

# Organizzazione progetto

Il progetto prevede l'implementazione di 7 differenti reti neurali per due differenti dataset con la seguente organizzazione:

Ogni dataset possiede una propria cartella (IMDB, SLS), ognuna contenente 9 differenti file notebook, tra cui due file che non riguardano l'implementazione dei modelli:

- *{IMDB/SLS}\_dataset.ipynb*: all'interno di questo file vengono effettuate tutte le analisi del dataset in questione. Nello specifico la prima operazione che viene fatta è quella di verificare se sono presenti elementi nulli che risultano inutilizzabili ai fini dell'addestramento. Successivamente viene analizzata la distribuzione dei dati, ovvero se le label positivo/negativo sono equamente distribuite all'interno del dataset. Infine viene fatto dove necessario una fase di pre-processing.
- *{IMDB/SLS}BiGRUAttHP\_complete.ipynb*: questo file nello specifico è stato creato per effettuare una analisi approfondita degli iper-parametri del modello, utilizzando uno strumento messo a disposizione dal framework tensorflow denominato **tensorboard**.

## Struttura modelli

Ogni modello costruito per entrambi i dataset è presente all'interno del relativo notebook della cartella associata al dataset in questione. Ad esempio tutti i modelli associati al dataset IMDB saranno contenuti all'interno della cartella IMDB denominati nella forma:

```
IMDB_TipoModello.ipynb
```

Inoltre all'interno di ogni notebook è possibile anche osservare tutte le informazioni inerenti alla fase di validazione tramite il comando

```
print(classification_report(y_test, np.argmax(y_pred, axis=1).astype("f
```

che restituisce la classificazione di ogni classe in termini di Accuracy, Precision, Recall e F1-Score

## tensorboard

Come descritto precedentemente, dopo aver effettuato l'addestramento dei modelli con i parametri descritti dall'articolo assegnato, è stato deciso di effettuare una ulteriore analisi del migliore modello trovato (BiGRU con meccanismo di attenzione) variando i seguenti parametri:

- Hidden Units: 16, 32, 64
- Dropout: 0.1, 0.2, 0.3, 0.4, 0.5, 0.6, 0.7, 0.8, 0.9
- Optimizer: "adam", "RMSProp"

Di cui la migliore configurazione per il dataset SLS risulta essere:

- Hidden Units: 16
- Dropout: 0.4
- Optimizer: "RMSProp"

Mentre per il dataset IMDB:

- Hidden Units: 16
- Dropout: 0.9
- Optimizer: "RMSProp"

Infine, una volta trovato la migliore configurazione per cercare di migliorare ulteriormente i risultati si è deciso di provare anche a modificare il learning rate con i seguenti valori:

- learning rate: 0.0001, 0.001, 0.005, 0.009, 0.1, 0.2, 0.25, 0.3

Di seguito dunque, vengono riportate le migliori configurazioni dei modelli BiGRU con meccanismo di attenzione per dataset.

SLS:

- Hidden Units: 16
- Dropout: 0.4
- Optimizer: "RMSProp"
- learning rate: 0.005

IMDB:

- Hidden Units: 16
- Dropout: 0.9
- Optimizer: "RMSProp"
- learning rate: 0.009

Di seguito vengono mostrati tutti i link a cui è possibile accedere per osservare i risultati ottenuti:

- Dataset SLS con variazione di Hidden Units, Dropout, Optimizer:  
<https://tensorboard.dev/experiment/kXqr7OGLQoi1VPUm1H6Nww/>
- Dataset SLS con variazione del learning rate:  
<https://tensorboard.dev/experiment/Yj2SeXwOS1K6x3A2T2IQmw/#hparams>
- Dataset IMDB con variazione di Hidden Units, Dropout, Optimizer:  
<https://tensorboard.dev/experiment/GVI9mRIMSD27E1FBhY1jnA/#scalars>
- Dataset IMDB con variazione del learning rate:  
<https://tensorboard.dev/experiment/0Qklwg2CQ1yQYqQoIZ0Ahg/#hparams>