ELECTRONICS ENGINEERING SOCIETY

UDYAM'25

MOSAIC

PS-1

Introduction

The Tale of the Clumsy Scholar in the Library of Mosaic '25

It was a peaceful evening in the Library of Mosaic '25. You were feeling all scholarly, flipping through books, when—whoops! You knocked over a whole tower of ancient texts with a single, careless move. The books crashed to the ground, pages flapping open, and to your horror, the words on the pages were now magically smudged! The <MASKED> tokens appeared all over the place where the words used to be.

Now, you're in a panic. You're the one who caused this disaster! If the great wizard finds out what you've done, you're doomed! The wizard is coming soon to inspect the library, and he'll surely notice the missing words. He's going to be so mad. But there's a way out! You can fix this, but it's going to take your best skills at figuring out what the missing words should be. The trick is—you can't just guess. You need to understand the sentence and the context to figure out the word that should go in each spot.

ELECTRONICS ENGINEERING SOCIETY

UDYAM'25

Your Mission:

Your job is to fix the mess you made! You need to fill in the missing words in the sentences and save yourself from the wizard's wrath. Each sentence has a <MASKED>, and you need to figure out which word fits best in that blank. It's up to you to make things right!

Examples of Your Blunders:

Sentence: "The cat sat on the <MASKED>."
Your Prediction: "mat"

Sentence: "He is a <MASKED> engineer."
Your Prediction: "software"

What's at Stake:

You did cause this chaos, and now it's up to you to fix it. If you make the wrong guesses, the wizard will definitely know something's off. But if you nail it, you might just get away with your clumsiness and become a hero of the library! So, don't blow it—your fate is in your hands.

Good luck, and may your quick thinking save you from the wizard's wrath!

ELECTRONICS ENGINEERING SOCIETY

UDYAM'25

File Format

Fortunately, There are so many different sources, from online WIZARD communities to public datasets to hidden archives—could one dataset be more trustworthy than another? Should you trust what the masses have contributed, or should you be more selective?

But one thing is for sure: The data is there. It's vast. It's valuable. It could help you, but only if you can use it wisely.

train_set.csv -

- IDS - Unique identifiers for English sentences.
- SENTENCES - Full English sentences.

test_set.csv -
a collection of smudged, masked sentences that you need to correct.

- IDS - Unique identifiers for English sentences.
- MASKED SENTENCES - English sentences with a word masked.

sample_submission.csv -

- IDS - Unique identifiers for English sentences.
- PREDICTED WORDS - The predicted masked word.

Dataset link：

1. Train
2. Test

ELECTRONICS ENGINEERING SOCIETY

UDYAM'25

Team Composition & Submission rules

- Team Size: ≤ 3 members
- Submission Deadline: 16th March EOD
- Deliverables:
  - Notebook (.ipynb) : The notebook must clearly document each step, including data preprocessing, model training, evaluation, and predictions.
  - Evaluation Metrics : Provide a clear definition and explanation of the evaluation metrics used to assess the model's performance, along with the rationale for selecting these metrics.
  - Report (pdf or doc file): Provide a detailed explanation of the approach, model design, training process, and hyperparameters used.
  - Submission File (submission.csv) : Ensure the CSV file follows the required format as specified in the file format description.

- Use of external dataset is not allowed.
- Direct use of pretrained model is strictly prohibited.
- Don't cheat, and have fun!

Mitul Agarwal
9368195131

Vidit Khandelwal
9811831312

Aditya Raj
8319678938