Practical 8:
(for Lecture 8: Evaluation)

Threshold	TP	FN	FP	TN	Correct	Incorrect	test Set
1	20	80	2	98	100	100	200
5	50	50	5	95	100	100	200
10	60	40	10	90	100	100	200
15	80	20	20	80	100	100	200
20	88	12	30	70	100	100	200
25	90	10	40	60	100	100	200
30	95	5	50	50	100	100	200
35	96	4	60	40	100	100	200
40	97	3	70	30	100	100	200
50	98	2	80	20	100	100	200

1) I developed a system that detects tweets about about different candidates in a general election. I have classified a gold-standard set of 200 tweets, 100 of which I identified to be about the election and 100 I classed as being about non-election things.

   When I vary the similarity threshold of the system from 1 – 50 I get different numbers of correct and incorrect answers, that is different numbers of True Positives (TP), False Negative (FN), False Positives (FP) and True Negatives (TN) tweets. For example, when my system correctly identifies the tweets as being about the election and it was indeed about the election, it's a True Positive. When my system says that the tweet is about the election and it is not, then I have got a False Positive.

   Taking this data, can you compute the Precision and Recall for the system at each threshold and identify the threshold values at which it does best, according to the F1 measure?

2) Now, can you plot the ROC for this data?

3) Now, can you plot the DET curve for the same data?