Report

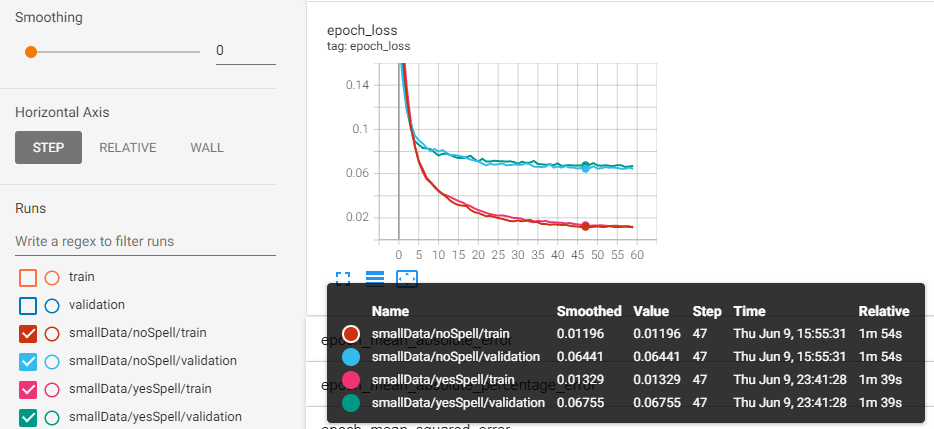
# \*Undersampling\* method

## All search terms

Dataset doesn’t filtered by search term of tweets

### Scenario1 (1500)

* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **2 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:60



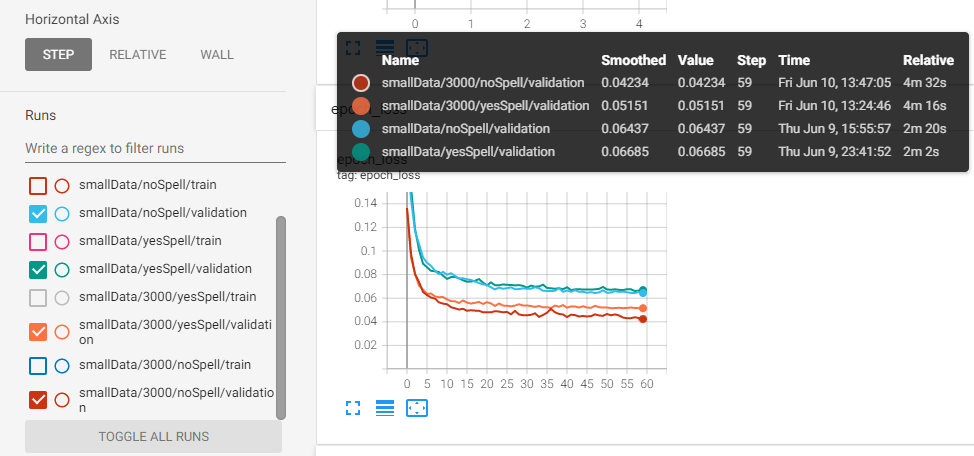
### Scenario2 (3000)

* size=3000 (1000 neg, 1000 neu, 1000 pos)
* Model: **2 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:60



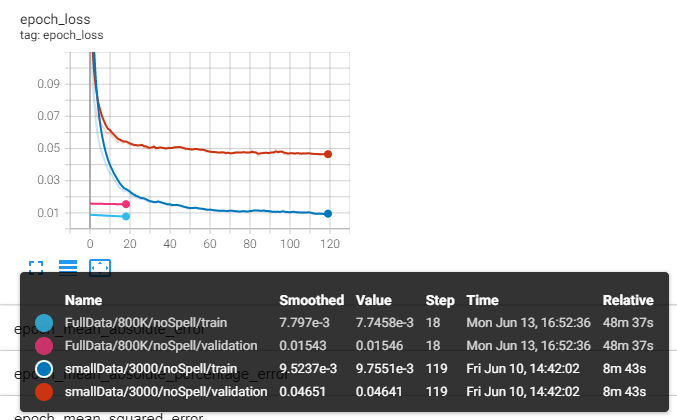
### Compare validations

Compare validation loss of two scenario in yesSpell and noSpell modes



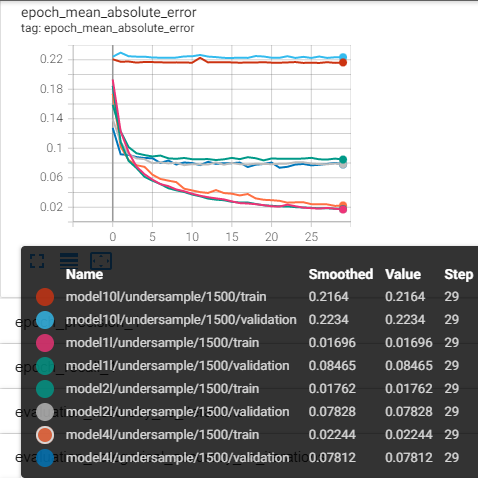
### Compare best with full data

* Full dataset means 820K samples. But it’s better because data is imbalanced.



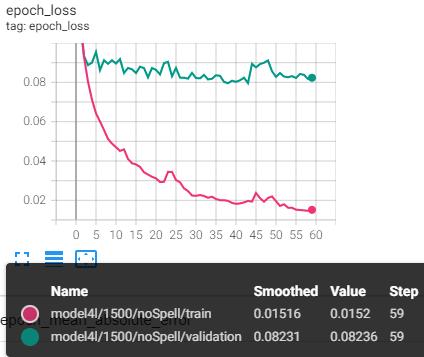
## Just \*bitcoin\* search term

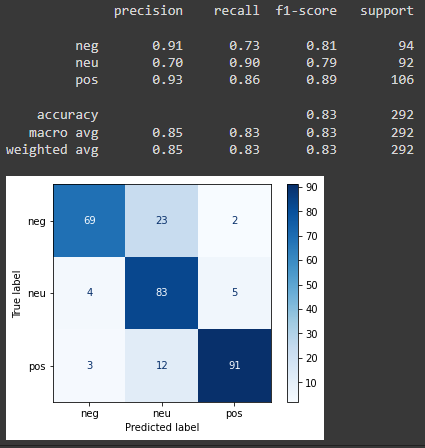
### Compare multiple layers of bi-lstm



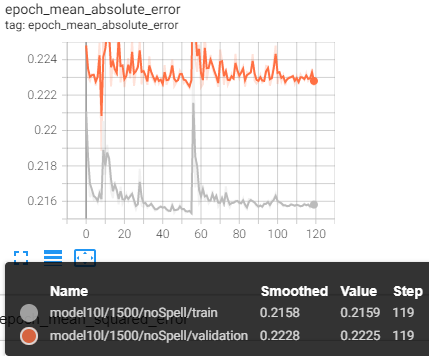
### 3L\_regression\_1500data

* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:60



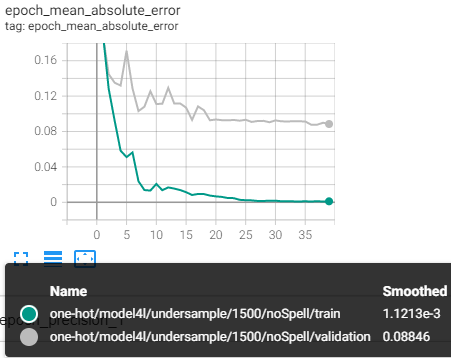


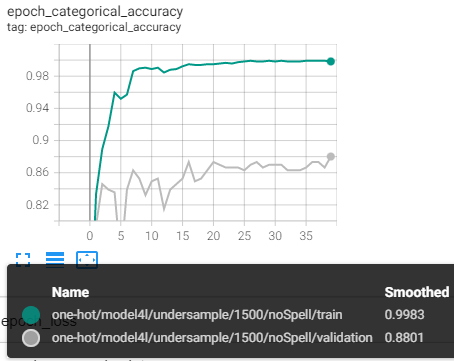
* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **10 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:120

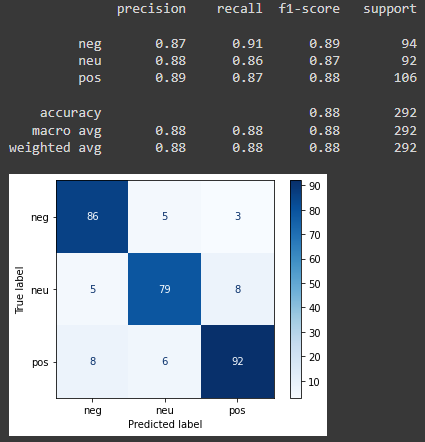


### 3L\_oneHot\_1500data

* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:categorical\_crossentropy, epochs:40

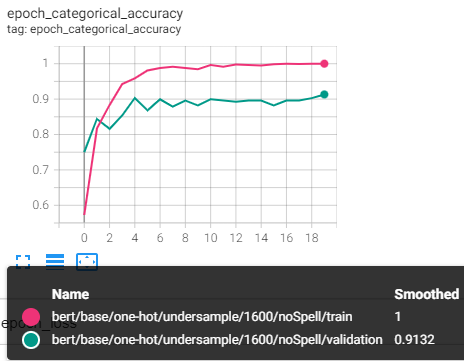


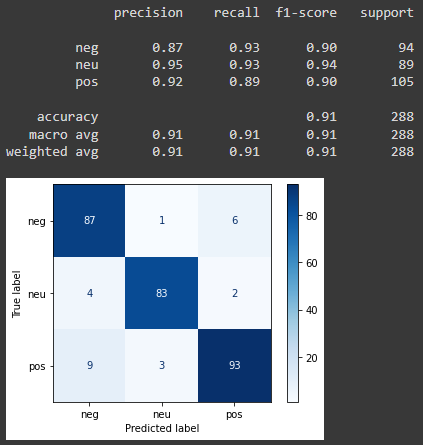




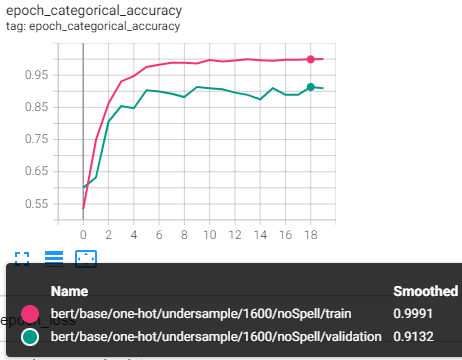
### BERT\_oneHot\_1500data

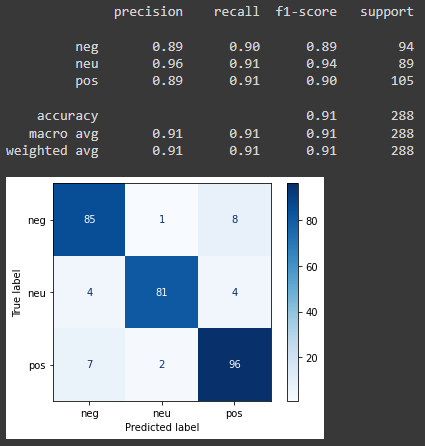
* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **bert\_en\_uncased\_L-12\_H-768\_A-12+batchNorm+dropout+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:30





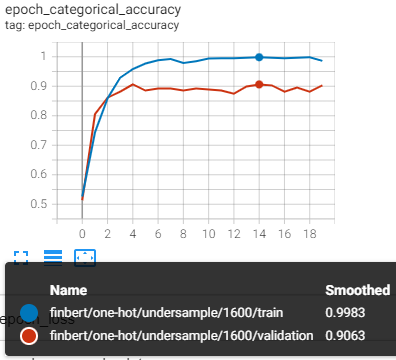
* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **bert\_en\_uncased\_L-12\_H-768\_A-12+batchNorm+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20

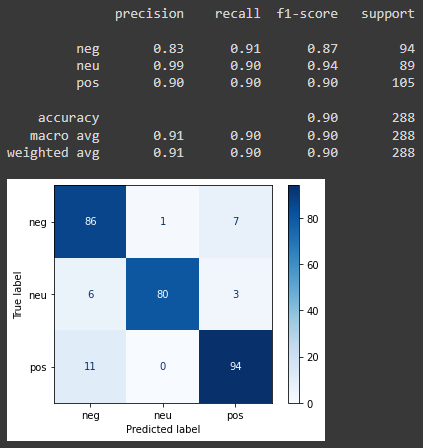




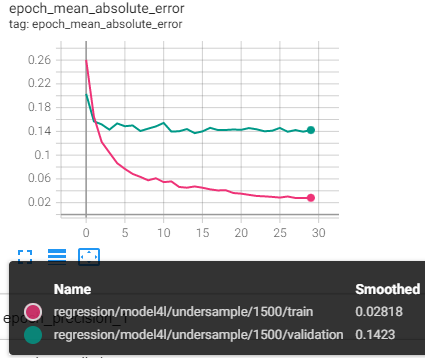
### FinBert\_oneHot\_1500data

* size=1500 (500 neg, 500 neu, 500 pos)
* Model: **ProsusAI/finbert : transformers bert+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20

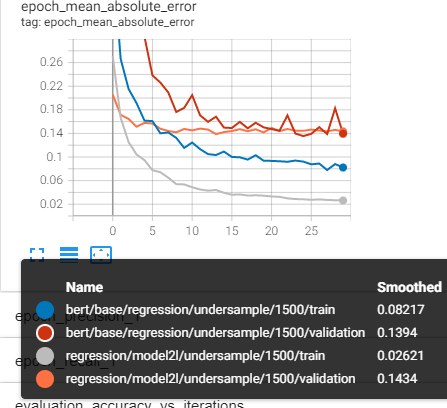




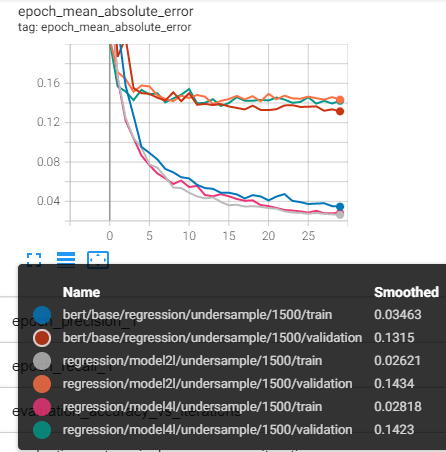
### 4lstm\_Regression(-1,1)\_1500data



Bert+batch+1dense



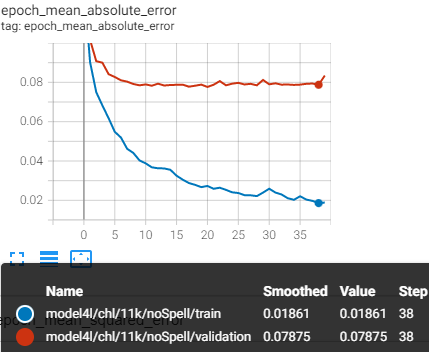
### Bert\_regression(-1,1)(Bert+1dense)

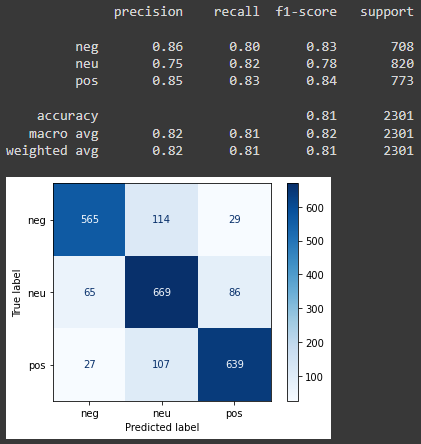


# \*Change label value\* method

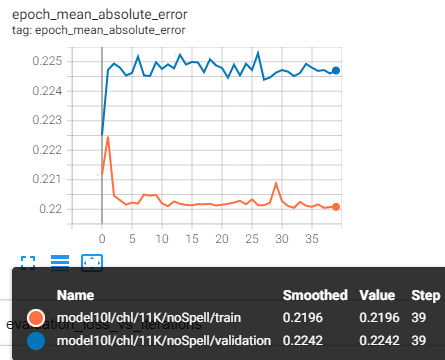
### 3L\_regression\_11K data

* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:40



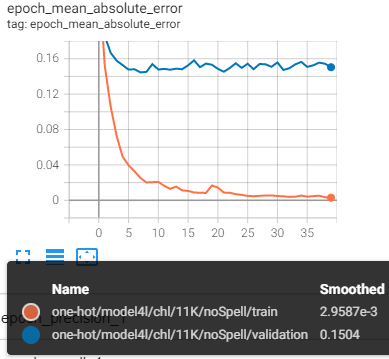


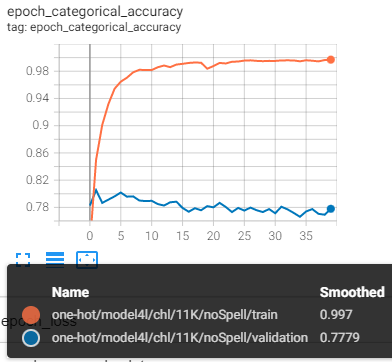
* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **10 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:40

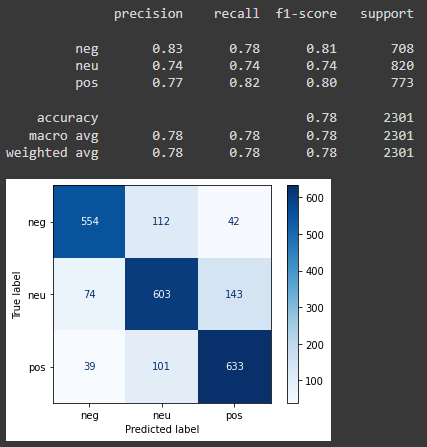


### 3L\_oneHot\_11K data

* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:categorical\_crossentropy, epochs:40

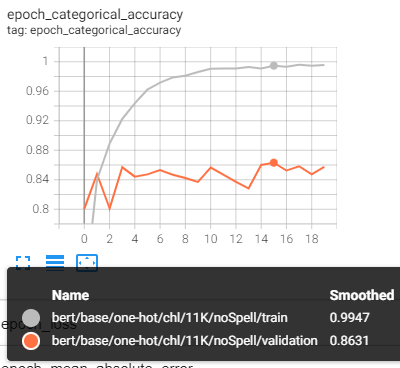


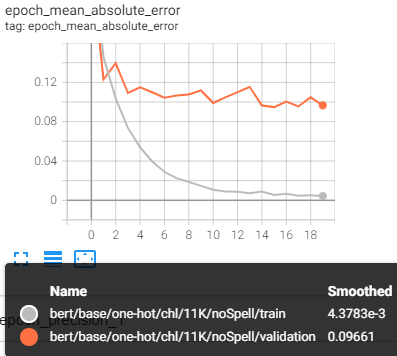


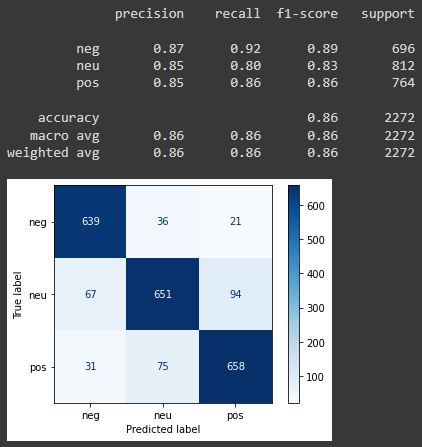


### BERT\_oneHot\_11K data

* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **bert\_en\_uncased\_L-12\_H-768\_A-12+batchNorm+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20

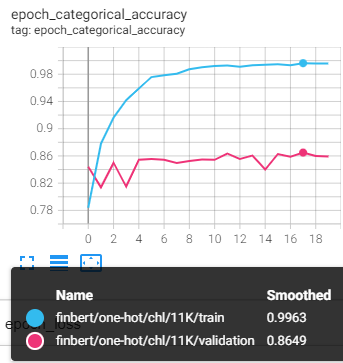


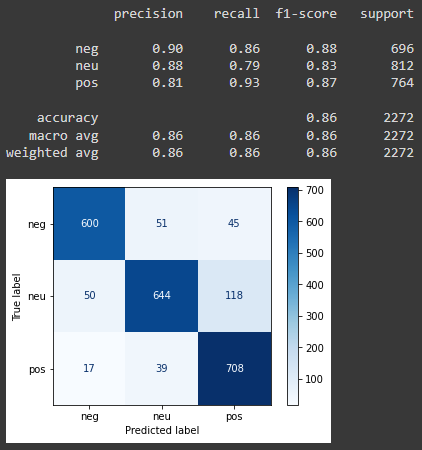




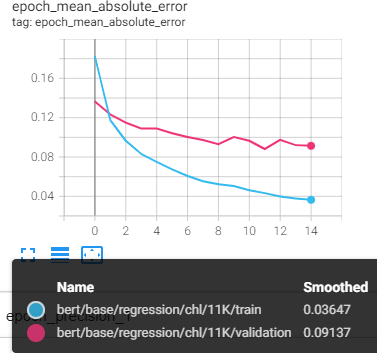
### FinBert\_oneHot\_11K data

* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **ProsusAI/finbert : transformers bert+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20

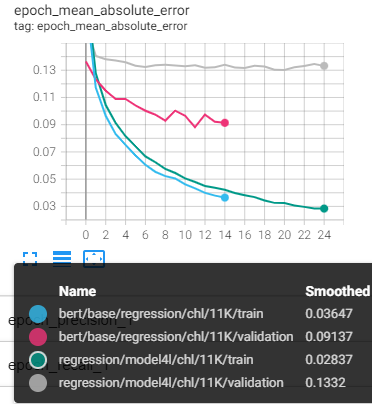




### bert\_regression(-1,1)



### 4lstm\_regression(-1,1)

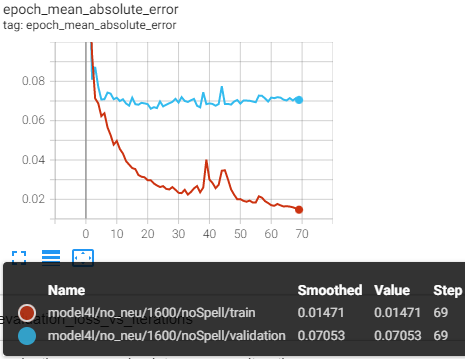


# \*removal of neutral data\*

## Main dataset (1600)

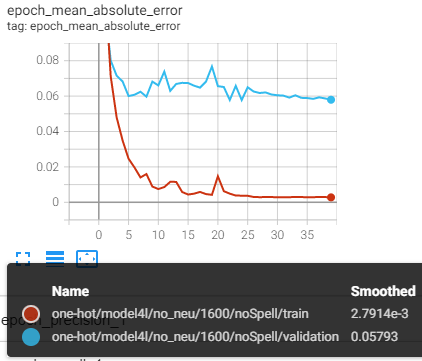
### 3L\_regression

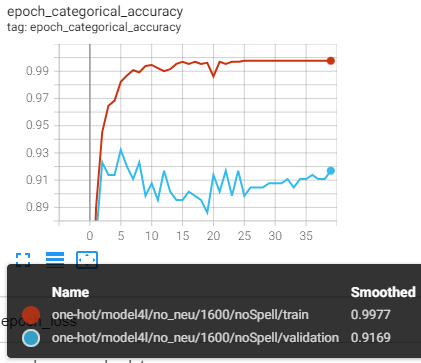
* size=1600 (500 neg, 1000pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:70

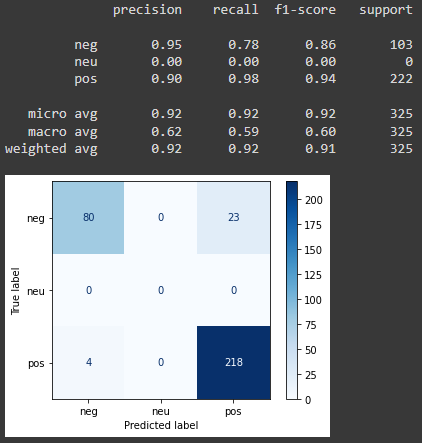


### 3L\_oneHot

* size=1600 (500 neg, 1000pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:categorical\_crossentropy, epochs:40

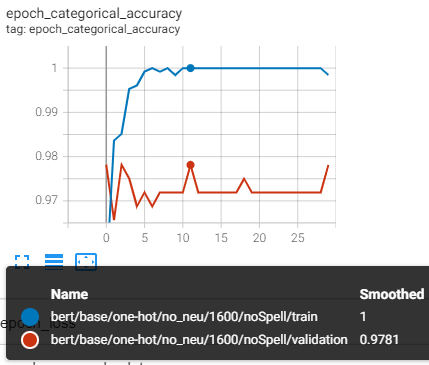


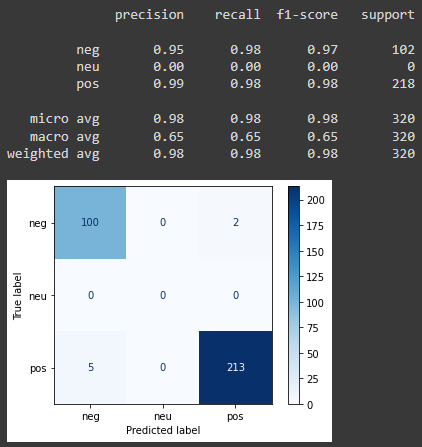




### BERT\_oneHot

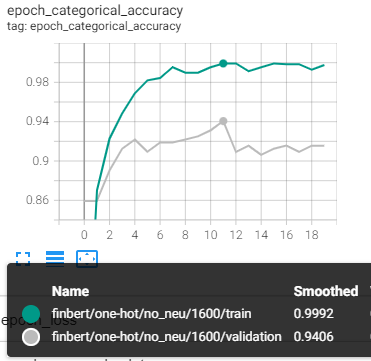
* size=1600 (500 neg, 1000pos)
* Model: **bert\_en\_uncased\_L-12\_H-768\_A-12+batchNorm+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:30





### FinBert\_oneHot

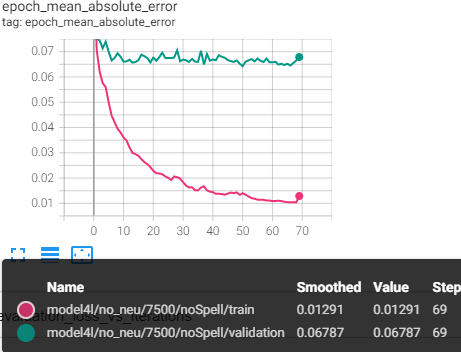
* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **ProsusAI/finbert : transformers bert+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20

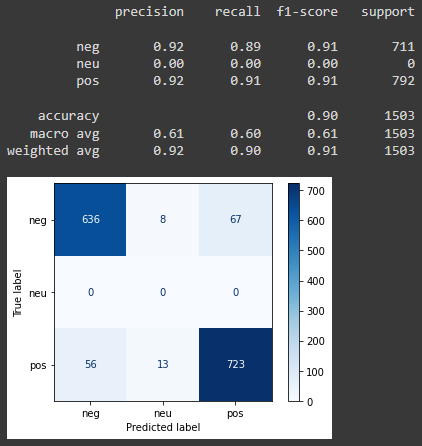


## Change label dataset (7500)

### 3L\_regression

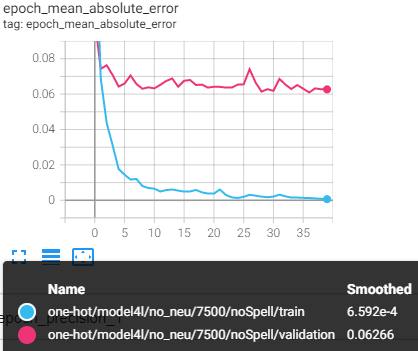
* size=7500 (3500 neg, 4000 pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:MAE, epochs:70

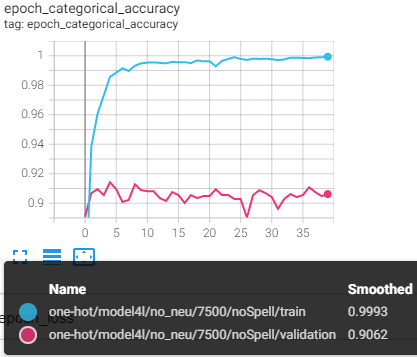


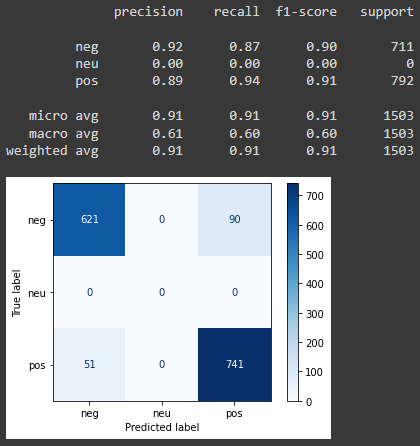


### 3L\_oneHot

* size=7500 (3500 neg, 4000 pos)
* Model: **4 layer BiLSTM**, optimizer:Adam, loss:categorical\_crossentropy, epochs:40

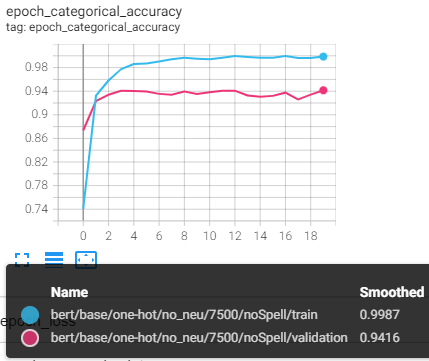


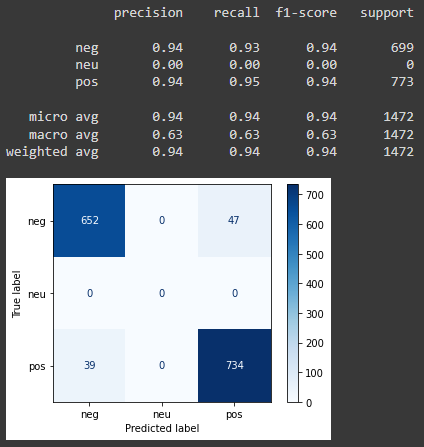




### BERT\_oneHot

* size=7500 (3500 neg, 4000 pos)
* Model: **bert\_en\_uncased\_L-12\_H-768\_A-12+batchNorm+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20





### FinBert\_oneHot

* size=11K (3500 neg, 4000neu, 4000pos)
* Model: **ProsusAI/finbert : transformers bert+dropout+128dense+3dense**, optimizer:Adam, loss:categorical\_crossentropy, epochs:20

