

# CS 838 - Project Stage 2

Jale Dinler, Riccardo Mutschlechner, Steven Lamphear

dinler@wisc.edu, riccardo@cs.wisc.edu, slamphear@wisc.edu

March 2, 2017

## 1 Questions to Answer

The entity type that we decided to go with for this portion of the project, is person names. Since we are using music reviews, this was a bit tricky as we had to exclude stage names, band names, song names, and album names. Thus, we ran into some issues which we will outline below. To tag them, we simply wrapped the name in angle brackets:  $\langle name \rangle$  and  $\langle /name \rangle$  tags. For example:  $\langle name \rangle$ Joan Jett  $\langle /name \rangle$

Our total number of mentions is 2715, across 200 reviews. In our Train/Dev set (of size 132 documents), I, we have 1932 mentions. In our Test set (of size 66 documents), we have 781 mentions.

Our best classifier originally, M, is Decision Tree. Our precision is 78.86 percent, our recall is 96.03 percent, and our f1 is 86.6 percent.

After doing some debugging on our development set, we improved the precision of M (now X), to be 91.04 percent; the recall to be 96.28 percent, and the f1 value to be 93.59 percent. We realized we had a lot of badly tagged data, so we had to spend a lot of time going back and fixing it.