

Jake Schinto

## STEP 1

## 1. Bag of Words

[illegible]

## 2. Term Freq

	L	a	b	L	i	t	r	g	b	i	a	a	m	t	i	f	P	L	t	k	f	b	r	w	n	s	w	m	e	a	t	p	e	a	s
	e	l	e	a	n	h	i	a	e	s	w	t	o	h	n	o	a	e	e	o	r	e	i	h	i	u	r	o	v	a	h	a	x		c
	h	w	a	f			v	m	c		f		s	i	c	t	h	a	k	f	a	v	a	c	t	p	i	s	t	r	y	n	t	e	
	g	y	t			e	a	e	a		u			n	l	b	'	m	w		t	n	h		r	s	g	l	i	n	s	t	c	s	e
	h	s	s			r	y	s	e		l			g	l	a	s		n		g	r			g			g							
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	1	0	0	1	1	0	0	0	1	2	0	0	0	0	0	1	0	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1	1	1

### STEP 2

1. Lehigh always beat Laf in the rival game because Laf is awful at most thing includ football pathetic
2. Lehigh football team is known for beat Laf in rival which is not surpris because Lehigh win most everything against that pathetic excuse for a school.

### STEP 3

## 1. Bag of Words

[illegible]

## 2. Term Freq

	L e h i g h	a l w a y s	b e a t	L a f	i n	t h e	r i v a l	g a m e	b e c a u s e	i s	a w f u l	a t	m o s t	t h i n g	i n c l u d	f o o t b a l l	p a t h e t i c	t e a m	k n o w n	f o r	w h i c h	n o t	s u r p r i s	w i n	m o s t	e v e r y t h i n g	a g a i n s t	t h a t	e x c u s e	a	s c h o o l	
1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
2	2	0	1	1	1	0	1	0	1	2	0	0	0	0	0	1	1	1	1	2	1	1	1	1	1	1	1	1	1	1	1	1

### STEP 4

The table became shorter because more words started being aggregated together instead of being counted as their own things. Preprocessing might be helpful to remove clutter and find an actual representation of the key words used without being fooled by different variants of the same word.