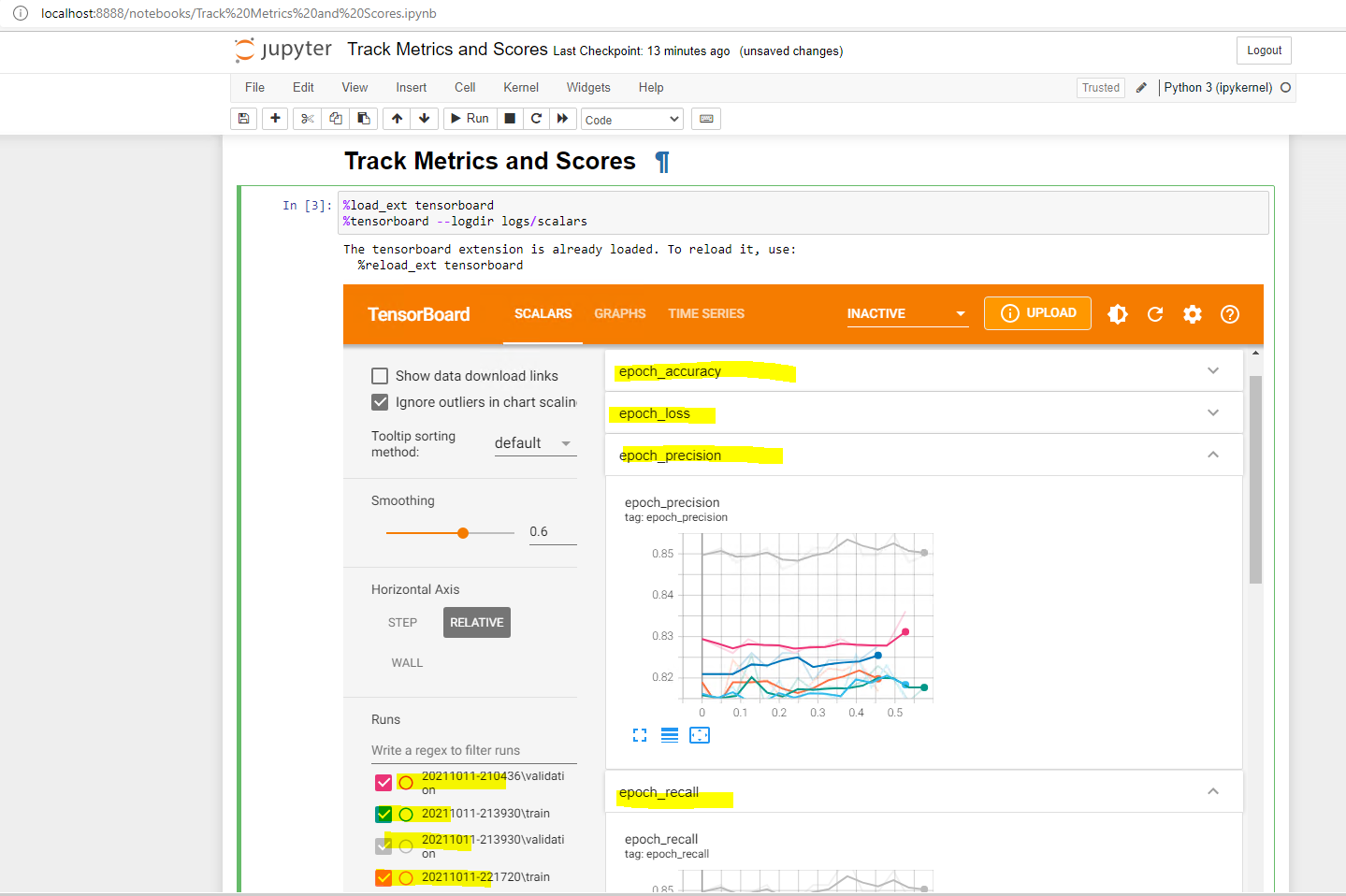
# Sentiment Analysis and Prediction

## Technicality and Approach

* Both the Sentence and Entity are embedded to form the input for training. The label is categorical 0 and 1 for “Negative” and “Positive” sentiments respectively.
* State of the art Bert architecture used to solve the problem statement.
* Stratified k-fold validation implemented during training.
* EarlyStopping implemented to avoid overfitting (loss monitored for each epoch and if there is no improvement for 10 epochs, training is stopped for that round of k-fold).
* Below Metrics used to track model performance.
  + Recall
  + Precision
  + Loss (Categorical Accuracy)
  + Accuracy

Metrics can be seen using Tensorboard (run the cells in Jupyter notebook file “Track Metrics and Scores.ipynb” check the document “How to setup and run the application.docx” ) .



## Basis for selecting the model.

It is ok if the model predicts positive reviews as negative as the companies can still reach out the customers for further action, but if negative reviews are falsely classified as positive, the company may not take any corrective actions on those and

may start losing customers pretty soon.

Therefore, preference to be given to Precision while selecting the model.

## Improvization to be done to the model.

1. Create a proper pipeline to separately embed Sentence and Entity for training.
2. Use Cosine similarity during training to handle contrastive sentences.
3. Precision can be improved by tweaking the early stopping mechanism.