SYS 6018-001

Kaggle: Twitter Analysis Regarding Self-Driving Cars

Competition 3-1 s

Elizabeth Homan

Ben Greenawald

Leelakrishna (Sai) Bollempalli

Reflection:

* What is the problem?
  + We want to know what people think about SDCs
* Who might care about this problem and why?
  + SDC developers would certainly care because current sentiment about SDC’s could influence new features they add to the vehicles or could change how they advertise these vehicles
  + Politicians would care because as they are crafting legislation for these cars, they want to ensure that their constituents are happy
  + Potential future SDC owners because they want to make sure the buzz is positive around these cars before spending an inordinate amount of money on them.
  + Addressing concerns via media campaigns, tapping into positive hype, creating legislation to either restrict or allow SDCs, “keeping up with the Joneses,” discovering problems or advantages that potential owners might not otherwise have considered
* Why might this problem be challenging?
  + It’s not just about referencing something, but about sentiment, which is hard to tease out
  + Text problems in general are difficult because it is hard to decide what features to extract from the data
  + The same words oscillate fluidly between positive and negative connotations based on which other words they’re paired with (i.e. incredible vs incredibly dangerous vs incredibly safe vs less safe vs less dangerous vs not incredibly dangerous)
  + Twitter language is particularly messy because people are squeezing their thoughts into a small number of words
  + People’s understanding of these cars is relatively limited, and so the reactions to them are limited/primitive
  + Only a certain demographic uses Twitter, and so it is very possibly a biased sampling that might skew our understanding of the larger “problem”
  + The data does not have an even distribution of classes (there are 600 3’s and about 20 1’s)
* What other problems resemble this problem?
  + Any other sentiment analysis using twitter feeds or even other social media feeds