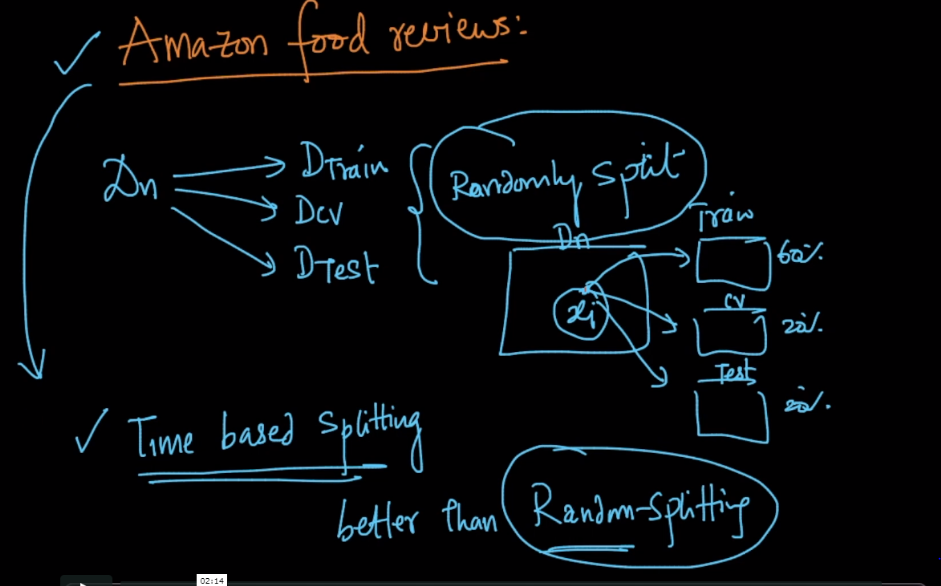
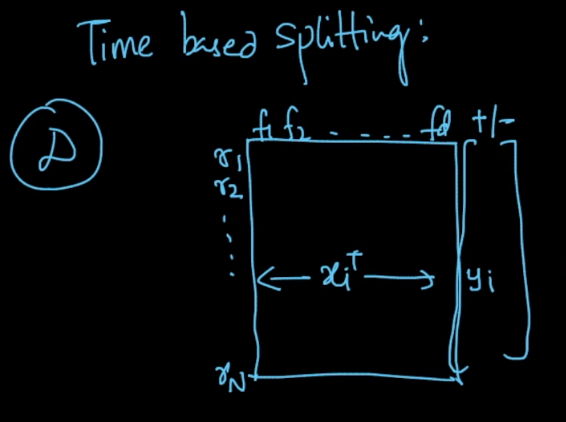
Till now we’ve seen random splitting, there some other splitting also, which are better than random splitting in some cases.

**Time Based splitting:**

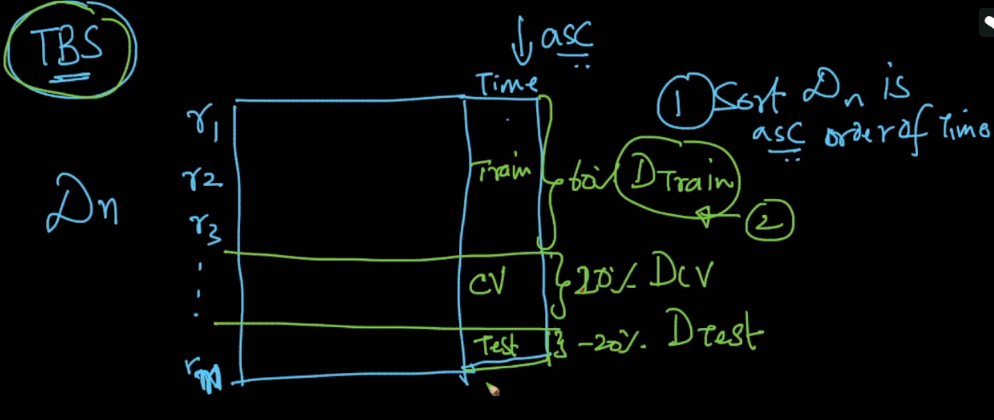
In this technique we split the data according to the time.





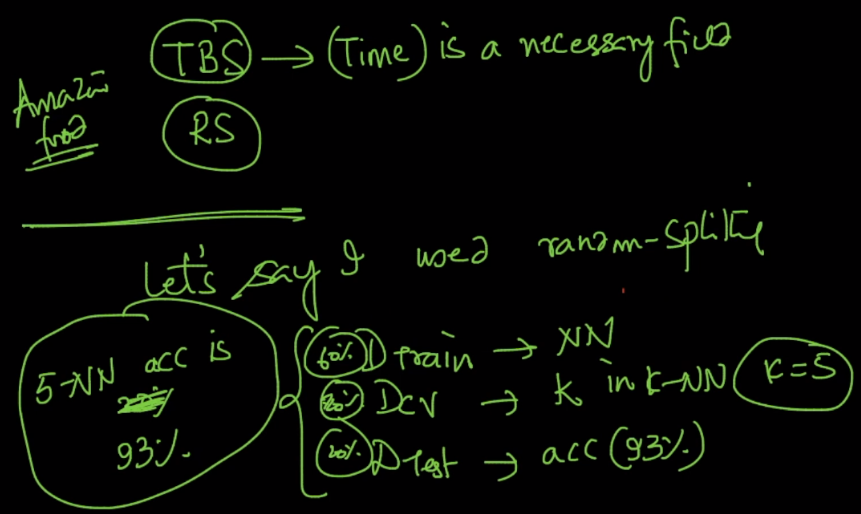
We sort data according to the time. Then we split data as:

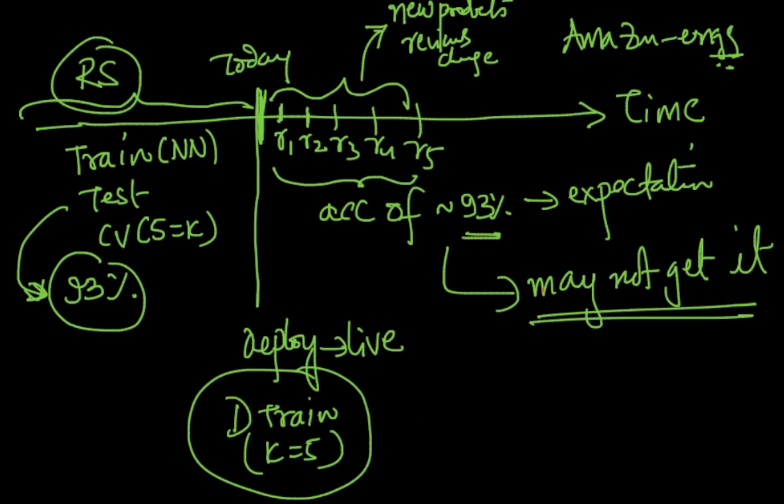
* Train: 60%
* CV: 20%
* Test: 20%

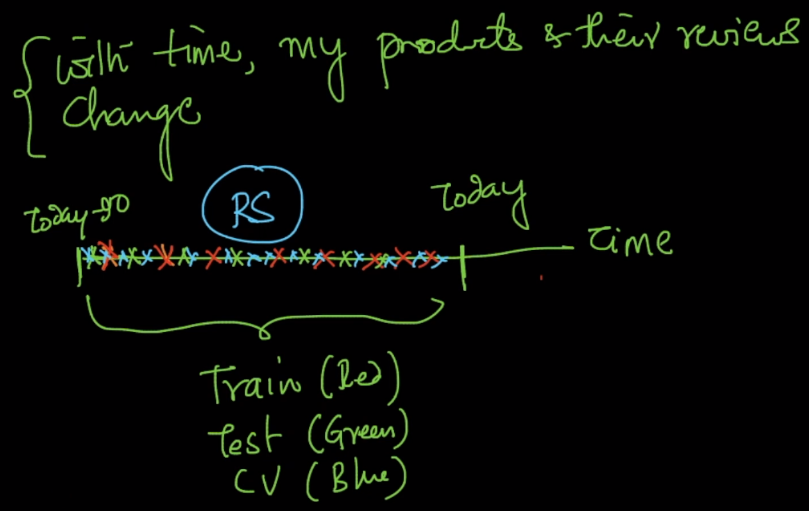


Let’s say we use random splitting for amazon fine food review. And gets 93% of accuracy so can we say that we would get similar accuracy for future data.

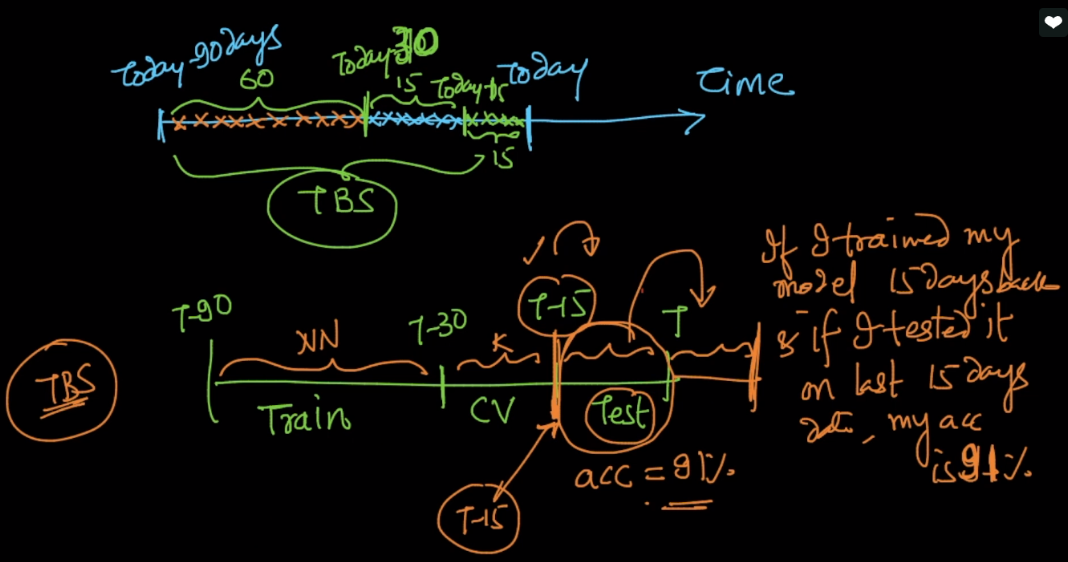
Answer is no, because we’ve trained tested on mixed timed data and we can’t say for future, because in future products change, review gets changed







Now we do time based splitting, so we train on oldest data(data before t-30), identified k using data before k-15 days and test accuracy using t-15 days, so now we can say that we’ll get 91% accuracy on future data.



**Where we should use time based splitting:**

Whenever time is available and if things/behavior of data changes over time then time based splitting should be used.

More details can be seen here: <https://discuss.analyticsvidhya.com/t/is-it-wise-to-split-training-and-test-dataset-based-on-time-year/2967/3>

