

Named Entity Recognition

Team: Srish, Himanshu, Aayush



Outline

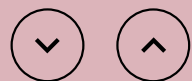
Introduction >

About Data >

Methodology >

Results >

Demo App >



F.B.I. Agent Peter Strzok **PERSON**, Who Criticized Trump
contentSkip to site indexPoliticsSubscribeLog InSubscribeLog
Who Criticized Trump **PERSON** in Texts, Is FiredImagePete
investigation after his disparaging texts about President Trump
TimesBy Adam Goldman **ORG** and Michael S. SchmidtAug
PERSON, the F.B.I. **GPE** senior counterintelligence agent
oversee the Hillary Clinton **PERSON** email and Russia **G**
said Monday **DATE**. Mr. Trump and his allies seized on the
Lisa Page — in **PERSON** assailing the Russia **GPE** inve
DATE at the F.B.I. **GPE** to become one of its most experi
inquiry. Along with writing the texts, Mr. Strzok **PERSON** wa
F.B.I. **GPE** had been under immense political pressure by
DATE from the staff of the special counsel, Robert S. Muel
Twitter **EVENT**, and on Monday **DATE** expressed satis
when Mr. Strzok **PERSON**'s conduct was laid out in a wide
Hillary Clinton's **PERSON** emails in the run-up to the 2016

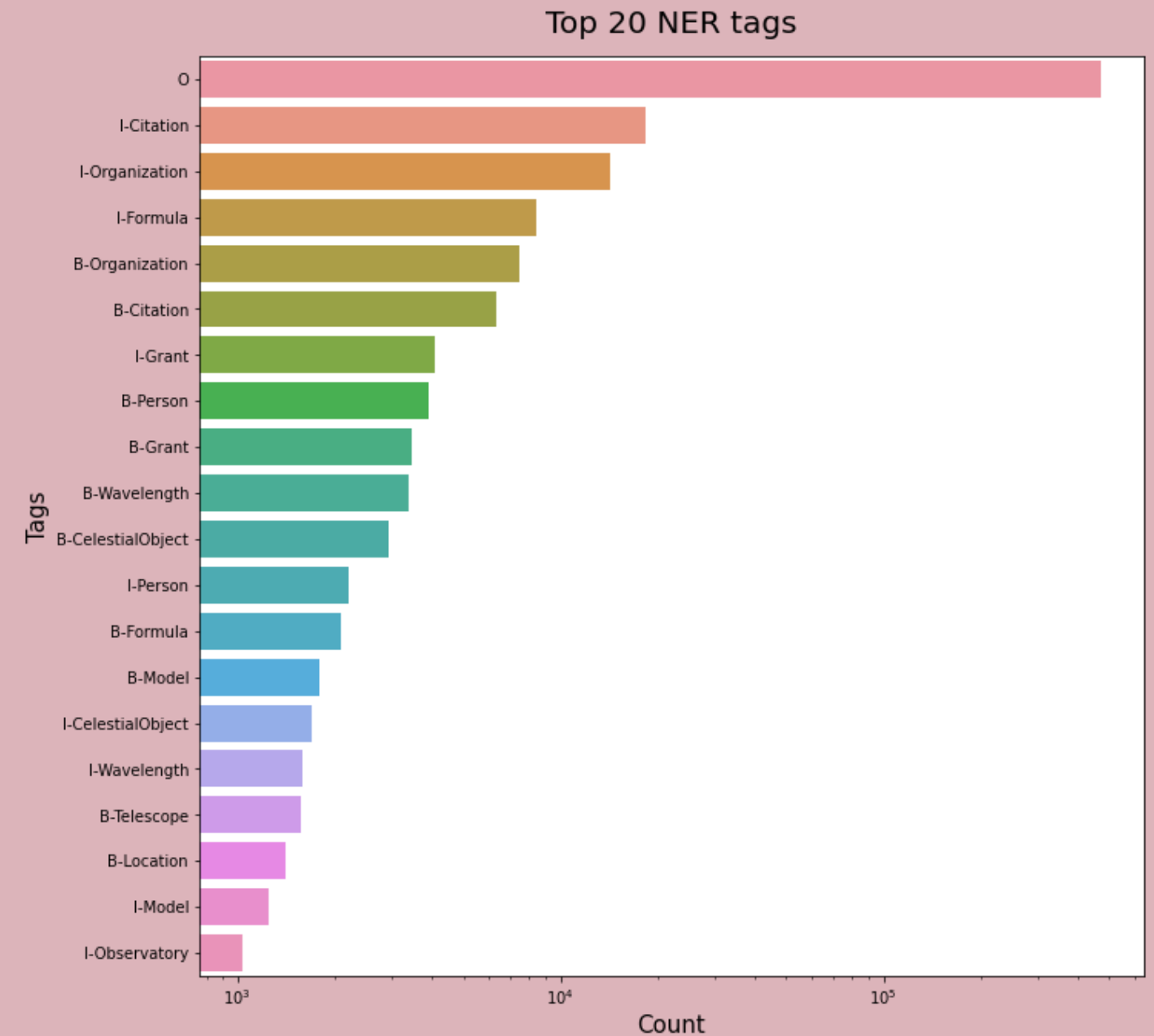
NEXT

Introduction

- The number of scientific papers published per year has exploded in recent years.
- Named - Entity Recognition can be used to optimize their storage and usage.
- This project involves identifying key tokens of scientific papers published in astronomy and classifying them into predefined entities.

Data

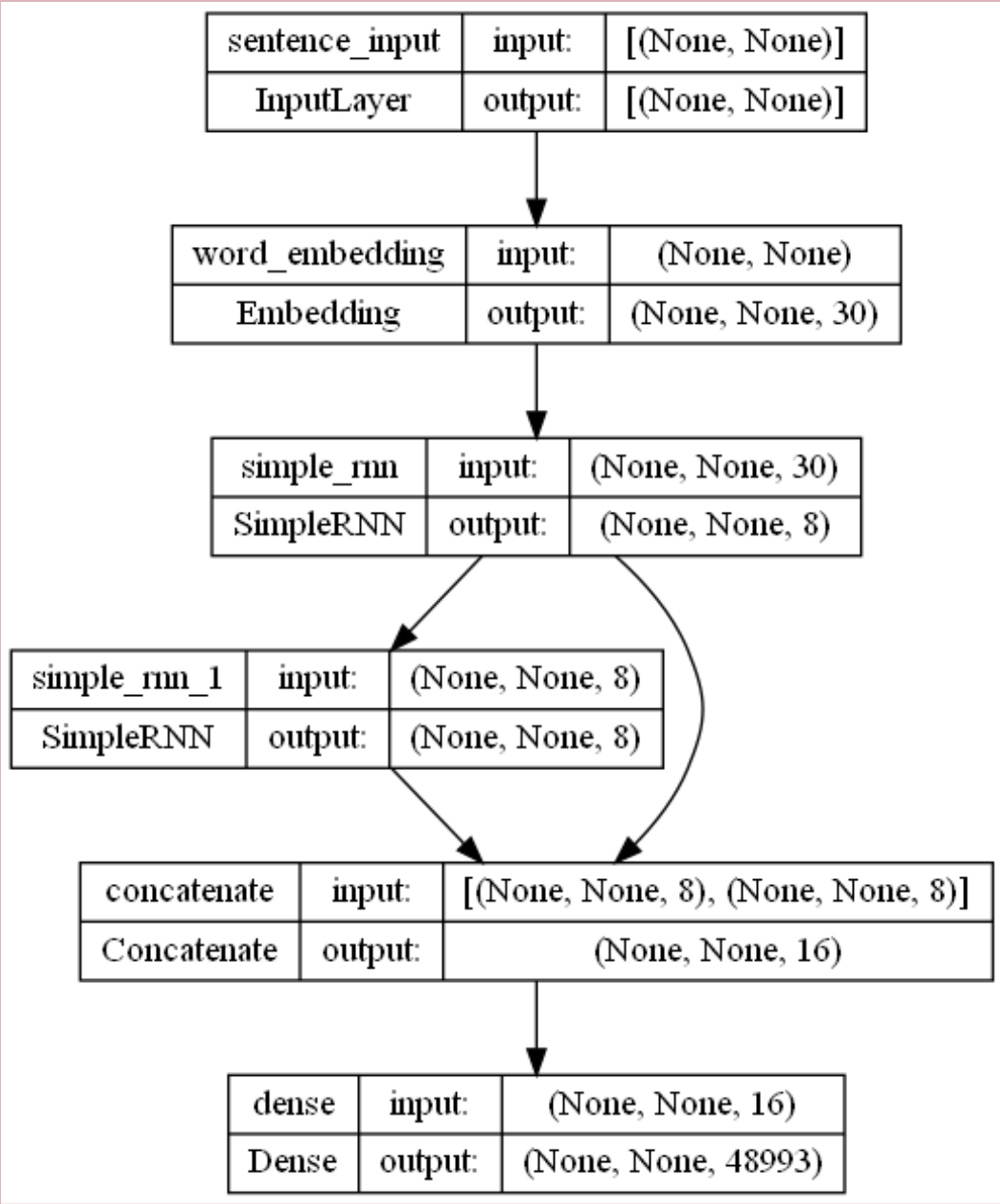
- We use the WIESP dataset(fgrezes/WIESP2022-NER · Datasets at Hugging Face).
- contains a collection of 1753 text fragments from astrophysics papers.
- The text fragments are in tokenized form along with the NER tags in IOB format.
- Fig shows the distribution of the top 20 tags.



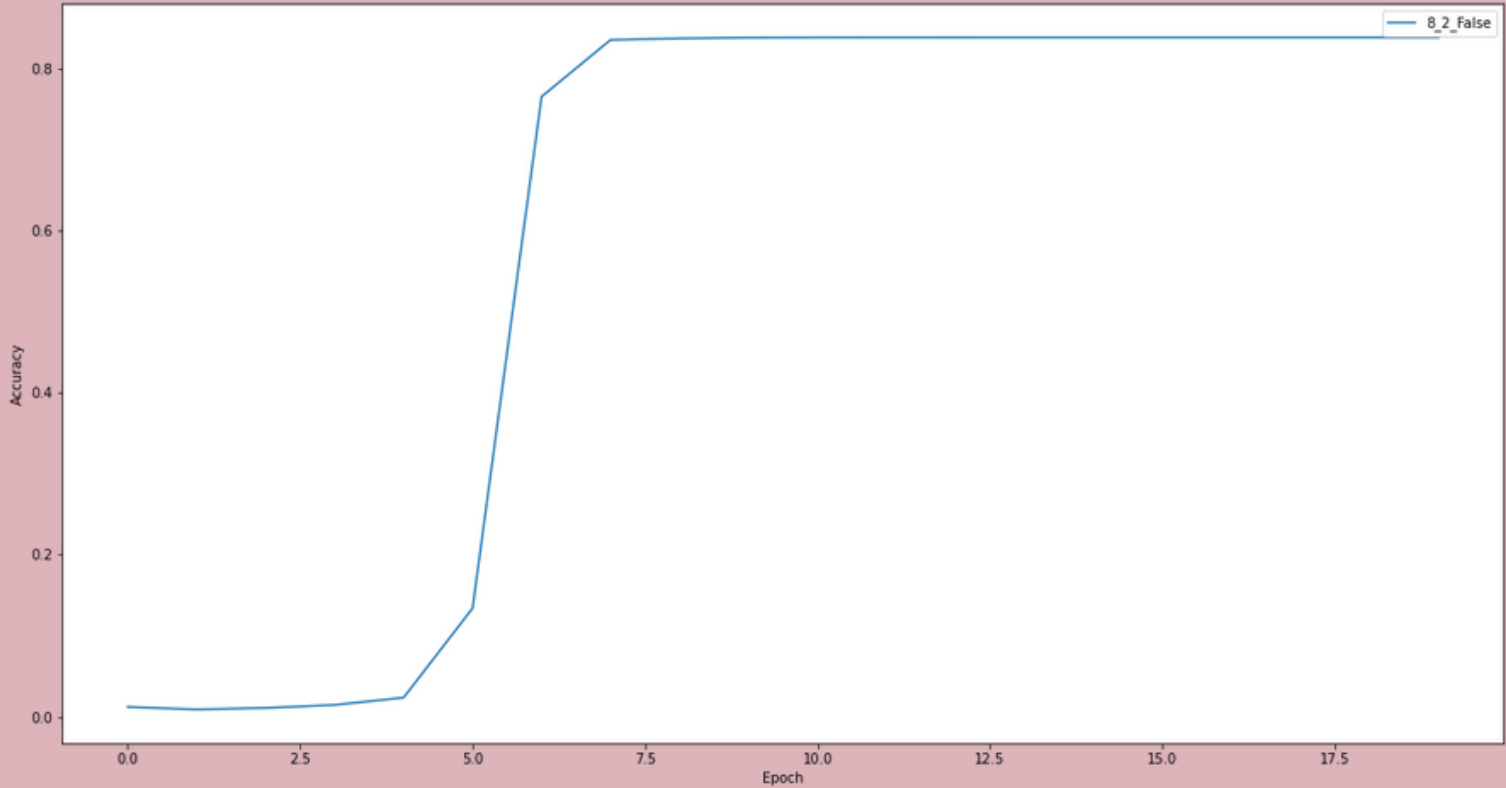
Methodology & Results

Baseline Model

→

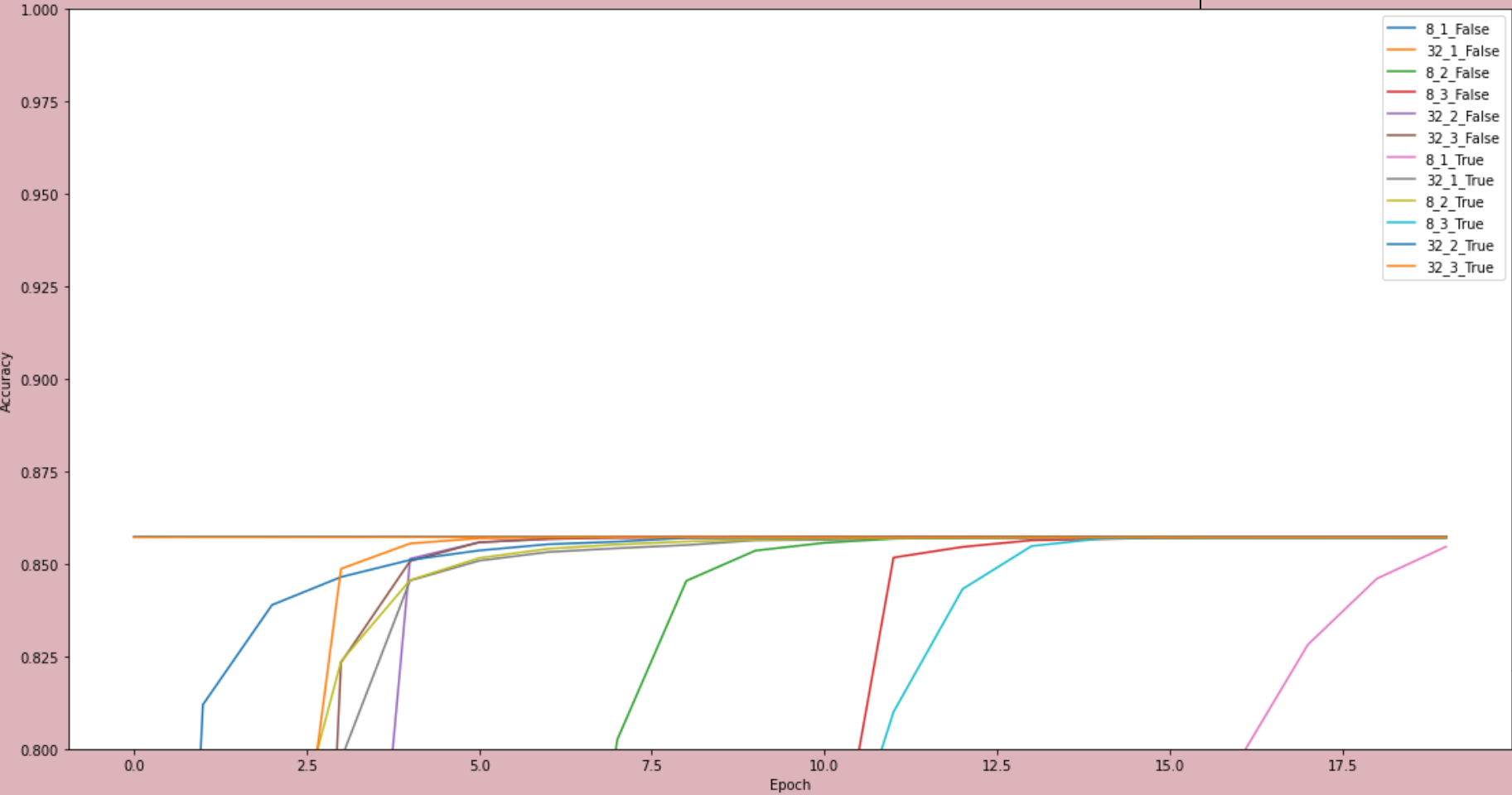
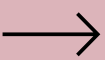


- We used 2 stacked SimpleRNNs as our baseline model.
- The accuracy of this model is approx 82%.



Methodology & Results

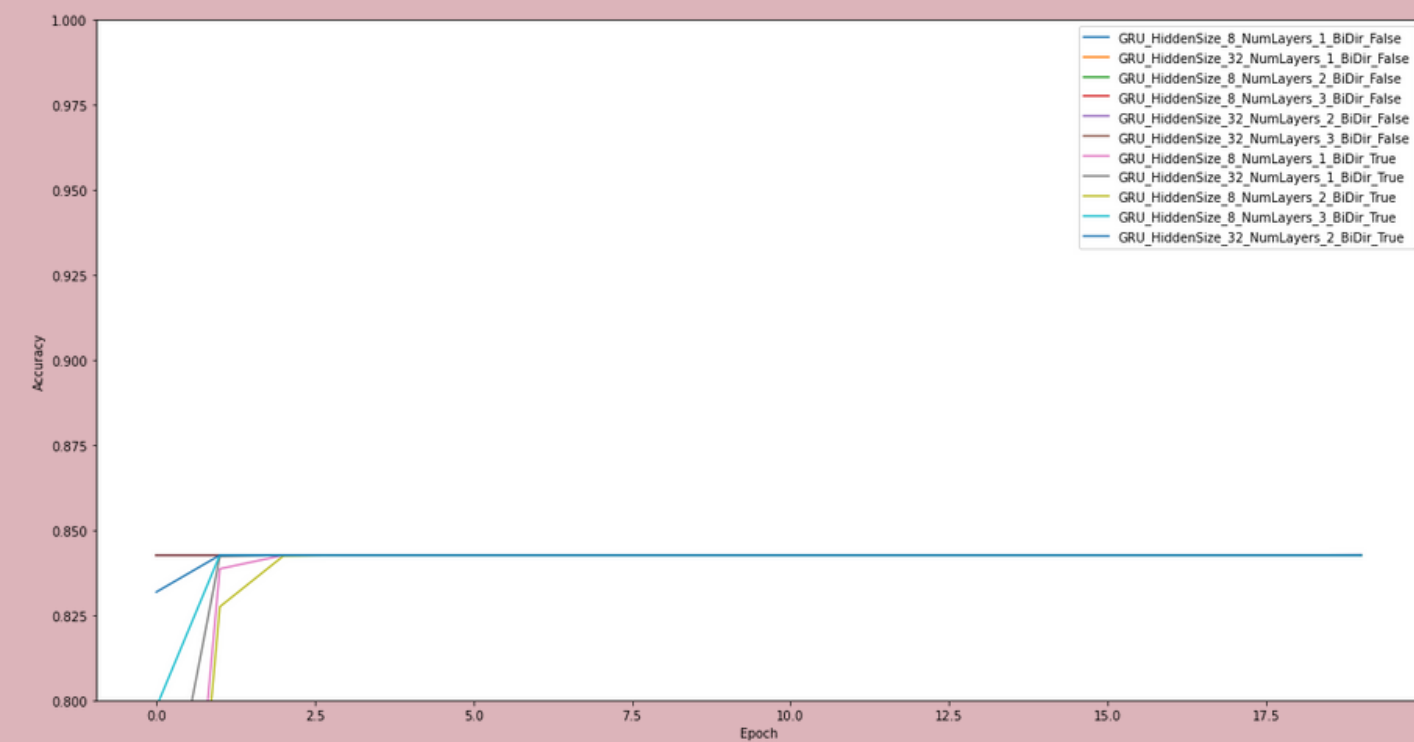
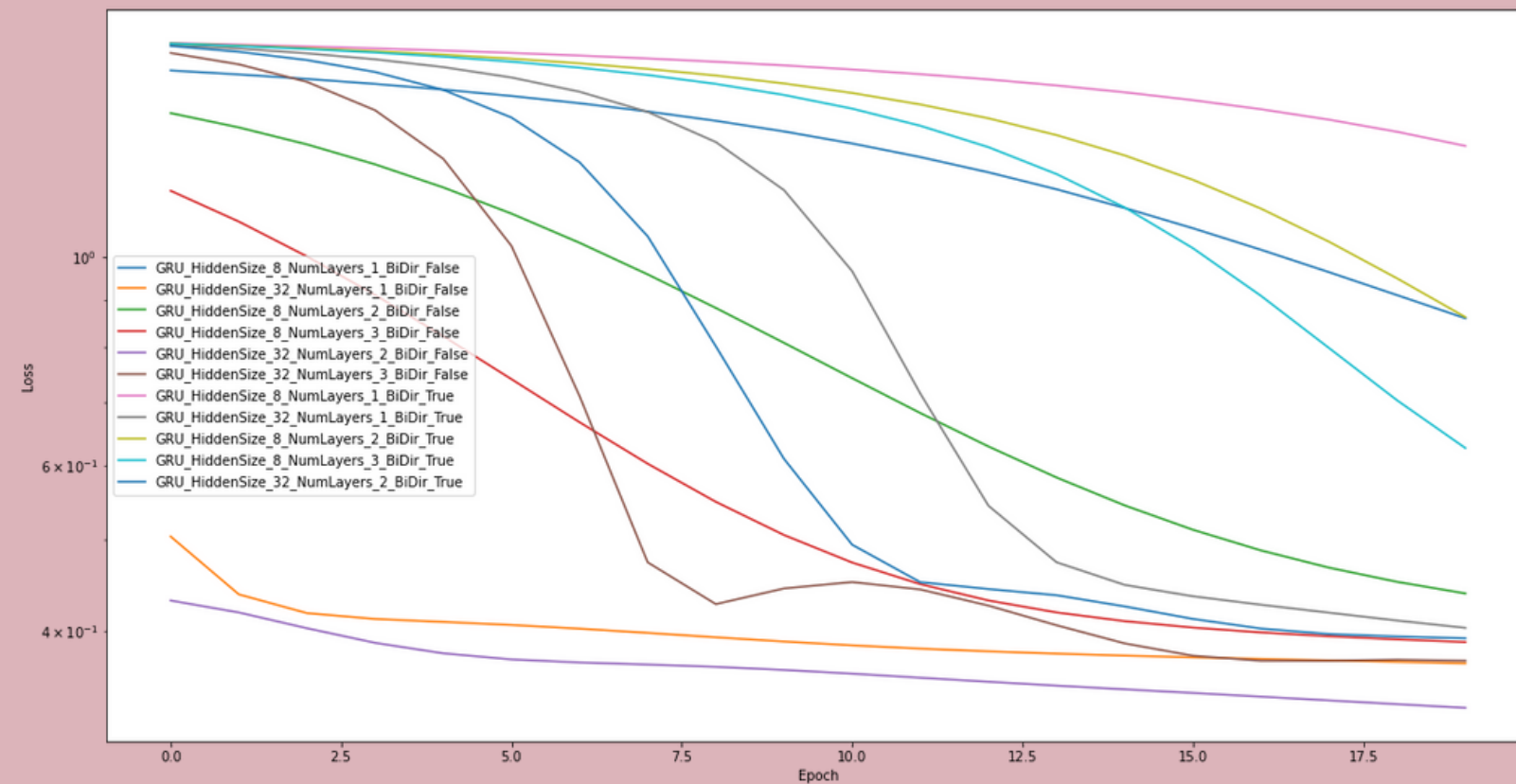
Hyper parameter tuning RNN model



- Using the baseline model we created a modular code to tune this model on different hyper parameters.
- The accuracy for all the experimentation is close to 86%

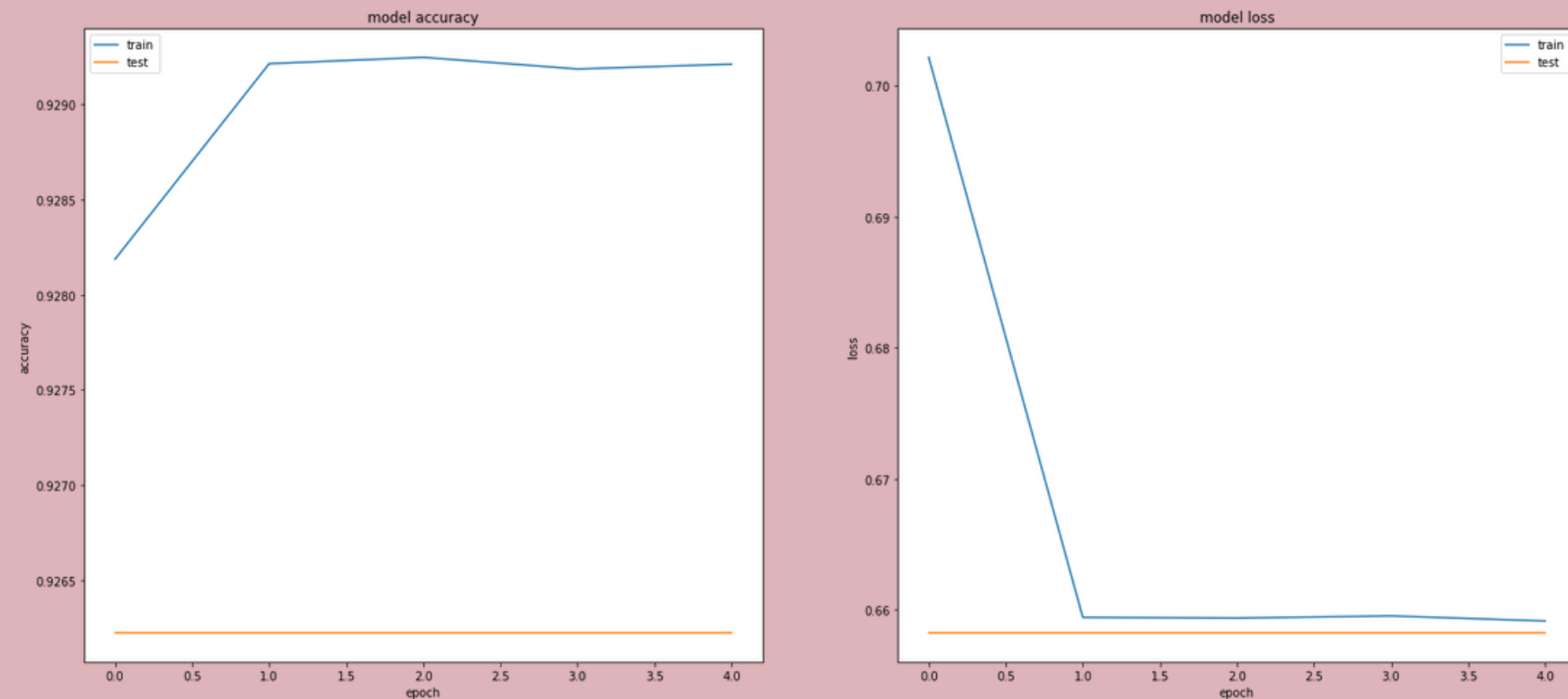
Methodology & Results

Gated Recurrent Unit(GRU)

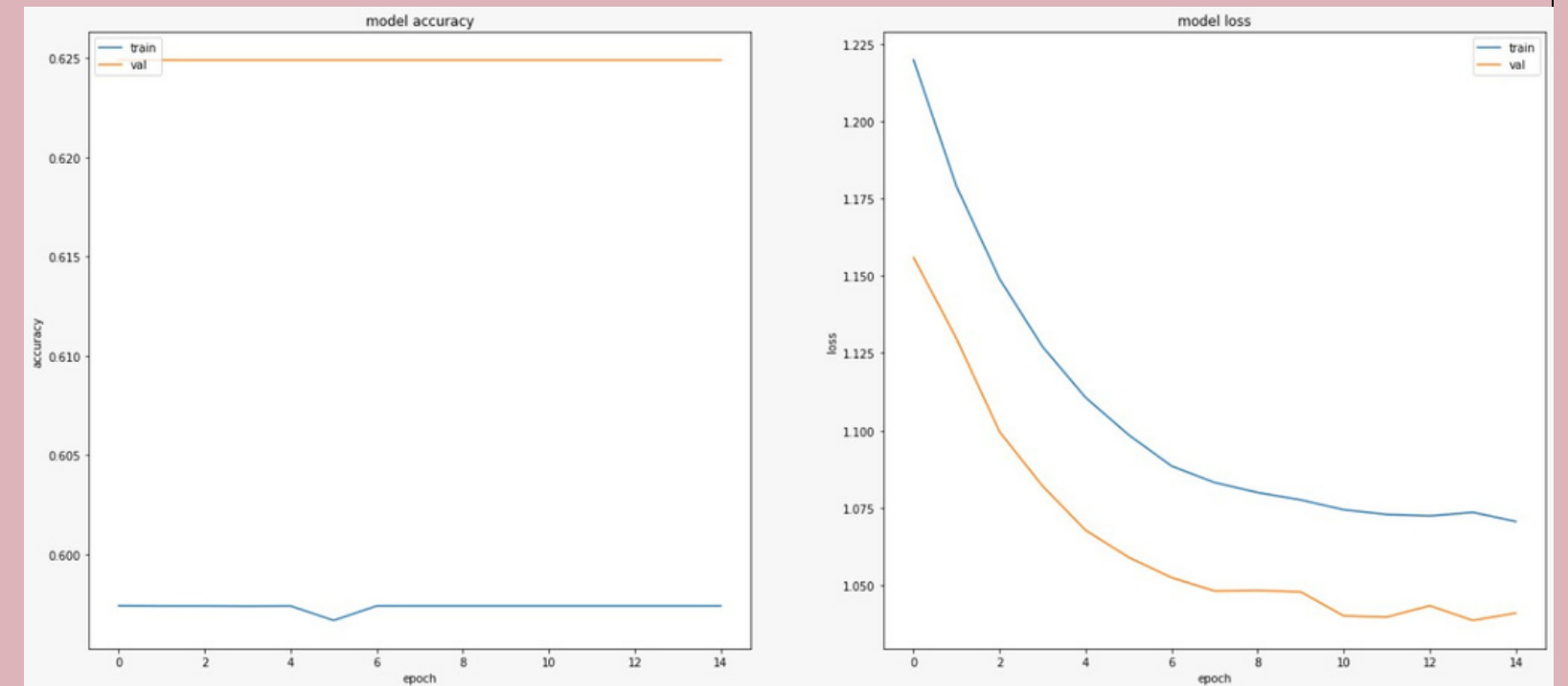


- Again we did hyper parameter tuning on GRU NN
- Accuracy obtained by this experimentation is close to 84%

Methodology & Results



BERT - Large (Huggingface)

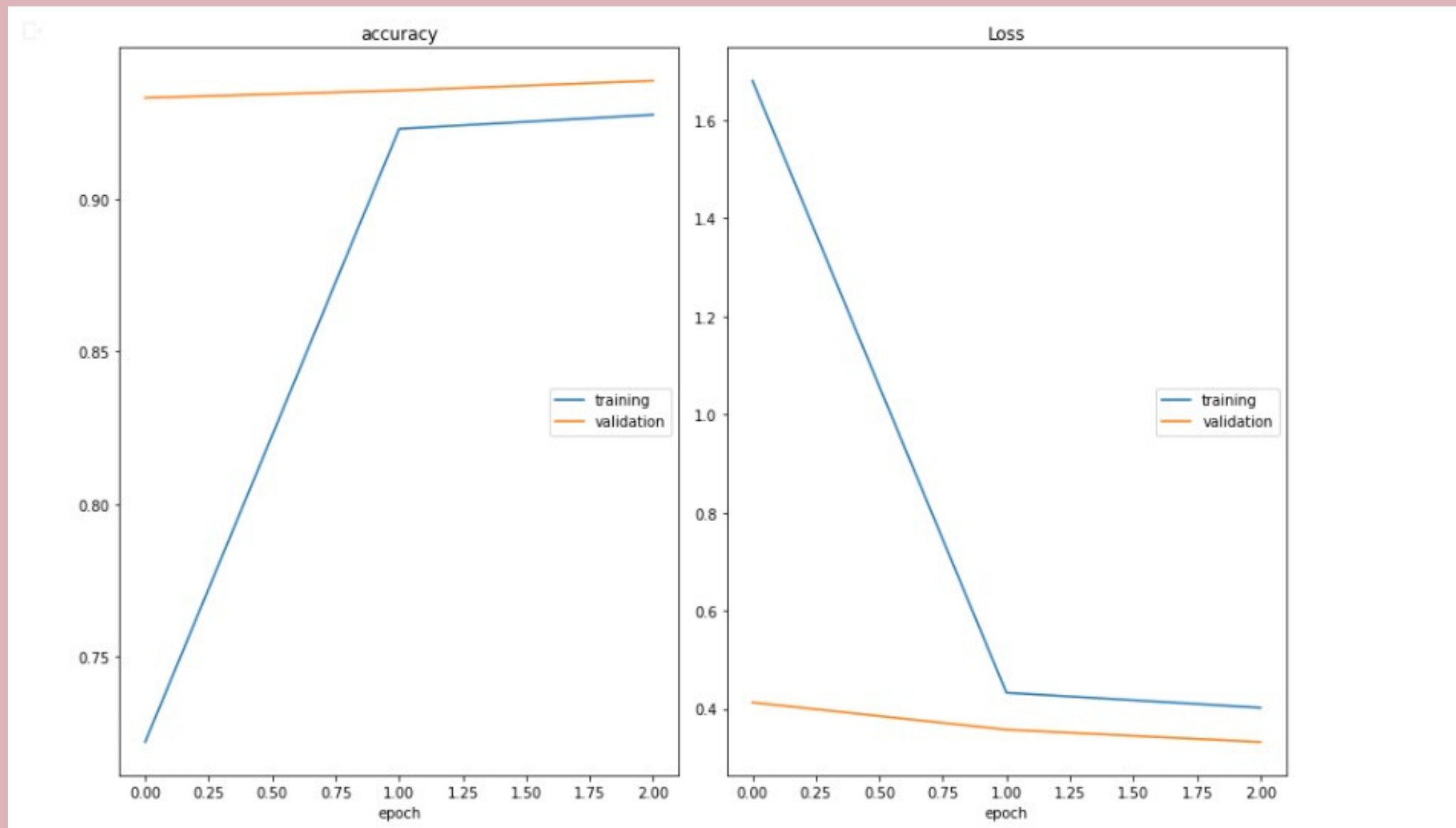


DistilBERT (Huggingface)

- We experimented with DistilBERT & BERT-Large transformer model by Huggingface
- Accuracy obtained by this model is approx 92% & 62%.

Methodology & Results

Bidirectional LSTM with Time Distributed Layer



- We used Bidirectional stacked LSTM along with Time Distributed Layer
- The TimeDistributed layers allow layer operation across every output over every time-step
- This model gave us an accuracy of 93%

Demo app using spaCY

Name Entity Recognition on WIESP Dataset



Enter a text

Predict

Total execution time: 0.5313 seconds

*Thank
you!*