belong To?

In this challenge, you are tasked with training a machine learning model that classifies a given line of text as belonging to one of the following 12 novels:

0. alice_in_wonderland
1. dracula
2. dubliners
3. great_expectations
4. hard_times
5. huckleberry_finn
6. les_miserable
7. moby_dick
8. oliver_twist
9. peter_pan
10. tale_of_two_cities
11. tom_sawyer

You are provided with a zip file (offline_challenge.zip) containing three text files:

- xtrain.txt
- ytrain.txt
- xtest.txt.

As you can see in the train files, we have applied an encoding to the text, but it is done such that each character has a deterministic mapping. Each line in xtrain.txt corresponds to a label in ytrain.txt.

Example:

```
line:
satwamuluhqgulamlrmvezuhqvkrpmletwulcitwskuhlemvtwamuluhiwiwenuhlrvimvqvkruh
ulenamuluhqgqvtwvimviwuhtwamuluhulqvkrenamcitwuhvipmpmqvuhskiwkrpmdfuhlrvimv
skvikrpmqvuhskmvgzenleuhqvmvamuluhulenamuluhqvletwtwvipmpmgzleenamuhtwamuluh
twletwdfuhiwkrxeleentwxeuhpmqvuhtwiwmvamdfuhpkeztwamuluhvimvuhqvtwmkpmpmlelr
uhgztwtwskuhtwlrkrpmlruhpmuluhqvenuhtwyplepmxeuhenuhamypkrqvuhamulmvdfuhqvsk
entwamletwlrlrpmiwuhtwamul
label: 7
```

Your Task

You are tasked with developing a machine learning model (deep learning preferred) that predicts the novel id of a given line of text.

Submission

As part of your submission, please include:

- Your model's predictions on xtest.txt (in the same format as ytrain.txt).
    - This file must be named as **ytest.txt**
- Source code **as a .zip file** (we prefer Jupyter notebooks, size limit is 10 MB)

Evaluation

Your submission will be evaluated based on the following criteria:

- Test set accuracy (80%)
- Explanation/documentation (10%)
- Implementation (10%)

Contents of Source Code

In your source code, please include the following:

- Implementation of the model
- Clear documentation of relevant parts of the code
- Training & validation accuracies
- **Explanation of strategy, methodology, and algorithms employed**

The last point is especially important, as we want to assess your reasoning and approach to this problem.

Good luck!