

Assignment Instructions: Assignment 3

Purpose

The purpose of this assignment is to use Naive Bayes for classification. The assignment requires you to:

- Use Naive Bayes classification to complete the task.
- Make predictions from your model.

Directions

Predicting flight delays can be useful to a variety of organizations. Here,

we look at five predictors. The outcome of interest is whether or not the flight is delayed (delayed here means arrived more than 15 minutes late). Our data consist of all flights from the Washington, DC area into the New York City area during January 2004. A record is a particular flight. The percentage of delayed flights among these 2201 flights is 19.5%. The data were obtained from the Bureau of Transportation Statistics (available on the web at www.transtats.bts.gov). The goal is to accurately predict whether or not a new flight (not in this dataset), will be delayed. The outcome variable is whether the flight was delayed, and thus it has two classes (1 = delayed and 0 = on time).

Day of Week	Coded as 1 = Monday, 2 = Tuesday, ..., 7 = Sunday
Sch. Dep. Time	Broken down into 18 intervals between 6:00 AM and 10:00 PM
Origin	Three airport codes: DCA (Reagan National), IAD (Dulles), BWI (Baltimore-Washington Int'l)
Destination	Three airport codes: JFK (Kennedy), LGA (LaGuardia), EWR (Newark)
Carrier	Eight airline codes: CO (Continental), DH (Atlantic Coast), DL (Delta), MQ (American Eagle), OH (Comair), RU (Continental Express), UA (United), and US (USAirways)

- Divide the data into 60% training and 40% validation
- Use the Naive Bayes model to predict whether the flight is delayed or not. Use only categorical variables for the predictor variables. Note that Week and Time variables need to be recoded as factors.
- Output both a counts table and a proportion table outlining how many and what proportion of flights were delayed and on-time at each of the three airports.
- Output the confusion matrix and ROC for the validation data

File Attached: FlightDelays.csv

Module Learning Outcomes

This assignment will help you with the following learning outcomes:

- Implement and optimize a Naive Bayes classification
- Use the constructed model to make predictions

Requirements

All due dates are included in the Assignment Schedule.

General Submission Instructions

All work must be your own. Copying other people's work or from the Internet is a form of plagiarism and will be prosecuted as such.

- Upload a pdf document to your git repository. Name your file Username_#.ext, where Username is your Kent State User ID (the part before @), and # is the Assignment number. In this case, 3.

Provide the link to your git repository in Blackboard Learn for the assignment.