

FUTURAMA

Bender's NLP

Isaac Kim

San Francisco Metis Fellow

APRIL 22, 2013 10:10am PT by Lesley Goldberg

'Futurama' Canceled for Third Time

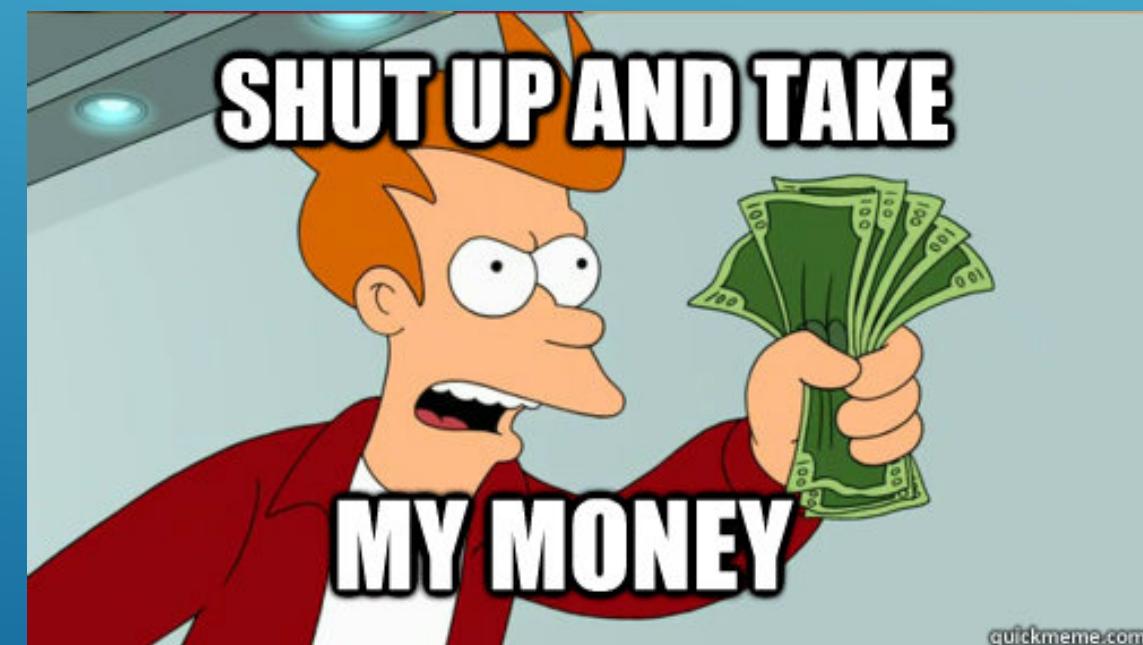
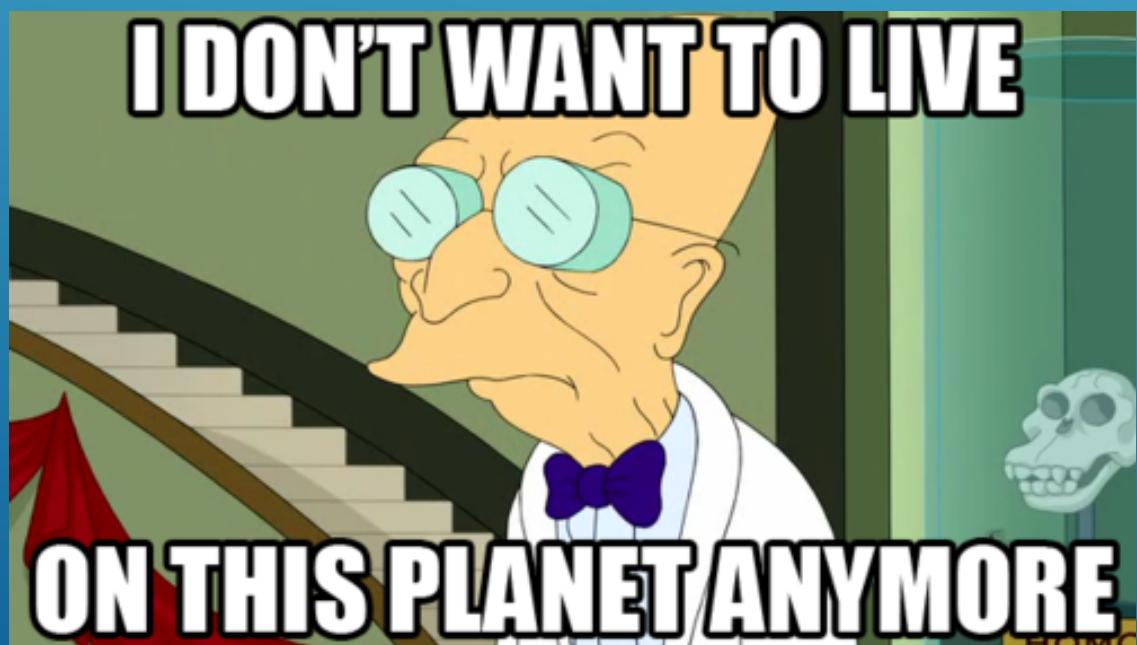
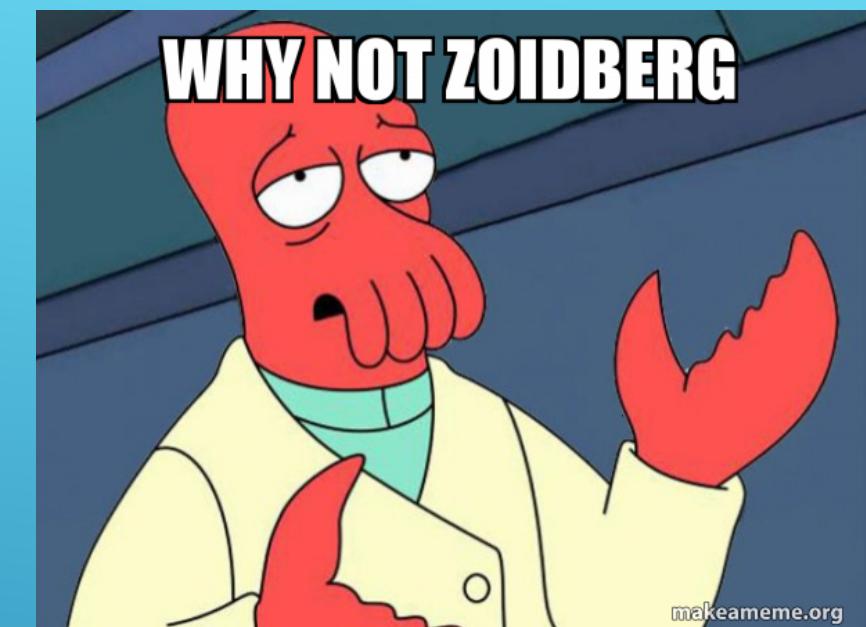
 Posted by u/Huntermbadley 3 years ago 

7.5k **ELI5: How did futurama win 6 emmys but got canceled twice?**

Explained

 1.9k Comments  Share  Save  Hide  Report





OBJECTIVES

►TOPIC MODELING

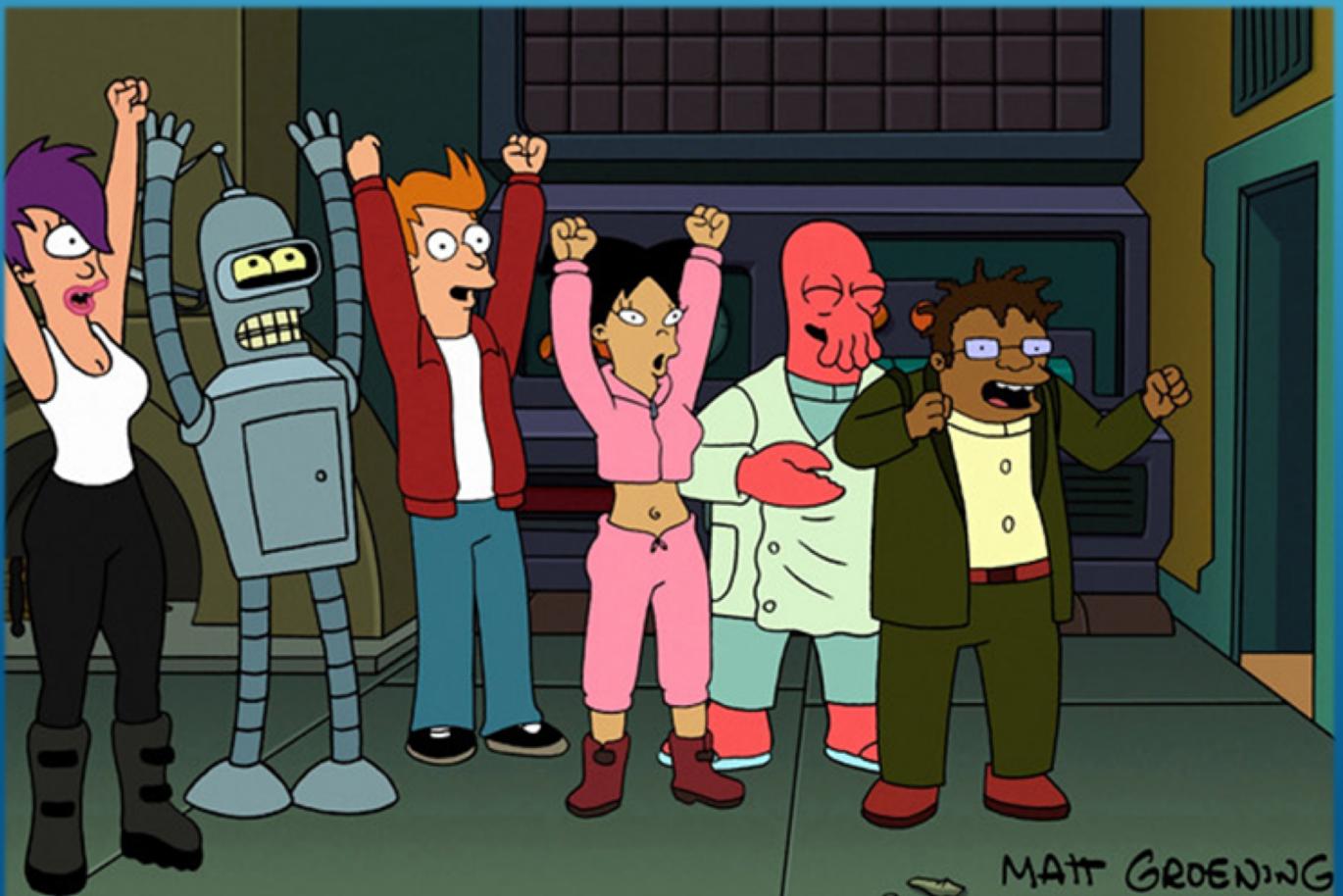
►CHARACTER
IDENTIFICATION

►TEXT GENERATION



Name	Counts
Fry	3926
Bender	3670
Leela	3116
Farnsworth	1519
Zoidberg	975
Amy	910
Hermes	843
Zapp	500
Kif	253
Mom	185

EXPLORATORY DATA ANALYSIS

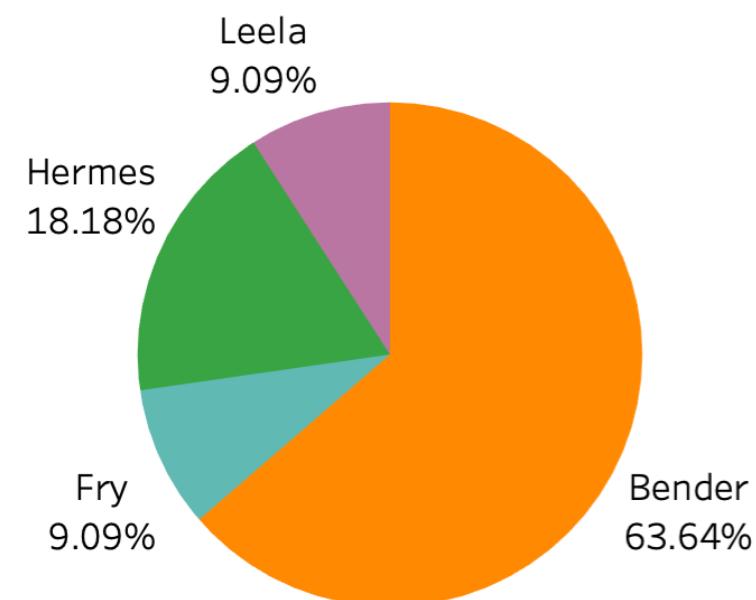


EXPLORATORY DATA ANALYSIS

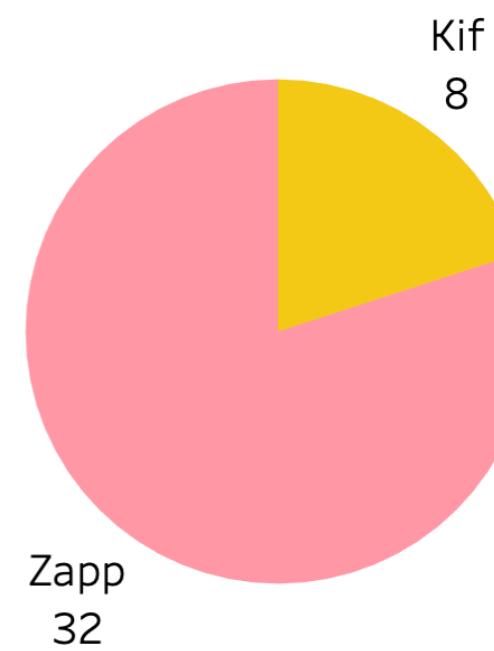
Names	Unigram1	Unigram 2	Bigram 1	Bigram 2
Fry	seymour	future	wanna hear	wait minute
Bender	elzar	ass	shiny metal	metal ass
Leela	surface	yah	save animal	dark matter
Farnsworth	news	wha	eh wha	good news
Zoidberg	hooray	claw	hermes friend	woop woop
Amy	buh	kif	fun feel	uh mom
Hermes	labarbara	mon	bird dog	mon hell
Zapp	neutral	kif	kif woman	sham pag
Kif	rustler	sir	um sir	um um



Utterances of the Word 'Ass' By Character

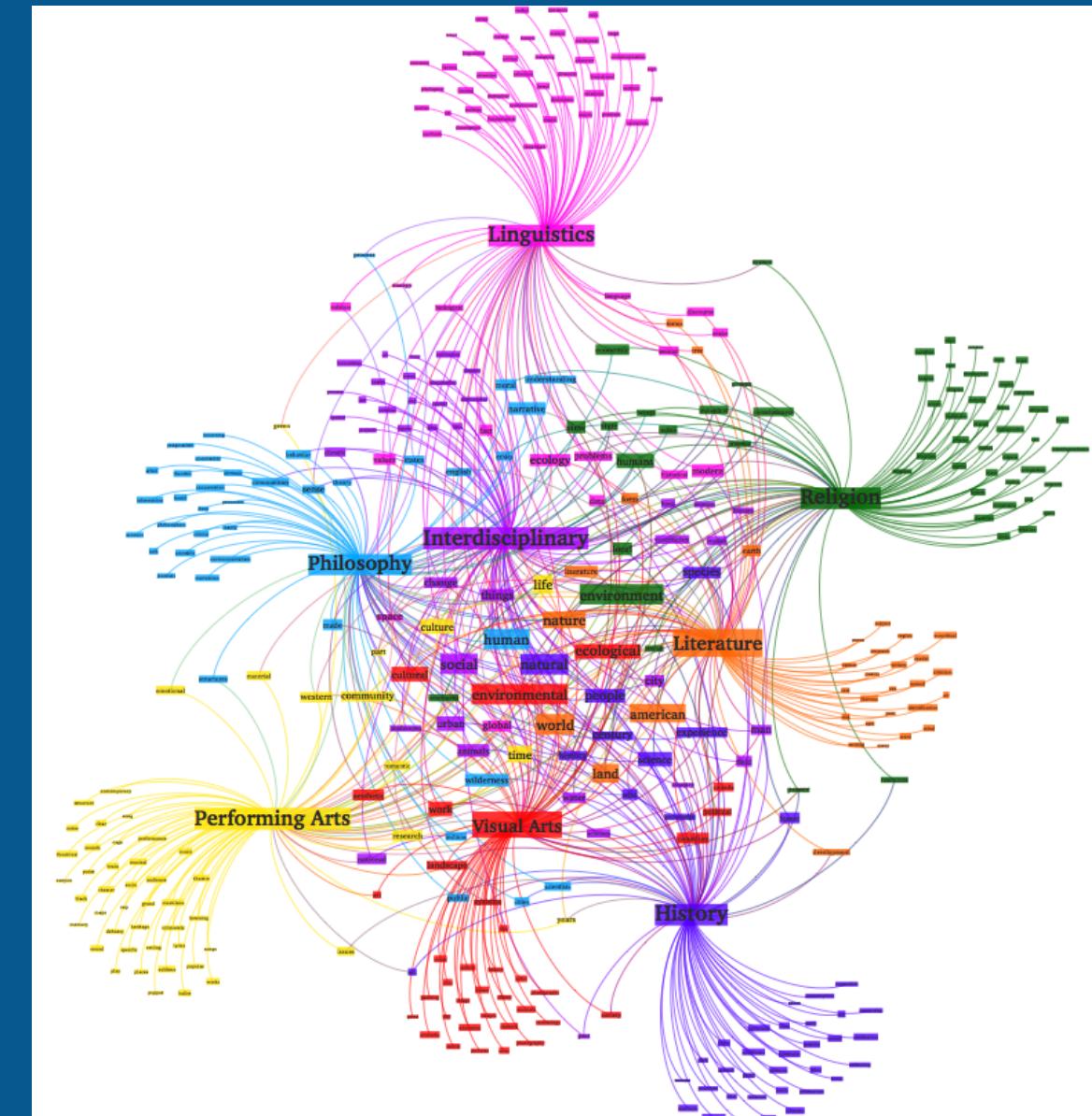


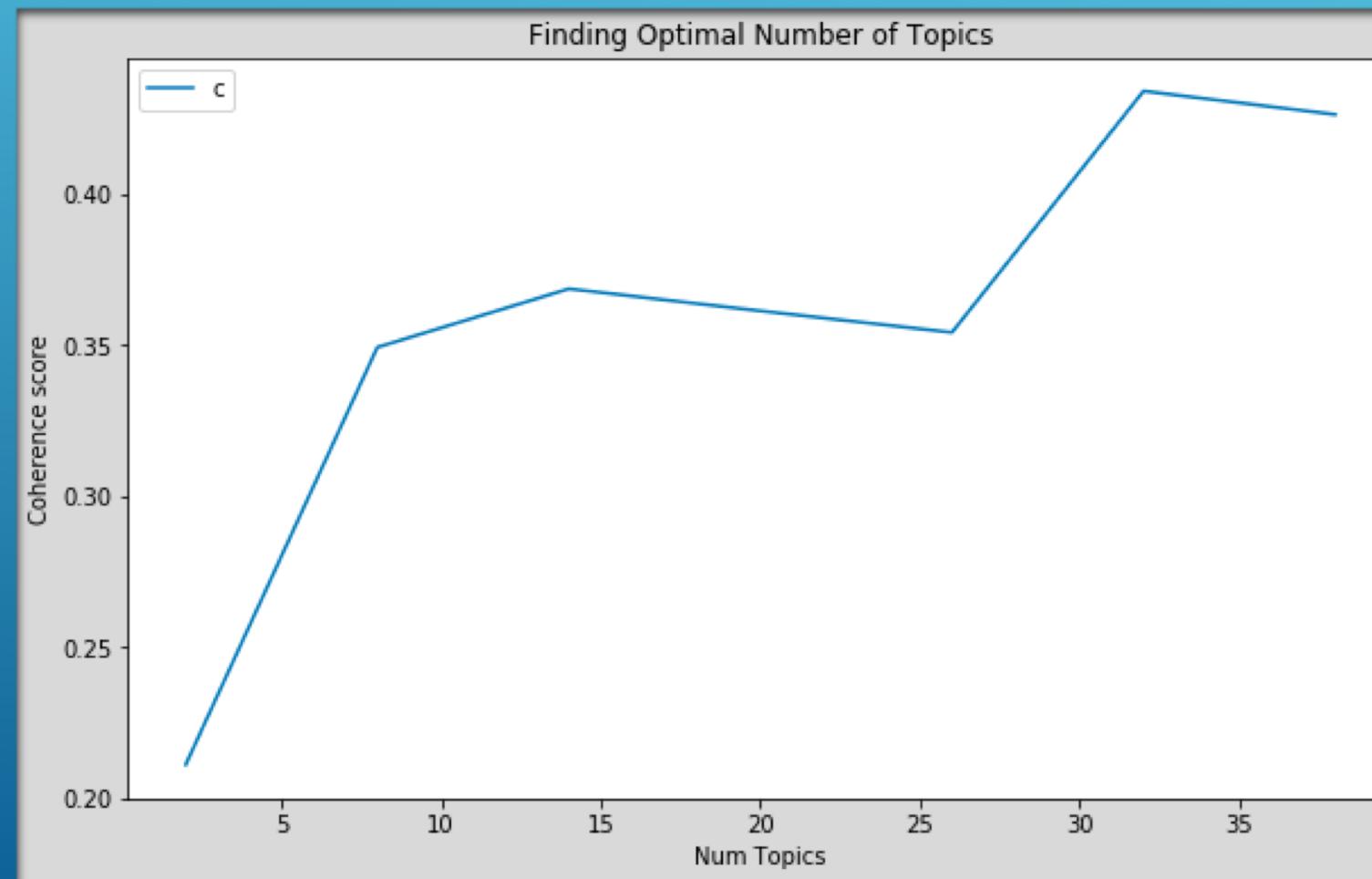
'Velour' Count by Character



TOPIC MODELING

spaCy





Topics:

1. Friends
2. Actions
3. Planet Express
4. Hermes
5. Leela
6. Amy

LDA COHERENCE SCORES

Out[54]:

Selected Topic: 2

Previous Topic

Next Topic

Clear Topic

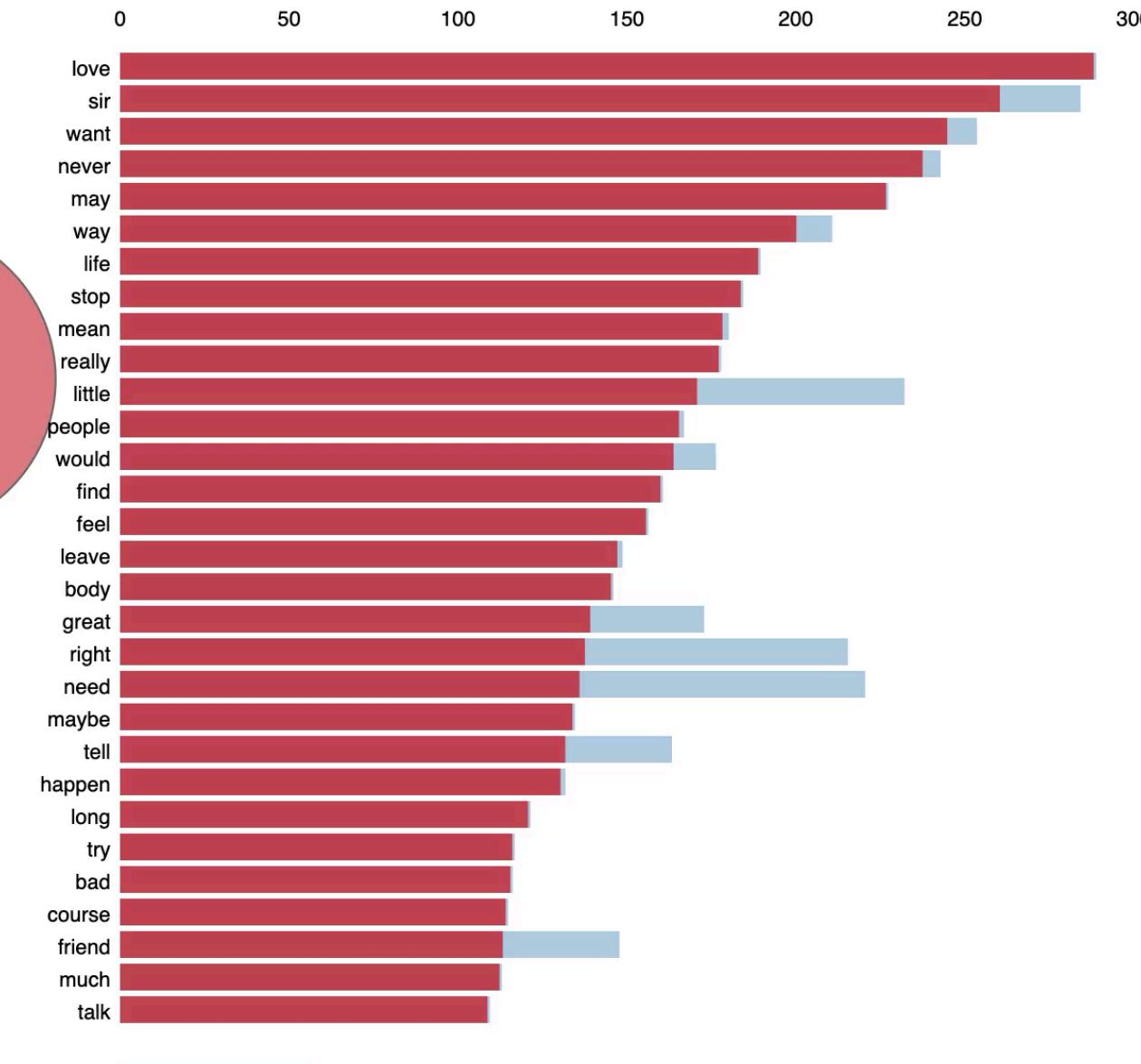
Slide to adjust relevance metric:(2)

 $\lambda = 1$

Intertopic Distance Map (via multidimensional scaling)



Top-30 Most Relevant Terms for Topic 1 (28.9% of tokens)



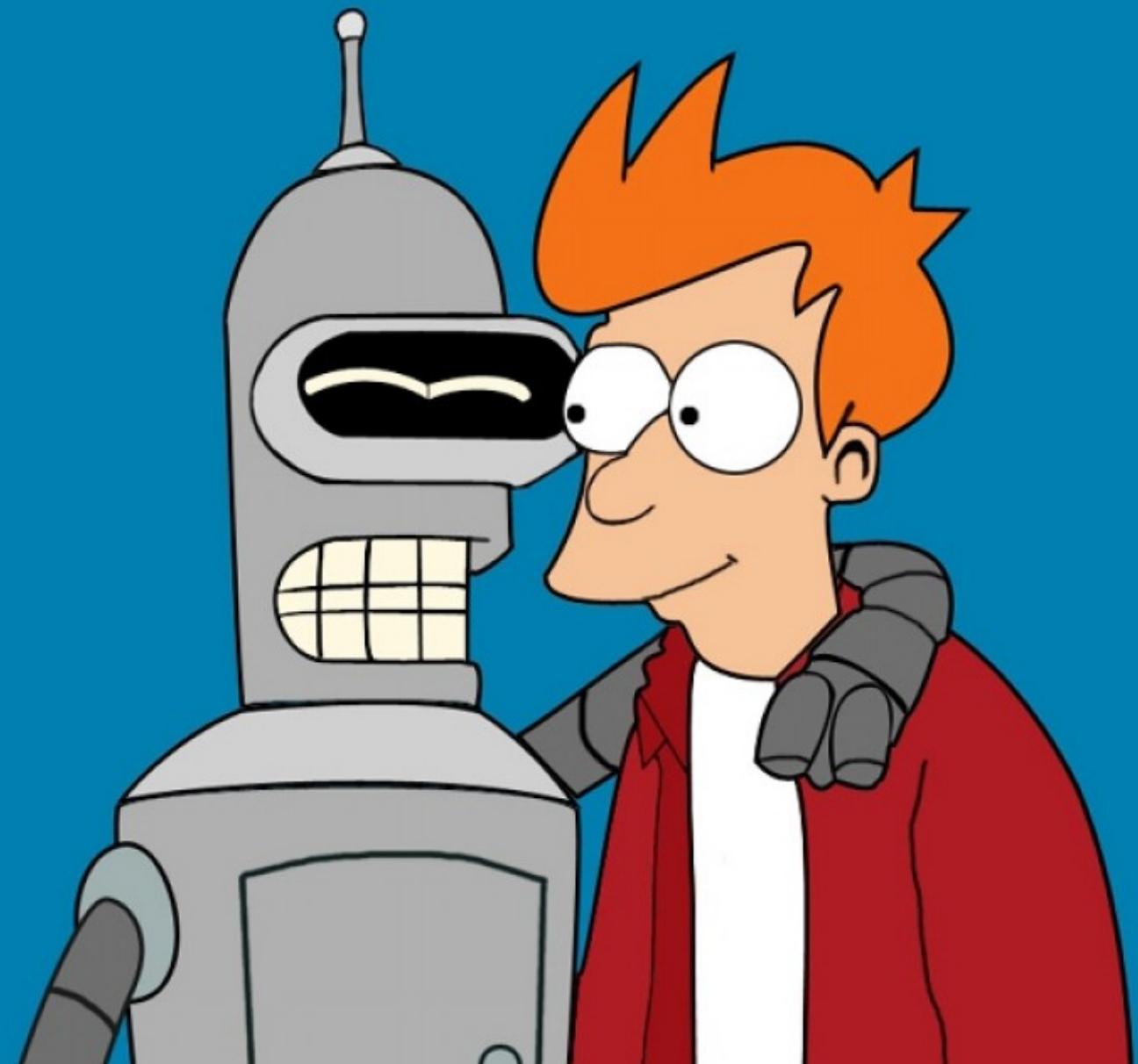
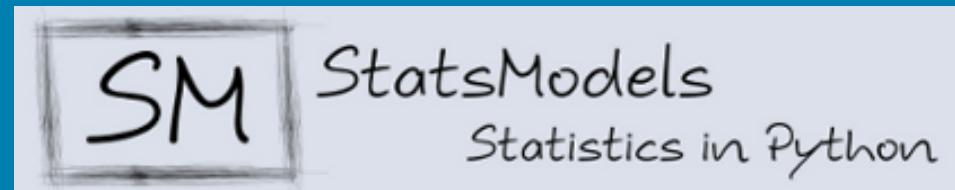
Overall term frequency

Estimated term frequency within the selected topic

1. saliency(term w) = frequency(w) * [sum_t p(t | w) * log(p(t | w)/p(t))] for topics t; see Chuang et. al (2012)

2. relevance(term w | topic t) = $\lambda * p(w | t) + (1 - \lambda) * p(w | t) / p(w)$; see Sievert & Shirley (2014)

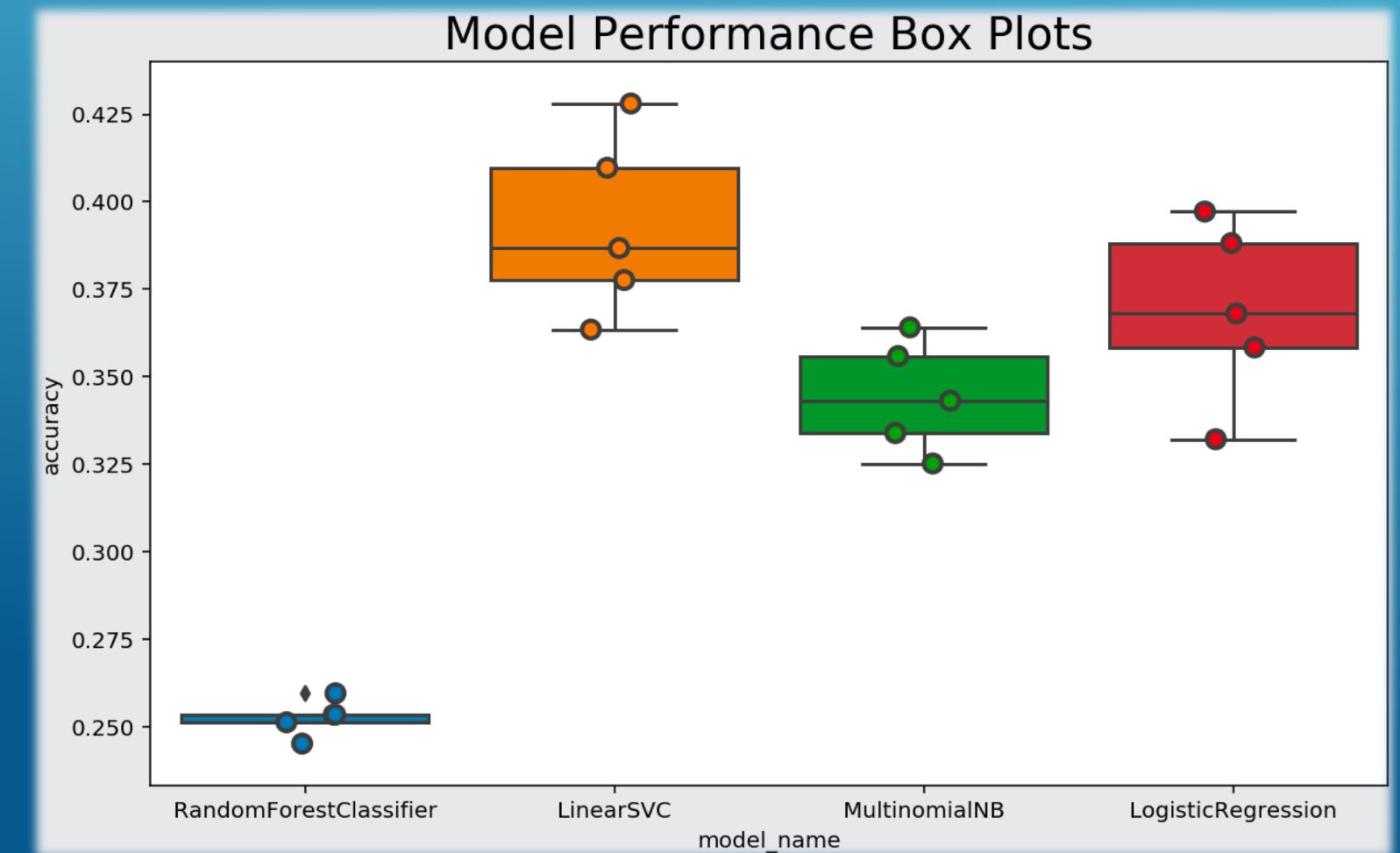
CHARACTER IDENTIFICATION



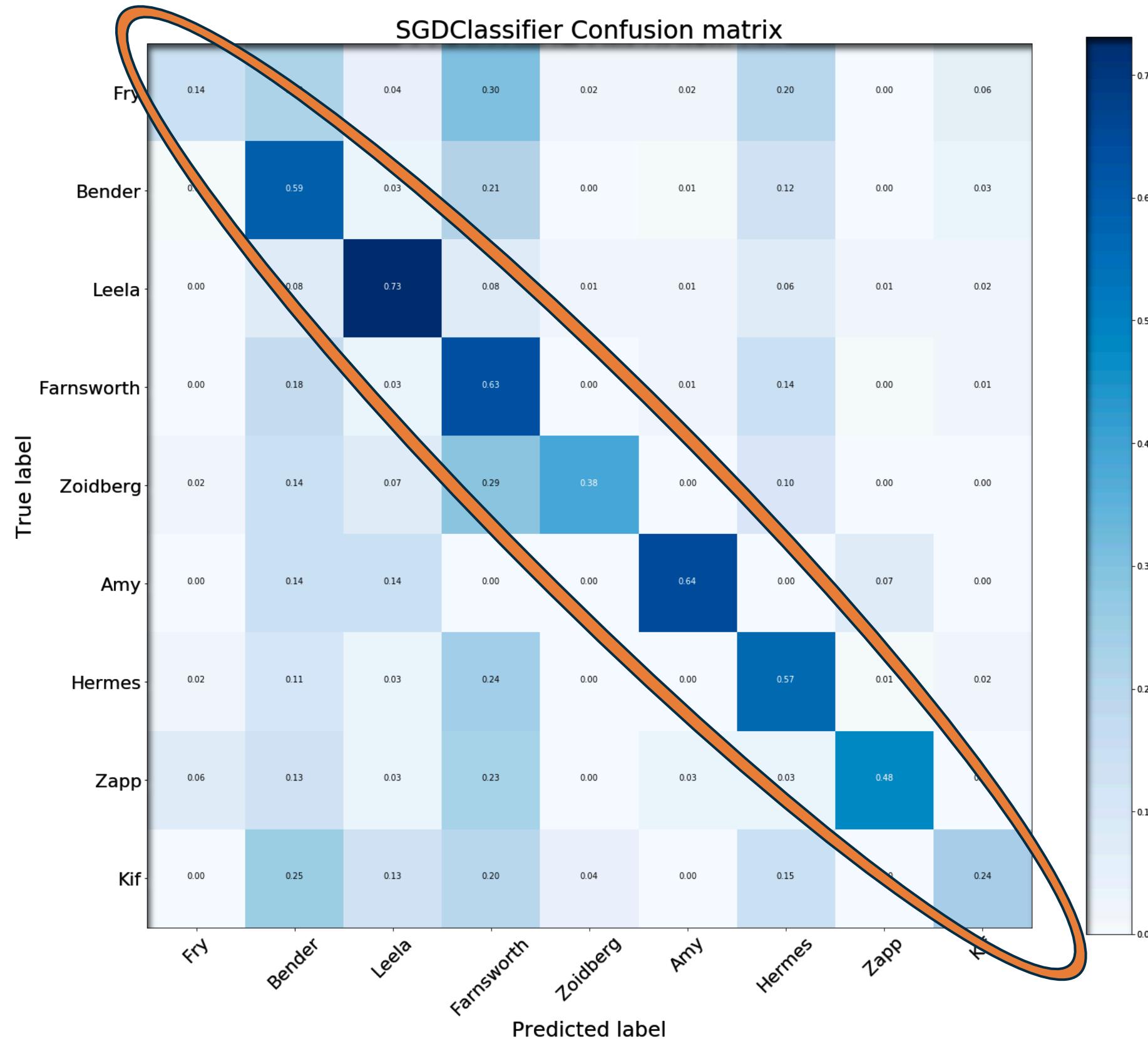
CHARACTER IDENTIFICATION

Classification:

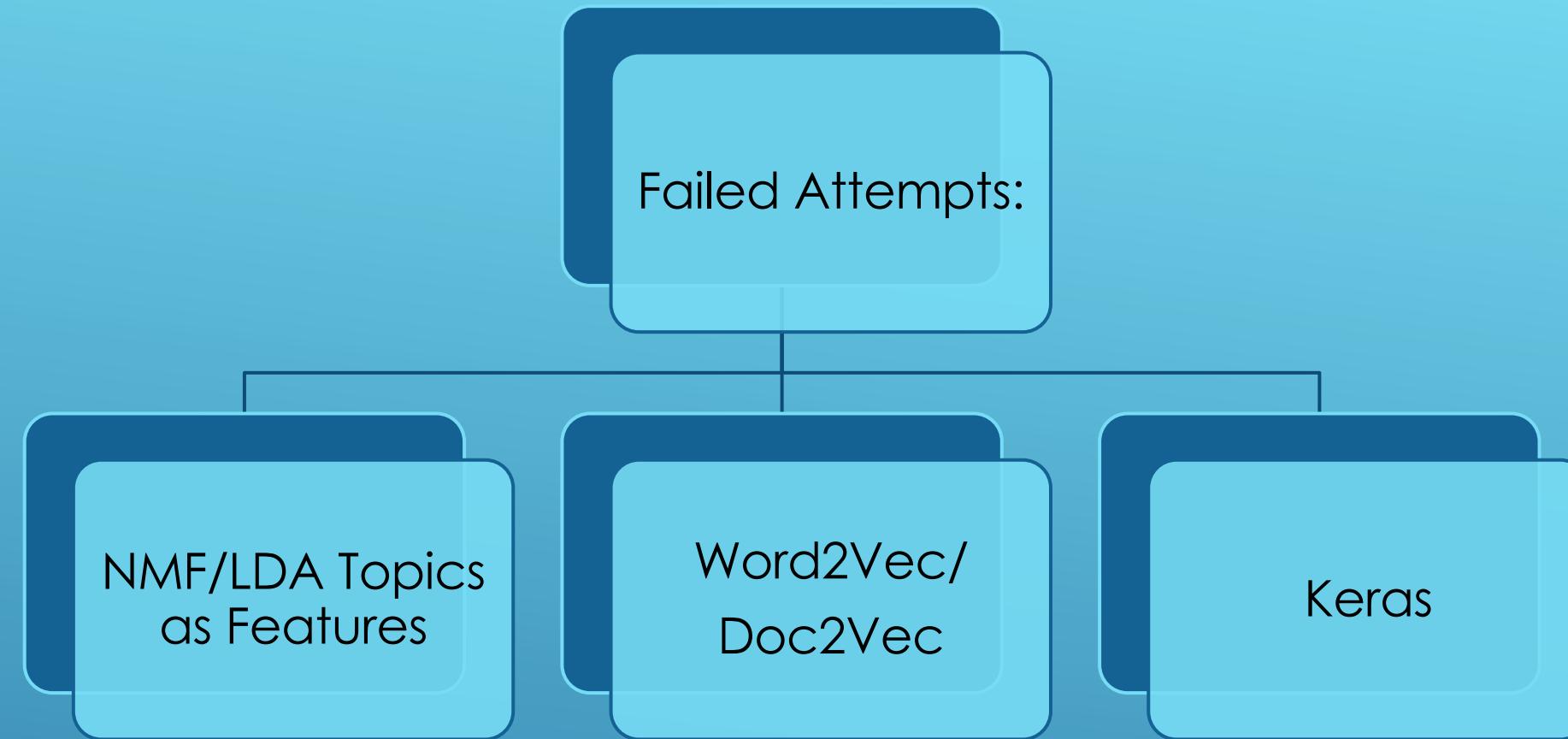
- TfidfVectorizer
- χ^2 Feature Selection
- SGDClassifier
 - 55% Accuracy

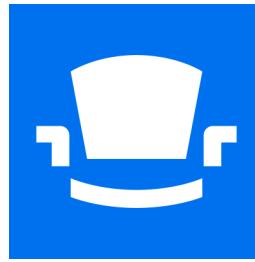


SGDClassifier Confusion matrix



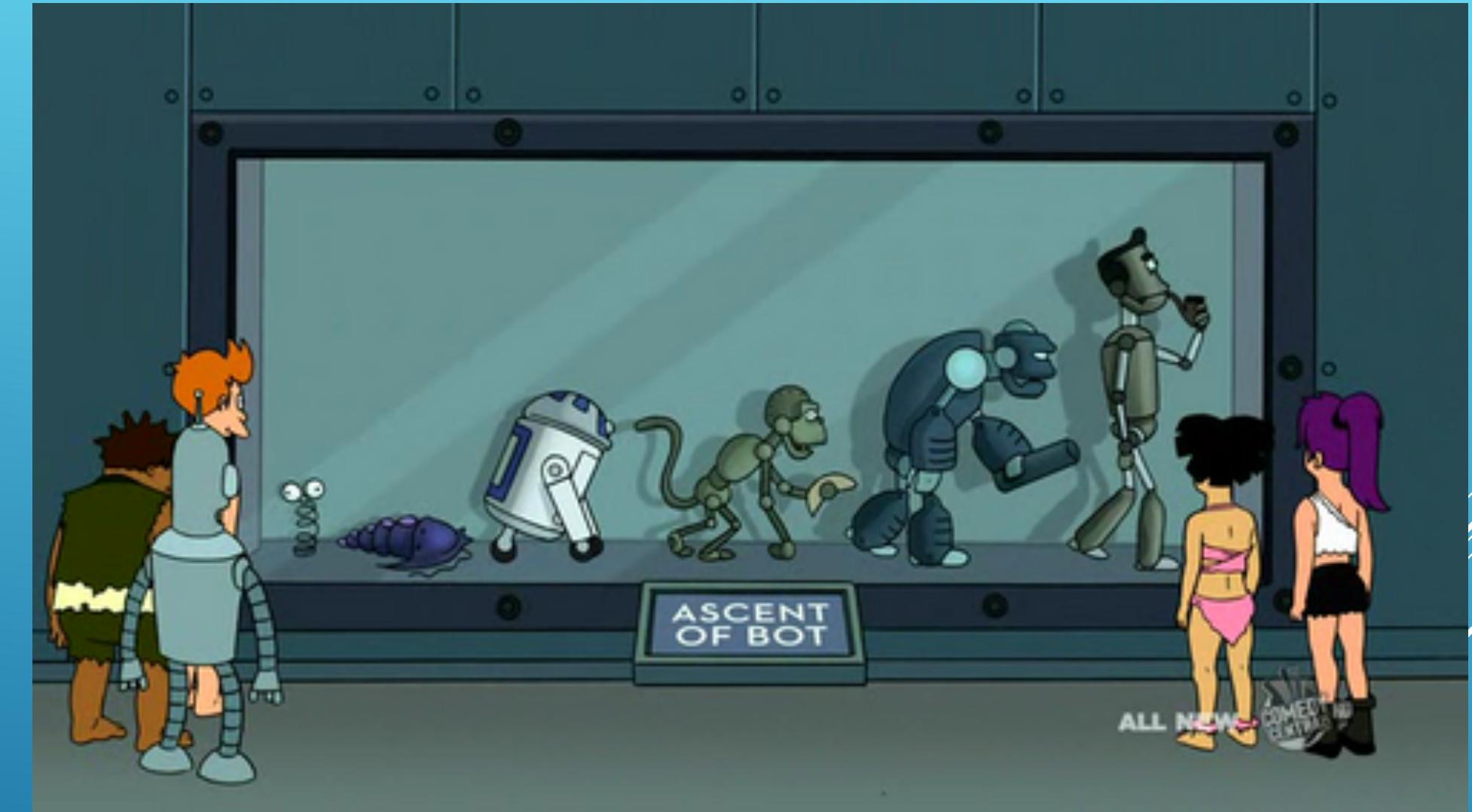
OTHER ATTEMPTS



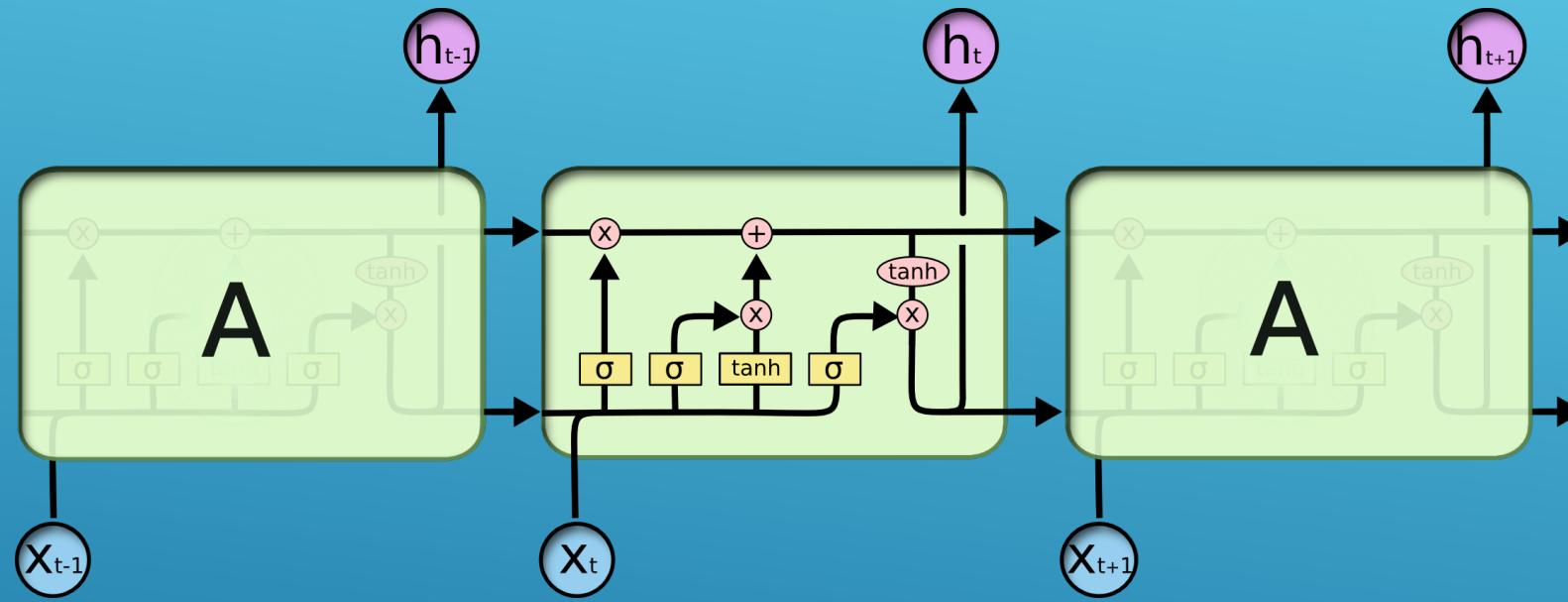


K Keras

T TensorFlow



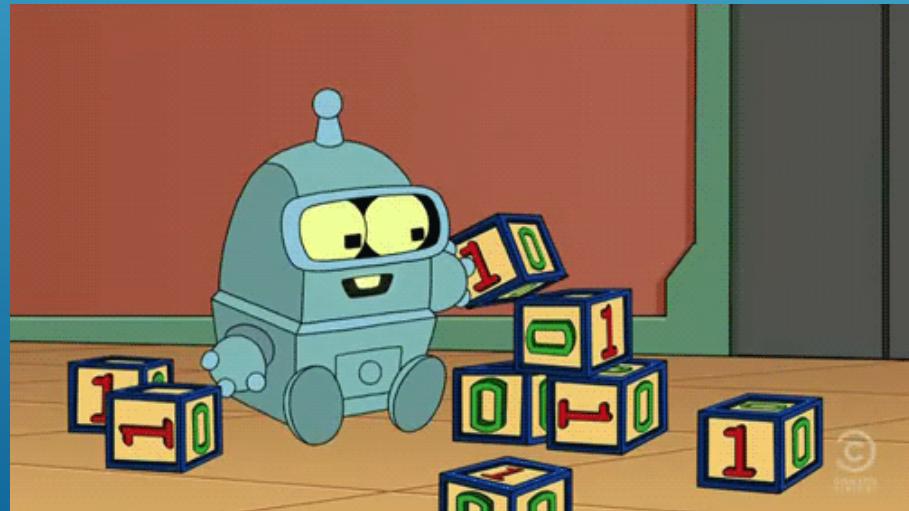
TEXT GENERATION



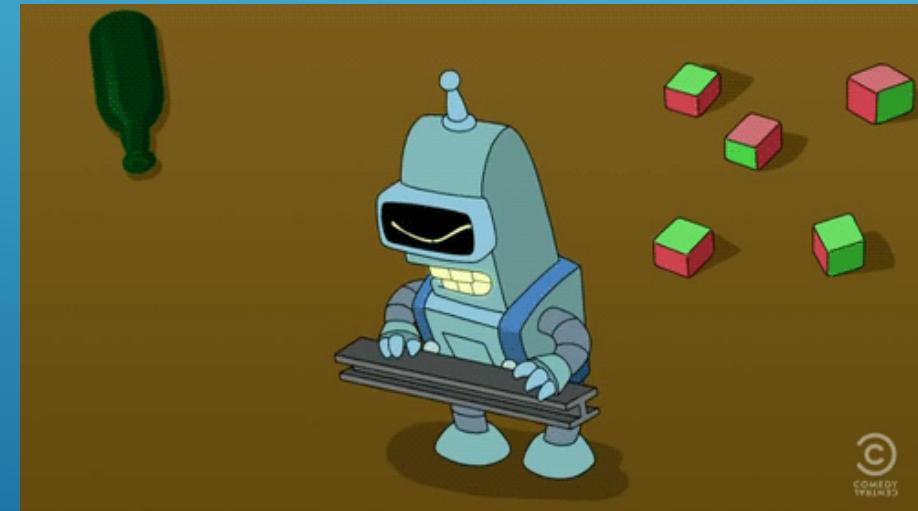
- ▶ Sequential LSTM
- ▶ Optimizer: RMSProp (Also Adam)
- ▶ 800 Epochs
- ▶ 1 Dropout Layer (0.2)
- ▶ Fuzzywords to massage data

LONG SHORT TERM MEMORY RECURRENT NEURAL NETWORK

TEXT GENERATION



My dumb LSTM model



30 Epochs, and still training

30 Epoch Output:

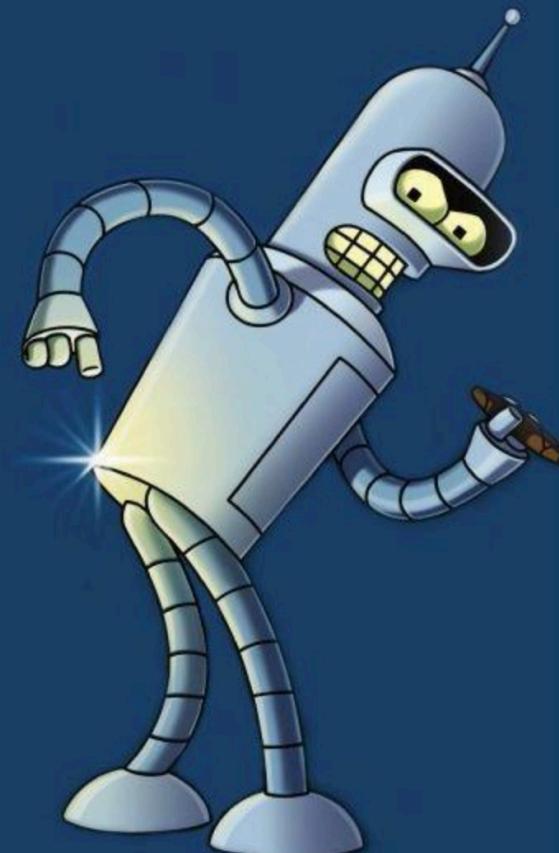
“pid the beat i dont kear the beat i dont kear the
beat i dont kear the beat i dont kear the beat i dont
kear the beat i dont ke”

Welcome to Bender's NLP Analysis

Bender Bot Beta 0.4

Click to get Bender Bot to Speak

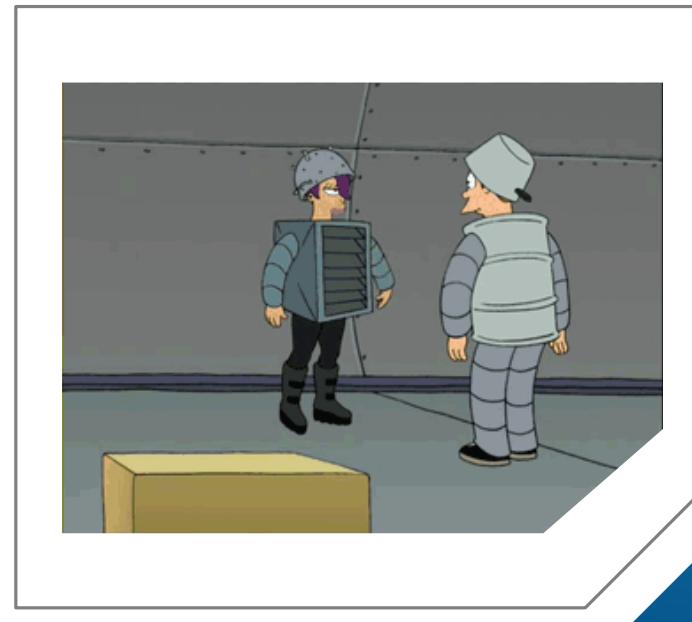
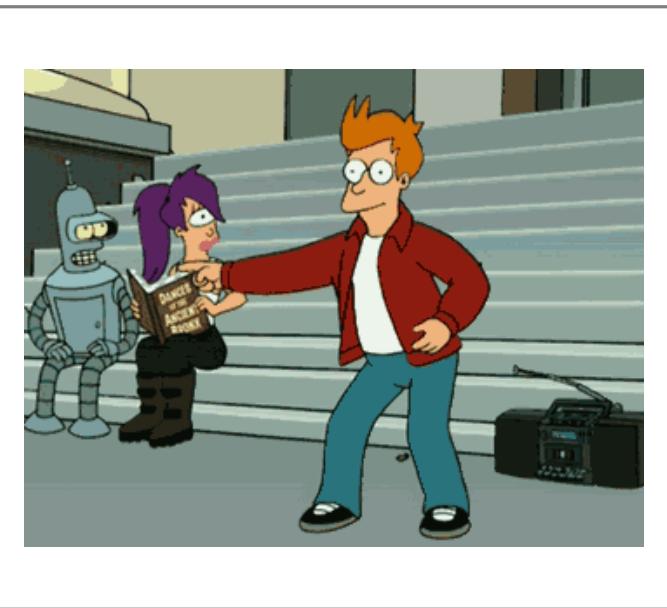
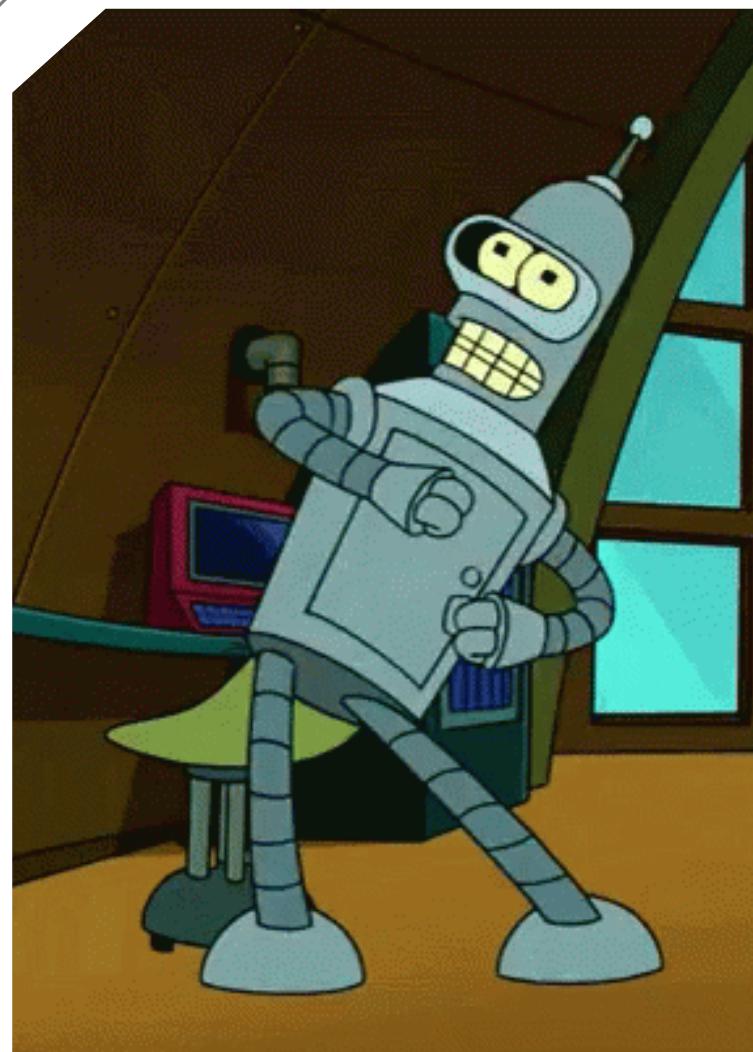
Predict



RECOMMENDATIONS



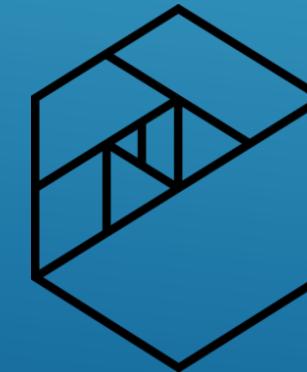
- Apply model concept to multiple shows
- Bender can yell at you when you are lonely
- Use the text generator to create better toys for children
- Identify people from spoken dialogue



THANK YOU!

<https://github.com/IsaacNewtonKim>

<https://www.linkedin.com/in/isaacnewtonkim/>



METIS



- Improve LSTM Model
 - More Layers
 - More Instances
 - Different Loss Functions/Optimizers
- Classification Across TV Shows / Genres
- Final 20 Episodes
- Manual Script Cleaning



FUTURE WORK



WORKS CITED

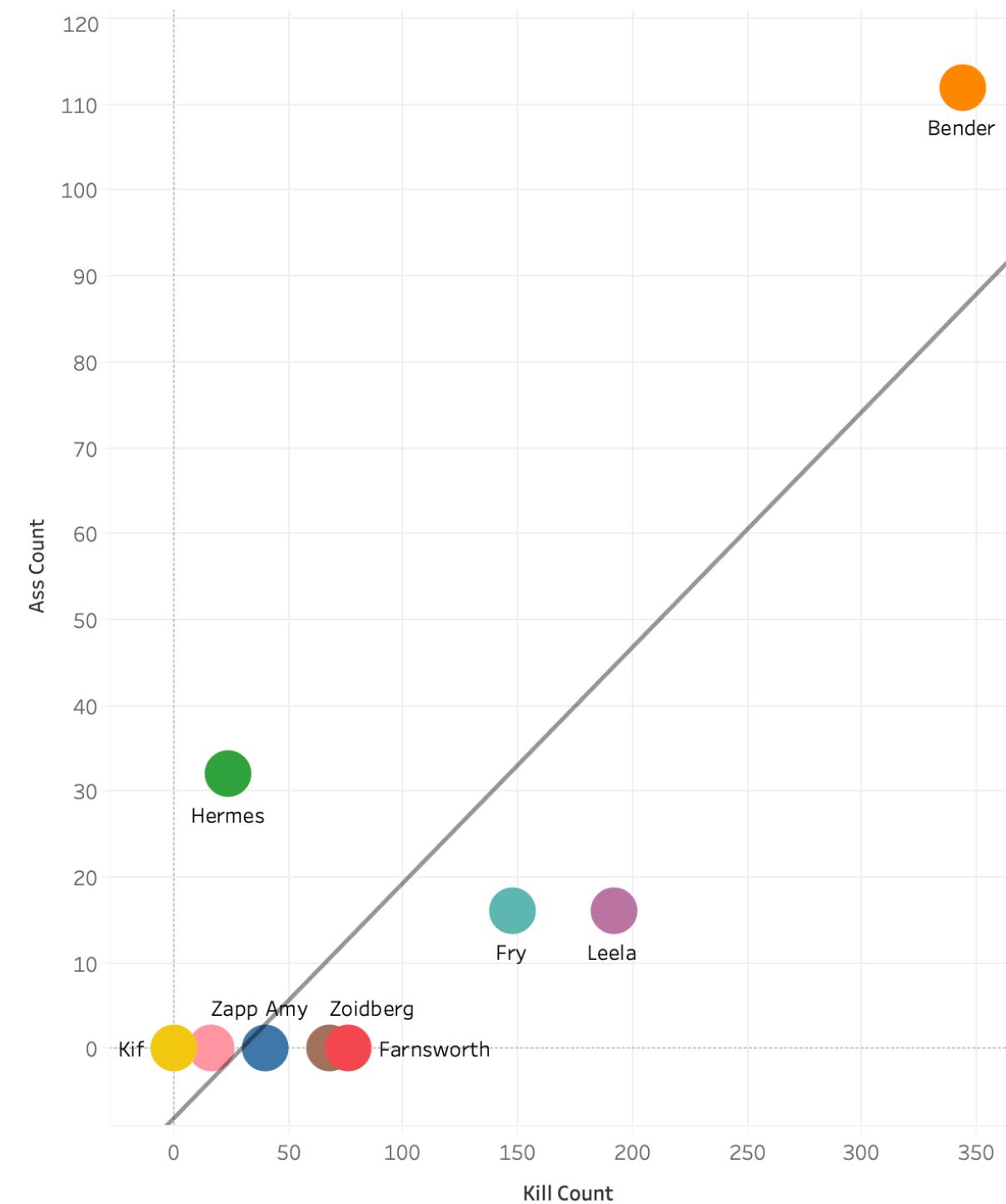
Data collected from: <https://theinfosphere.org>

- LDA: <https://www.machinelearningplus.com/nlp/topic-modeling-gensim-python/>
- Classification: <https://towardsdatascience.com/multi-class-text-classification-model-comparison-and-selection-5eb066197568>
- LSTM: <https://colah.github.io/posts/2015-08-Understanding-LSTMs/>
- LSTM: <https://machinelearningmastery.com/text-generation-lstm-recurrent-neural-networks-python-keras/>

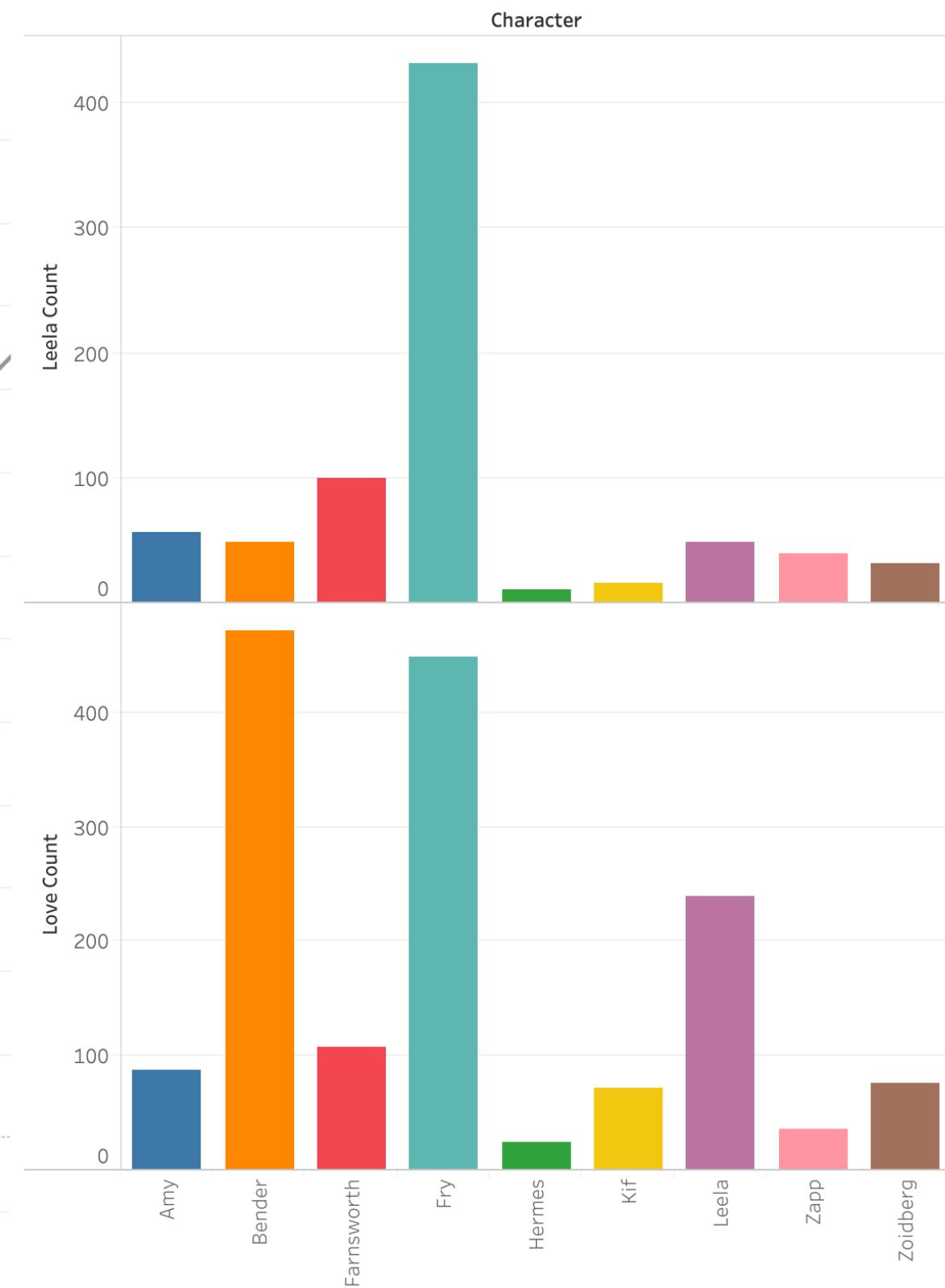


Amy	Fry	Leela
Bender	Hermes	Zapp
Farnsworth	Kif	Zoidberg

'Ass' vs 'Kill' Correlation



Love and Leela Count





NMF

- ▶ Topic 0 earth, wow, world, tv, okay, minute, kid, cool, lot, miss
- ▶ Topic 1 sir, sir sir, um, earth, yes sir, actually, aye, um sir, actually sir, sir shall
- ▶ Topic 2 whoa, whoa whoa, whoa whoa whoa, look like, wash, whoa slow, offensive, whoa whoa slow, slow, object
- ▶ Topic 3 kif, woman, snu, betsy, ready, party, kroker, poor, forever, wave
- ▶ Topic 4 huh, uh huh, slurm, huh hey, blah, huh uh, uh huh uh, calculon, blah blah, huh uh huh
- ▶ Topic 5 body, switch, switch body, mind, bender body, use body, robot body, nixon, fabulous, original
- ▶ Topic 6 human, kill human, point, light, know human, calculon, eye, beep, blood, race
- ▶ Topic 7 ass, metal, shiny, shiny metal, bite, metal ass, shiny metal ass, bite shiny metal, bite shiny, cmon
- ▶ Topic 8 captain, zapp, brannigan, zapp brannigan, animal, new captain, mutiny, space, leela captain, know zapp