

Classification Trees

2018-07-20

Classification Tree: Model 1

You have been given a set of index cards that give data on 30 passengers on the Titanic. Each card represents one passenger. You have also been given a *classification tree* (see Classification Tree: Model 1 on the next page) that provides rules for classifying whether or not each passenger survived. The tree is a visual model of a classification algorithm. We will call this classification algorithm *Model 1*.

1. For each passenger, use the data provided along with the classification algorithm to predict whether each of the 30 passengers *Died* or *Survived*. Fill in the classification prediction for each of the 30 passengers in Tables 1 and 2 (under column “Model 1”).
2. In the table where you put the predictions you can find out whether that passenger actually survived (under the “Fate” column). The classification algorithm correctly classified some of the passengers, and incorrectly classified others. You will use the classifications from Model 1 and the passengers’ actual fates to create a **classification table** for Model 1. To do this, fill in the table below by counting and recording the number of passengers in each cell of the table. Also compute and record in the table the percentage of the 30 passengers in each cell.

True Class	Predicted Class	
	Died	Survived
Died		
Survived		

Figure 1: Classification Table: Model 1

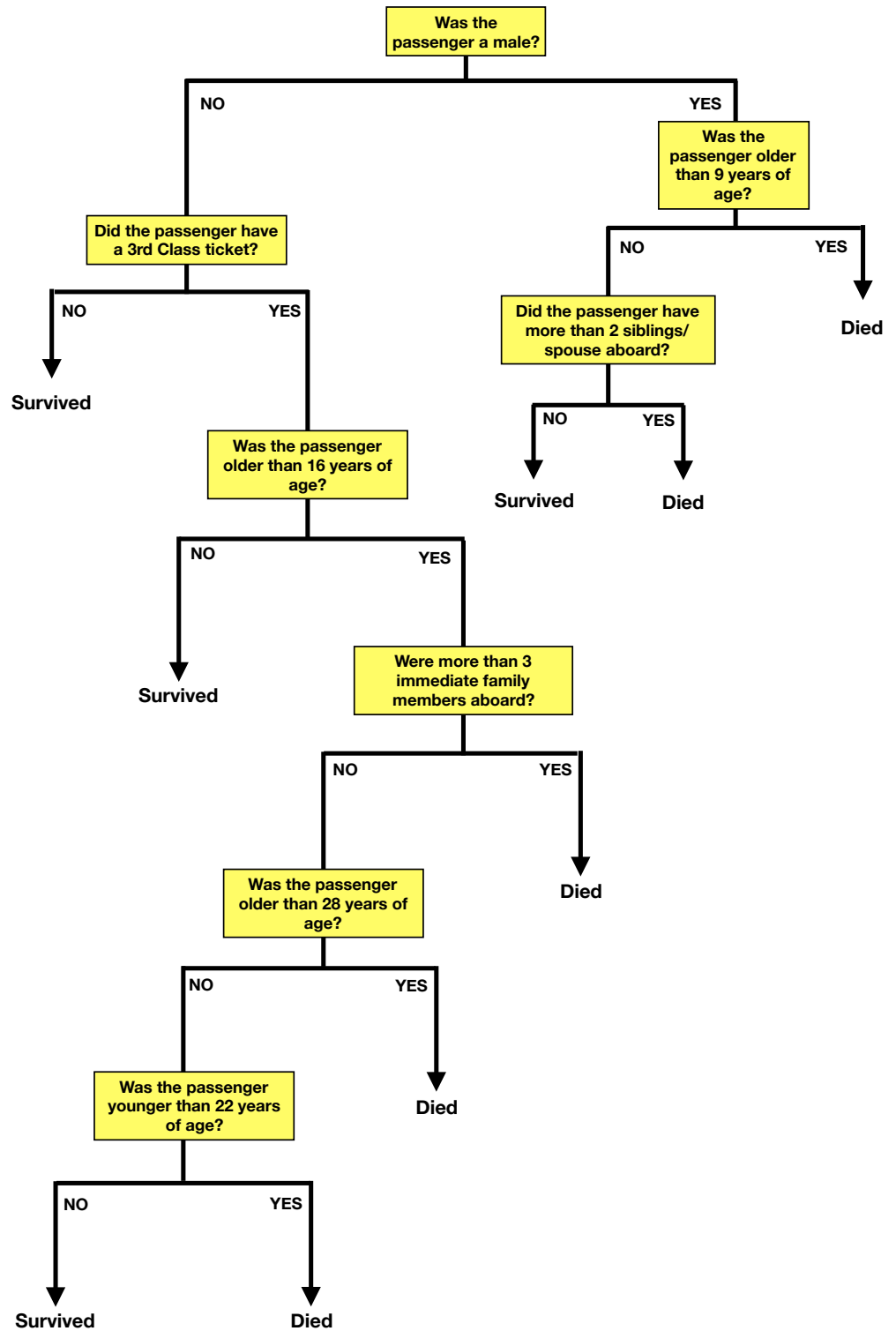


Figure 2: Classification Tree: Model 1

3. Use the information from your classification table to come up with a quantitative measure of the classification accuracy of the tree algorithm.

4. Based on your measure of classification accuracy, was the tree algorithm a “good” model? Explain.

Single Classification Models

5. Almost all of the passengers aboard the Titanic died. One algorithm we could use to classify passengers is one that has a single classification: Classify everyone as “Died” (Model 2). Fill in the “Model 2” column of Tables 1 and 2 based on this algorithm.
6. Fill in the classification table below for Model 2 based on this algorithm and the passengers’ actual fates. Again, compute the counts and percentages for each cell.

True Class	Predicted Class	
	Died	Survived
Died		
Survived		

Figure 3: Classification Table: Model 2

7. Create the classification tree for Model 2.

8. Model 2 is one of the two simplest classification algorithms we could create; after all there are no decisions and only a single classification. There is one other classification algorithm that also has a single classification. Write the algorithm for this model (Model 3) and fill in the classifications from this algorithm in the “Model 3” column of Tables 1 and 2. Also, compute the counts and percentages for each cell in the classification table for this algorithm.

True Class	Predicted Class	
	Died	Survived
Died		
Survived		

Figure 4: Classification Table: Model 3

9. Of the two single classification models, Model 2 and Model 3, which is the better classification model? Explain.

Baseline Model for Comparisons

The single classification model that has the better classification accuracy can be used as a baseline model to which we can compare the classification accuracy of other models.

10. Using the better single classification model, re-evaluate whether you think that the first model (Model 1) was “good”. Explain why or why not.

11. Quantify how much better (or worse) Model 1 is than your chosen baseline model.

Two Additional Classification Algorithms

12. You have been given the tree models for two additional classification algorithms (Models 4 and 5). For each,
 - a) Fill in the classification predictions in Tables 1 and 2;
 - b) Construct a classification table; and
 - c) Quantify how much better (or worse) the model is than your chosen baseline model.

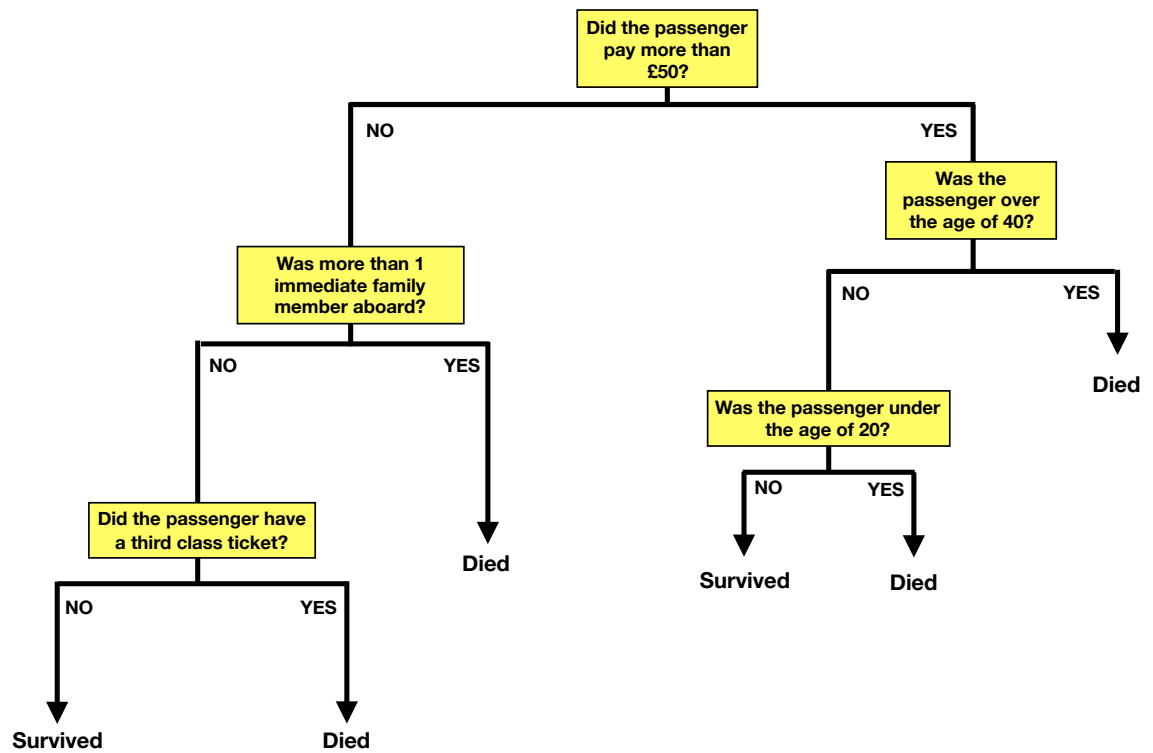


Figure 5: Classification Tree: Model 4

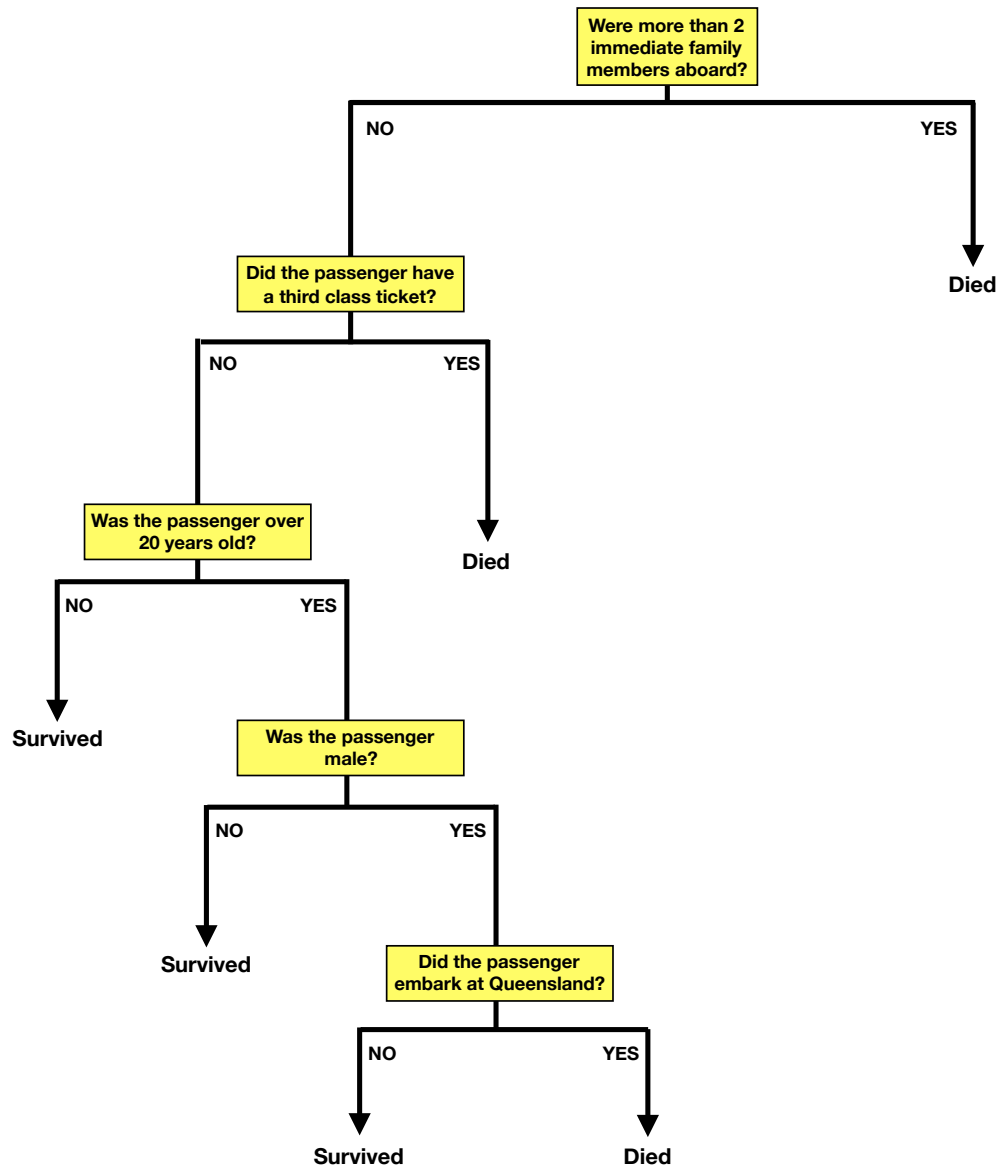


Figure 6: Classification Tree: Model 5

13. You have now looked at five different classification algorithms. Rank order the five algorithms from “worst” to “best”. Explain how you are making these rankings.

Table 1: Classifications and Actual Fate for Passengers 1–15 in the Test Data Set

Passenger	Model 1	Model 2	Model 3	Model 4	Model 5	Fate
Angle, Mrs. William A (Florence "Mary" Agnes Hughes)						Survived
Barbara, Miss. Saide						Died
Barber, Miss. Ellen "Nellie"						Survived
Beckwith, Mr. Richard Leonard						Survived
Brown, Mr. Thomas William Solomon						Died
Buss, Miss. Kate						Survived
Caldwell, Master. Alden Gates						Survived
Carbines, Mr. William						Died
Colley, Mr. Edward Pomeroy						Died
Crease, Mr. Ernest James						Died
Duff Gordon, Lady. (Lucille Christiana Sutherland) ("Mrs Morgan")						Survived
Francatelli, Miss. Laura Mabel						Survived
Gilinski, Mr. Eliezer						Died
Gustafsson, Mr. Anders Vilhelm						Died
Hakkarainen, Mrs. Pekka Pietari (Elin Matilda Dolck)						Survived

Table 2: Classifications and Actual Fate for the Passengers 16–30 in the Test Data Set

Passenger	Model 1	Model 2	Model 3	Model 4	Model 5	Fate
Hoyt, Mr. Frederick Maxfield						Survived
Lang, Mr. Fang						Survived
Larsson, Mr. August Viktor						Died
Lindblom, Miss. Augusta Charlotta						Died
Millet, Mr. Francis Davis						Died
Mitchell, Mr. Henry Michael						Died
Nakid, Mr. Sahid						Survived
Nye, Mrs. (Elizabeth Ramell)						Survived
Pickard, Mr. Berk (Berk Trembisky)						Survived
Roths, the Countess. of (Lucy Noel Martha Dyer-Edwards)						Survived
Salkjelsvik, Miss. Anna Kristine						Survived
Sawyer, Mr. Frederick Charles						Died
Sheerlinck, Mr. Jan Baptist						Survived
Skoog, Master. Harald						Died
Strandberg, Miss. Ida Sofia						Died