

Data Science HW 2

1a	2
1b	2
2	not done
3a	2
3b	2
4	2
5	2
6	not done
7	2
8a	2
8b	2
9	2
10	2
11	2

1a. 0.202

1b. 0.798

2.

3a. The dropped element is d . It had a popularity score of 0.5.

3b. The most popular element is a and its score is 2.1.

4. 7 8 9 10 works. The hash values are:

7 6 0110

8 9 1001

9 1 0001

10 4 0100

So, the longest tail length is that of $h(10)$, which is 4, or 0100, and $2^2 = 4$.

5.

A \vee (B \wedge C)

		A	
1	/	\	0
1		B	
	1	/	\
		C	0
1	/	\	0
1		0	

6.

7.

Accuracy: 0.0950

Precision: 0.0871

Recall: 0.8500

F1 Score: 0.1580

8a. The accuracy is 0.99. This is a poor fraud detection system because we would rather have a classifier that produces some false positives but produces no false negatives than the opposite.

8b. Recall: 0. Precision: 0.

9. ROC

10. The tuple where a decision tree beats native bayes is Led7. 0.16

11. Always switch.