

Results by Sentences non-random data

Test Fold	Result
0	88.8
1	89.8
2	90.0
3	91.6
4	90.4
5	90.9
6	91.0
7	91.5
8	91.0
9	91.3

Results by Sentences random data

Test Fold	Result
0	92.4
1	92.5
2	92.3
3	92.3
4	92.4
5	92.3
6	92.2
7	92.4
8	92.2
9	92.3

Results by Fileld non-random data

Test Fold	Result
0	88.6
1	90.1
2	89.7

3	91.2
4	91.4
5	90.4
6	91.0
7	90.9
8	91.1
9	91.2

Results by Fileld random data

Test Fold	Result
0	91.4
1	91.1
2	91.1
3	90.9
4	90.7
5	90.6
6	90.6
7	90.5
8	91.0
9	90.8

Section 1:

There both have pretty close percentage, the random shuffle data is a little bit more accurate compare to non-random shuffle data. In performance wise, the data without shuffle produce the result slightly faster than the other.

Section 2:

Result of split by sentence is really close to split by source, random shuffle data has higher percentage compare to non-shuffle one, the only different.

Section 3:

Split by sentences data has higher accurate compare to split by source, it has higher percentage and it has more data that are over at least 90 percent.