Jeffrey Levesque

jlevesqu@syr.edu | Topic Modeling

IST-736: HW8

PRofessor gates

A drawing of a cartoon character

Description automatically generated

Table of Contents

[Introduction 2](#_Toc9955148)

[Analysis 3](#_Toc9955149)

[Data Preparation 5](#_Toc9955150)

[Amazon Mechanical Turk 6](#_Toc9955151)

[Stop Words 7](#_Toc9955152)

[Results 7](#_Toc9955153)

[Unigram Models 7](#_Toc9955154)

[Unigram and Bigram Models 9](#_Toc9955155)

[Unigram (100% train) 11](#_Toc9955156)

[Conclusions 12](#_Toc9955157)

# Introduction

# Analysis

# Data Preparation

## Stop Words

A rich set of stop words was created[[1]](#footnote-1), utilized prior to vectorization:

* http
* https
* nhttps
* rt
* amp
* co
* think
* twitter
* ye
* el
* la
* lo
* en
* es
* est
* deo
* se
* s8
* que
* para
* guion
* hecho
* trono
* ser
* year
* today
* time
* thank
* make
* people
* going
* say
* like
* new
* need
* know
* chairman
* list
* act
* years
* house
* congress
* speaker
* madam
* colleagues
* gentleman
* let
* said
* way
* want
* support
* state
* yield
* time
* good
* right
* urge
* important
* come
* rise
* dont
* members
* help
* country
* american
* legislation
* president
* united
* states
* mr
* hr
* thats
* things
* day
* use
* ensure
* look
* got
* fact
* really
* thing
* percent
* trying
* tell
* im
* "im'",
* sure
* saying
* understand
* floor
* actually
* came
* says
* talking
* believe
* didnt
* congressman
* great
* lets
* heard
* rule
* amendment
* committee
* went
* ago
* conference
* section
* provision
* able
* yes
* republican
* republicans
* democrat
* democrats
* vote
* majority
* democratic
* north
* south
* bipartisan
* washington
* carolina
* debate
* question
* talk
* friend
* district
* issue
* certainly
* appreciate
* making
* middle
* including
* res
* program
* doesnt
* ought
* simply
* talked
* weve
* wanted
* order
* party
* place
* body
* coming
* kind
* point
* little
* lot
* aisle
* maybe
* ms
* seen
* hear
* getting
* theyre
* provide
* area

# Results

## Latent Dirichlet Allocation

## Non-negative Matrix factorization

## Overall Sentiment

## Overall Topic Model

# Conclusions

1. <https://github.com/jeff1evesque/ist-736-hw/blob/master/hw8/utility/stopwords.py> [↑](#footnote-ref-1)