**CS6005-DEEP LEARNING**

**MINI PROJECT ON NLP**

**BY**

**S.KARTHICKRAJA**

**2018103547**

**NAMED ENTITY RECOGNITION**

**PROBLEM STATEMENT:**

[Named Entity Recognition and Classification](https://en.wikipedia.org/wiki/Named-entity_recognition) (NERC) is a process of recognizing information units like names, including person, organization and location names, and numeric expressions including time, date, money and percent expressions from unstructured text. The goal is to develop practical and domain-independent techniques in order to detect named entities with high accuracy automatically.

**DATASET DESCRIPTION:**

**Annotated Corpus for Named Entity Recognition** using GMB(Groningen Meaning Bank) corpus for entity classification with enhanced and popular features by Natural Language Processing applied to the data set.

*Essential info about entities*:

* geo = Geographical Entity
* org = Organization
* per = Person
* gpe = Geopolitical Entity
* tim = Time indicator
* art = Artifact
* eve = Event
* nat = Natural Phenomenon

Total Words Count = 1354149  
Target Data Column: "tag"

**MODULES:**

## **Loading the Data for Named Entity Recognition**

## **Data Preparation for Neural Networks**

## **Training Neural Network for Named Entity Recognition**

## **Testing the Named Entity Recognition (NER) Model**

## ****MODULE DESCRIPTION & SNAPSHOTS :****

## MODULE I :

## LOADING THE DATA FOR NER:

## Initially the dataset is uploaded. And the data of the some samples are viewed.

## 

## MODULE II:

## **Data Preparation for Neural Networks:**

## I will train a Neural Network for the task of Named Entity Recognition (NER). So we need to do some modifications in the data to prepare it in such a manner so that it can easily fit into a neutral network. I will start this step by extracting the mappings that are required to train the neural network.

## 

## Now I will transform the columns in the data to extract the sequential data for our neural network

## 

## Now I will split the data into training and test sets. I will create a function for splitting the data because the LSTM layers accept sequences of the same length only. So every sentence that appears as integer in the data must be padded with the same length.

## 

Summary produced:



**MODULE III:**

## ****Training Neural Network for Named Entity Recognition:****

## Now, I will proceed with training the neural network architecture of our model. So let’s start with importing all the packages we need for training our neural network

## 

## The layer below will take the dimensions from the LSTM layer and will give the maximum length and maximum tags as an output

## 

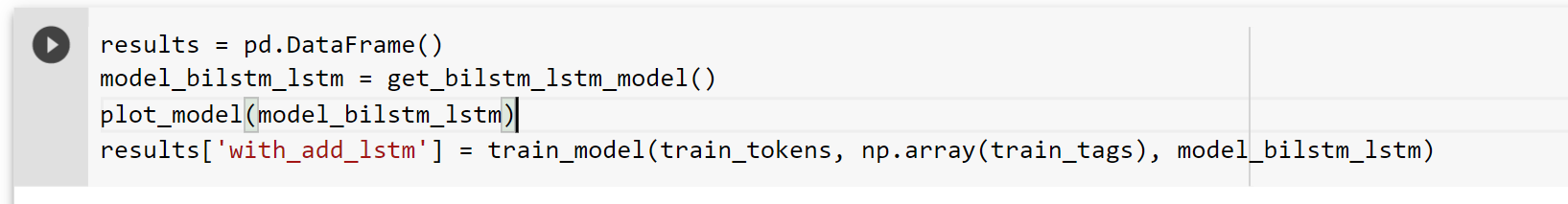
## Now I will create a helper function which will help us in giving the summary of every layer of the neural network model for Named Entity Recognition (NER)

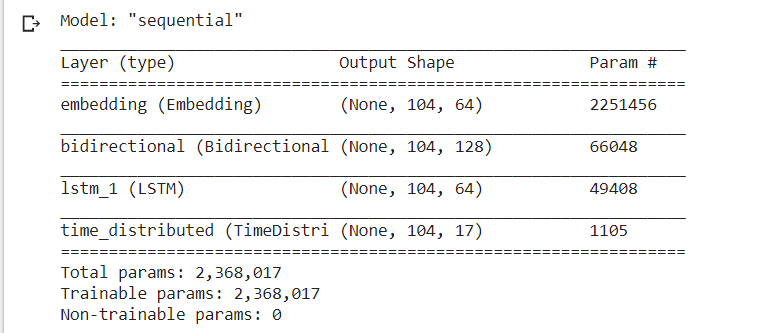
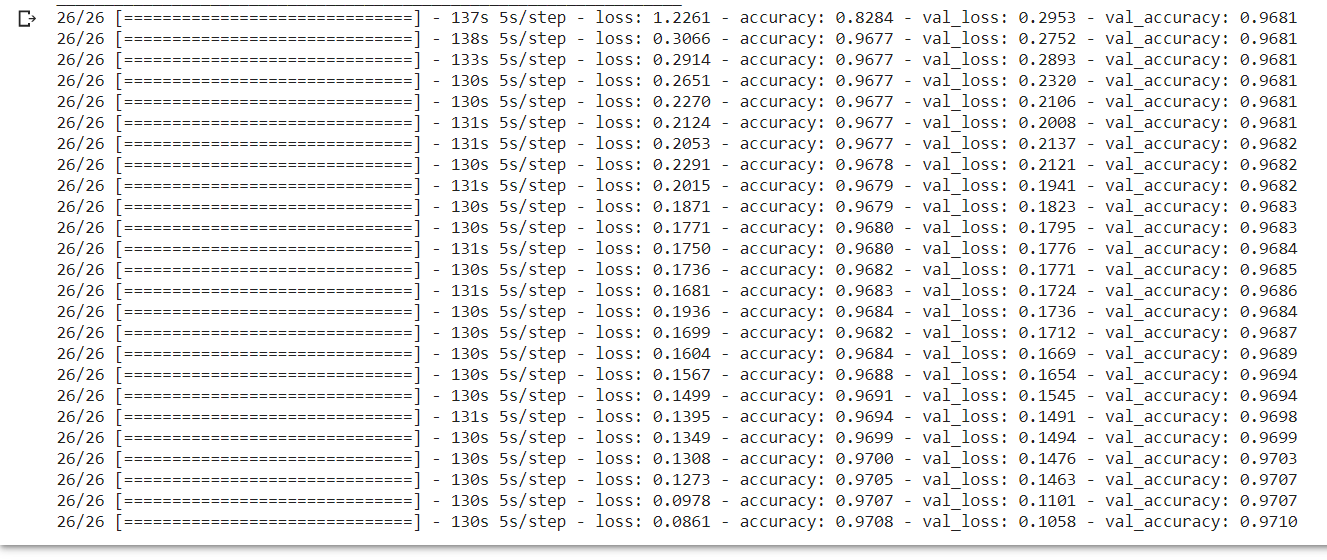
## 

## Now I will create a helper function to train the Named Entity Recognition model

## 

#### **Driver code:**



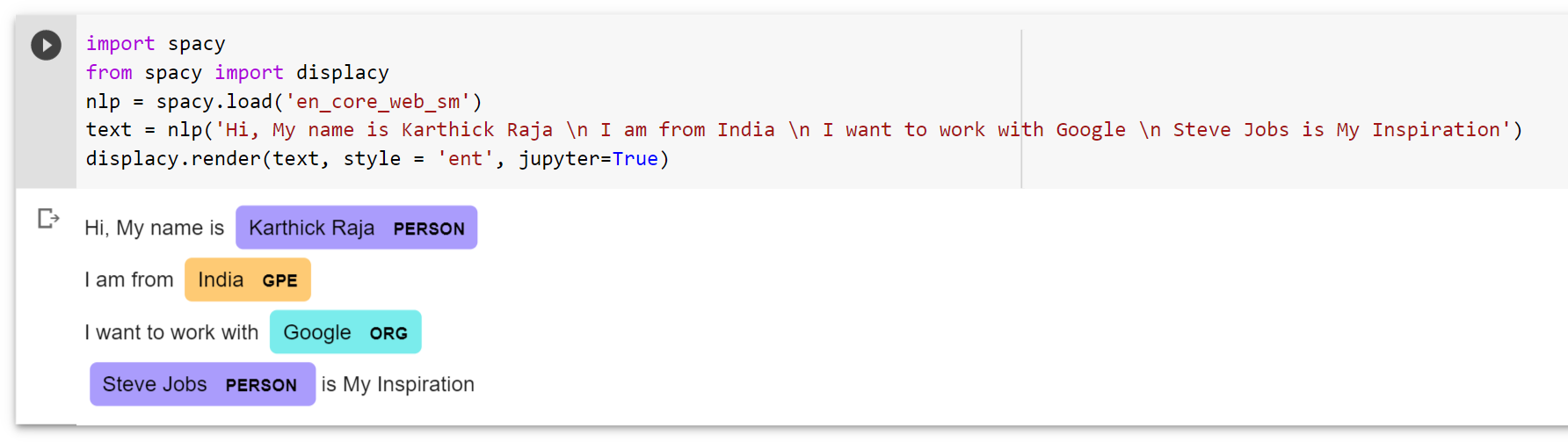
 

**ACCURACY: 97%**

**MODULE IV:**

**Testing the Named Entity Recognition (NER) Model:**

Now let’s test our model on a piece of text.



* ORG = Organization
* PERSON = Person
* GPE = Geopolitical Entity

## ****CONCLUSION:****

## **Thus the model developed successfully predicted the entities in the given text.**