Political Lean Classification

Chica Morrow and Bryce Samwel

11/11/2020

Introduction

We're gonna try to tell people whether they lean democratic or republican with a Naive Bayes classifier. We will look at sex, race, generation, and community type.

First, let's look at the posteriors from the Pew Research Center.

```
#Probabilities given sex

probRepGivenMale <- 0.51
probDemGivenMale <- 0.41
probIndGivenMale <- 0.08

probRepGivenFemale <- 0.38
probDemGivenFemale <- 0.54
probIndGivenFemale <- 0.08

#Probabilities given race/ethnicity

probRepGivenWhite <- 0.54
probDemGivenWhite <- 0.39
probIndGivenWhite <- 0.07

probRepGivenBlack <- 0.07
probDemGivenBlack <- 0.87
probIndGivenBlack <- 0.06

probRepGivenHispanic <- 0.27
probDemGivenHispanic <- 0.63
probIndGivenHispanic <- 0.10

probRepGivenAsian <- 0.27
probDemGivenAsian <- 0.66
probIndGivenAsian <- 0.07

#Probabilities given generation

probRepGivenMillenial <- 0.36
probDemGivenMillenial <- 0.57
probIndGivenMillenial <- 0.07
```

```
probRepGivenGenX <- 0.42
probDemGivenGenX <- 0.48
probIndGivenGenX <- 0.09

probRepGivenBoomer <- 0.49
probDemGivenBoomer <- 0.45
probIndGivenBoomer <- 0.06

probRepGivenSilent <- 0.53
probDemGivenSilent <- 0.40
probIndGivenSilent <- 0.07

#Probabilities given community type

probRepGivenUrban <- 0.33
probDemGivenUrban <- 0.60
probIndGivenUrban <- 0.07

probRepGivenSuburban <- 0.48
probDemGivenSuburban <- 0.44
probIndGivenSuburban <- 0.07

probRepGivenRural <- 0.55
probDemGivenRural <- 0.37
probIndGivenRural <- 0.08
```

Now, let's see the class priors.

```
#Class priors

probRep <- 0.44
probDem <- 0.48
probInd <- 0.08
```

And the predictor priors...

```
#Sex priors

probMale <- 0.492
probFemale <- 0.508

#Race/ethnicity priors

probWhite <- 0.62
probBlack <- 0.123
probHispanic <- 0.173
probAsian <- 0.052

#Generation Priors

probMillenial <- 0.197
```

```
probGenX <- 0.23
probBoomer <- 0.408
probSilent <- 0.145

#Community Type priors
probUrban <- 0.336
probSuburban <- 0.466
probRural <- 0.171
```

Now that we have all of these, we can used the naive bayes theorem to calculate each of the likelihoods which will be critical in making our predictions.

```
#Likelihoods given Rep

probMaleGivenRep <- probRepGivenMale*probMale/probRep
probFemaleGivenRep <- probRepGivenFemale*probFemale/probRep

probWhiteGivenRep <- probRepGivenWhite*probWhite/probRep
probBlackGivenRep <- probRepGivenBlack*probBlack/probRep
probHispanicGivenRep <- probRepGivenHispanic*probHispanic/probRep
probAsianGivenRep <- probRepGivenAsian*probAsian/probRep

probMillenialGivenRep <- probRepGivenMillenial*probMillenial/probRep
probGenXGivenRep <- probRepGivenGenX*probGenX/probRep
probBoomerGivenRep <- probRepGivenBoomer*probBoomer/probRep
probSilentGivenRep <- probRepGivenSilent*probSilent/probRep

probUrbanGivenRep <- probRepGivenUrban*probUrban/probRep
probSuburbanGivenRep <- probRepGivenSuburban*probSuburban/probRep
probRuralGivenRep <- probRepGivenRural*probRural/probRep

#Likelihoods given Dem

probMaleGivenDem <- probDemGivenMale*probMale/probDem
probFemaleGivenDem <- probDemGivenFemale*probFemale/probDem

probWhiteGivenDem <- probDemGivenWhite*probWhite/probDem
probBlackGivenDem <- probDemGivenBlack*probBlack/probDem
probHispanicGivenDem <- probDemGivenHispanic*probHispanic/probDem
probAsianGivenDem <- probDemGivenAsian*probAsian/probDem

probMillenialGivenDem <- probDemGivenMillenial*probMillenial/probDem
probGenXGivenDem <- probDemGivenGenX*probGenX/probDem
probBoomerGivenDem <- probDemGivenBoomer*probBoomer/probDem
probSilentGivenDem <- probDemGivenSilent*probSilent/probDem

probUrbanGivenDem <- probDemGivenUrban*probUrban/probDem
probSuburbanGivenDem <- probDemGivenSuburban*probSuburban/probDem
probRuralGivenDem <- probDemGivenRural*probRural/probDem

#Likelihoods given Ind
```

```
probMaleGivenInd <- probIndGivenMale*probMale/probInd
probFemaleGivenInd <- probIndGivenFemale*probFemale/probInd

probWhiteGivenInd <- probIndGivenWhite*probWhite/probInd
probBlackGivenInd <- probIndGivenBlack*probBlack/probInd
probHispanicGivenInd <- probIndGivenHispanic*probHispanic/probInd
probAsianGivenInd <- probIndGivenAsian*probAsian/probInd

probMillenialGivenInd <- probIndGivenMillenial*probMillenial/probInd
probGenXGivenInd <- probIndGivenGenX*probGenX/probInd
probBoomerGivenInd <- probIndGivenBoomer*probBoomer/probInd
probSilentGivenInd <- probIndGivenSilent*probSilent/probInd

probUrbanGivenInd <- probIndGivenUrban*probUrban/probInd
probSuburbanGivenInd <- probIndGivenSuburban*probSuburban/probInd
probRuralGivenInd <- probIndGivenRural*probRural/probInd
```

So, if we come up with an example person, we can see which class (political leaning) they are most likely to fall into! Let's suppose that we are interested in knowing a black woman's leaning. It's as simple as multiplying the prior for a given class by the likelihoods for that person given that class. The class with the highest outcome is our prediction! Let's do the calculations.

```
#Republican

probRep*probBlackGivenRep*probFemaleGivenRep
```

```
## [1] 0.003777442
```

```
#Democrat

probDem*probBlackGivenDem*probFemaleGivenDem
```

```
## [1] 0.06115621
```

```
#Independent

probInd*probBlackGivenInd*probFemaleGivenInd
```

```
## [1] 0.00374904
```

As we can see, the value for leaning democratic is higher than all of the other values, so we would predict that a black woman leans democratic based on this simple model! Now that might have been rather obvious, but what if we test for something else.

```
#Republican

probRep*probWhiteGivenRep*probMaleGivenRep*probGenXGivenRep*probSuburbanGivenRep
```

```
## [1] 0.02130919
```

```
#Democrat

probDem*probWhiteGivenDem*probMaleGivenDem*probGenXGivenDem*probSuburbanGivenDem
```

```
## [1] 0.009983647
```

```
#Independent

probInd*probWhiteGivenInd*probMaleGivenInd*probGenXGivenInd*probSuburbanGivenInd
```

```
## [1] 0.002252834
```