

Impact Analytics - Assignment (NLP)

Build an LSTM model for named entity recognition(NER). [Dataset](#) has been attached with the email or you can use any other publicly available datasets for NER. You can find more information about dataset in the link provided. You can choose any framework based on your comfort.

Instructions:

Please create a repository in Github and send us the link to it as soon as you start. We should be able to see your check-ins/commits from the beginning and *not* just like a single commit at the end. Include instruction on how to run the project and set up the requirements.

Note: The assignment is to check your approach and implementation skills and not to get a trained model.

Explain the following things in README of Github

1. Explain in detail the process of data preparation and feature extraction?
2. Describe the hyperparameter choices?
3. How does the batch size affect your model?
4. Report Recall, precision, and F1 measure
5. Apart from LSTM if given a choice what model do you think works better? and why?
6. Describe the problems you faced during the implementation and how you tackled them.

Bonus(not mandatory):

- Report entity wise Recall, precision, and F1 measure like PER, LOC
- Report the effect of the imbalanced dataset if any

Disclaimer: We expect your code/approach to be free from plagiarism.

References: <https://arxiv.org/pdf/cs/0306050.pdf>