**Brief approach**

I noticed pretty early that increasing the max sequence length increased the score sufficiently. This observation more or less dictated my approach. I used a basic XLNet model with hardly any feature engineering.

**Data pre-processing / Feature engineering**

Lower cased all sentences and masked the relevant drug in the sentences. Then proceeded to take the first 1380 tokens after the sentence piece tokenization.

I chose to the fill the GPU RAM with as much max sequence length as possible and refrained from using extra features. I tried to add variations to data but made implementation mistakes and ran out of time.

**Final Model**

I used 6 seeds to average the XLNet base cased model's predictions. Time didn't permit for any hyperparameter tuning.

**Key takeaways, if any?**

Really surprised that in a competition where deep learning could be the best solution, I could be competitive with just 8 GB GPU RAM machine.

**According to you, what are the 5 things a participant must focus on while solving such problems?**

Identifying the differences in train and test distribution may be crucial.

Most of the things are same as other competiitions