Practice Problem on Decision Trees: Happy Halloween!

In the U.S. and many countries around the world, children in costumes go from place to place in the evening of October 31st, asking for treats with the phrase "Trick or treat". The "treat" is usually candy, while the "trick" refers to a "trickery" performed to the homeowner or their property if no treat is given. We would like to predict whether a trick or a treat will be performed for different houses given the availability of candy (Candy) and the degree of "spookiness" of one's costume (Spooky). The data that we have collected from previous years are the following:

	Sample	Features		Outcome
		Candy	Spooky	Treat
Train	**S1**	No	Little	No
	S2	No	Very	No
	S3	Moderate	Little	Yes
	S4	Great	Little	Yes
	S5	Great	Little	Yes
	S6	Great	Very	No
	S7	Moderate	Very	Yes

Based on the above data, we will build a decision tree using the information entropy as a splitting criterion. The input features are **Candy** and **Spooky**, while the outcome variable is **Treat**.

(a) Compute the entropy splitting criterion of the outcome **Treat** conditioned on the **Candy** and **Spooky** features. Which feature will be used as the splitting attribute in root of the tree? Show all your calculations.
Note: You **do not** need to perform arithmetic calculations for logarithms, e.g. if one of your equations contains $\log(\frac{1}{3})$, you can leave it like that and still solve the problem.

(b) Create the decision tree using only one node, i.e., the tree will only have the root. Please show the **splitting criterion of the node**, as well as **the decisions from each possible outcome of the corresponding criterion**. Please **describe how the decisions were made**.

(c) Which of the training samples will be classified correctly only using the above tree and which not?

(d) Expand the tree until all samples are correctly classified.