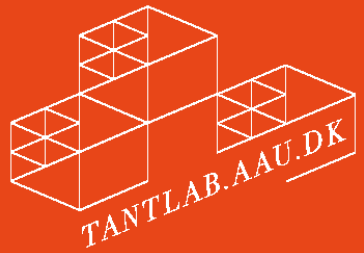
Thick Machines

Anders Kristian Munk, Mathieu Jacomy & Asger Gehrt Olesen

The Techno-Anthropology Lab
AALBORG UNIVERSITY (COPENHAGEN)

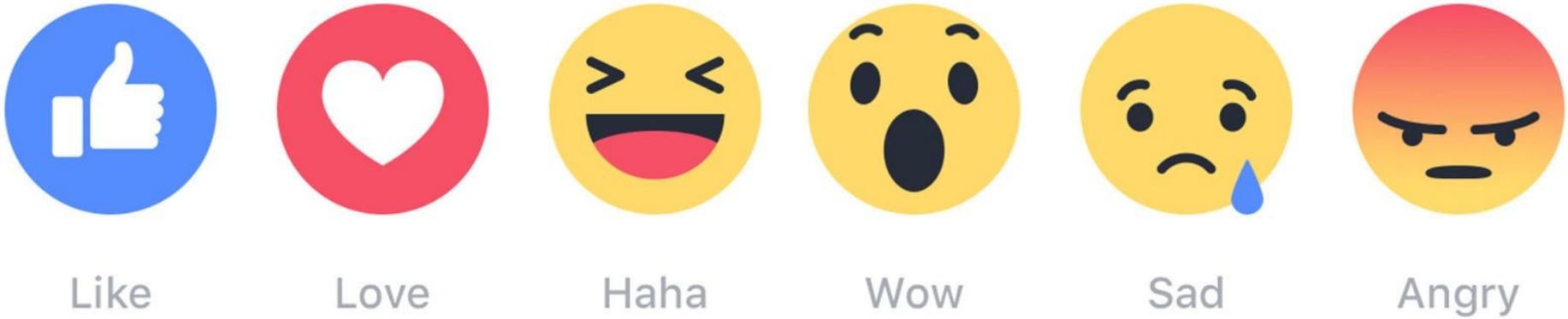
twitter: @TANTlab
web: www.tantlab.aau.dk



Nu er jeg ude igen 😢 kald mig hvad I vil 😢 jeg fjerner bladene 😠 men ser at de fleste af jer 😢 er hamrendende ligeglade med at cypernsvej 😢 ser ud ad pommern til 😢😢😢 kom nu ud og vis at vores vej skal se ordentlig ud 😠

👍😢 2 8 kommentarer Set af 68

'Me again 😢 call me what you want 😢 I remove the fallen leafs 😡 but I see that most of you 😢 couldn't care less that our road 😢 looks awful 😢 come, now, and show that our road should look proper 😡'


Like
Love
Haha
Wow
Sad
Angry

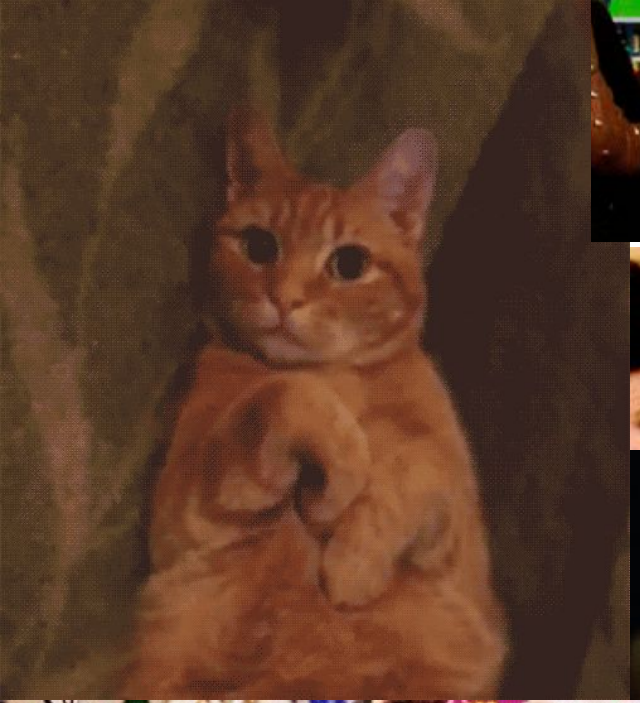
'Does butter go beneath Nutella?'

'Eww. No!'

'Totally!'

TOP: 'My dog just died … crying'

BOTTOM: 'Flemming's funeral will be held at 1pm in the Church of Holte'

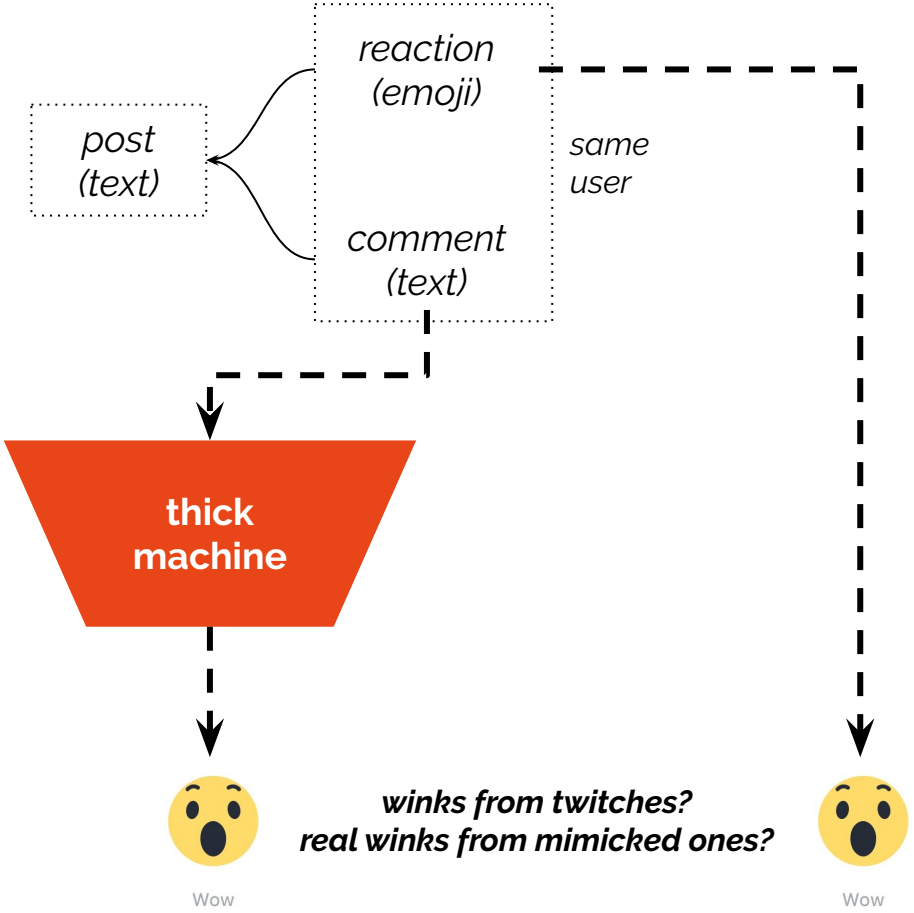


COMEDY CENTRAL

Can we have a Turing test for thick machines?

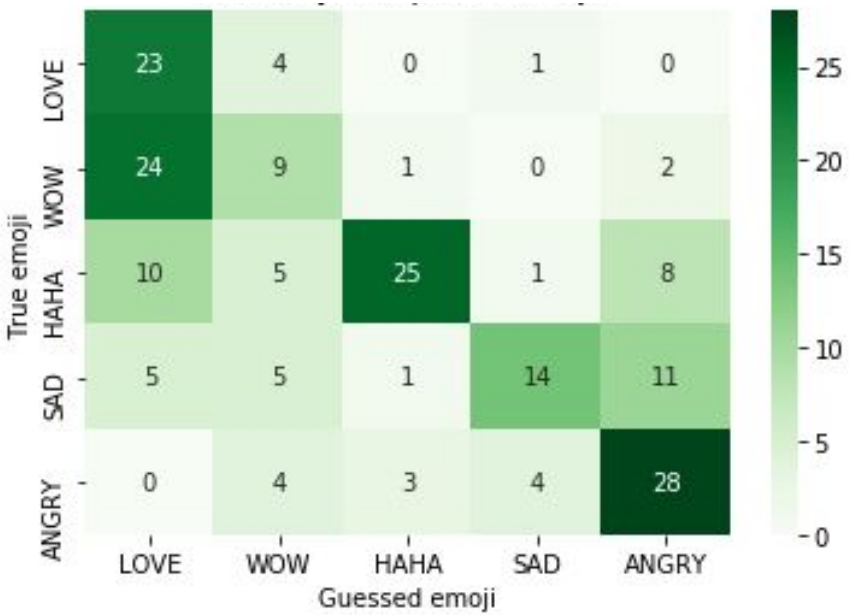
Geertz, C. (1973). ***The Interpretation of Cultures,*** p.16

The claim to attention of an ethnographic account does not rest on its author's ability to capture primitive facts in faraway places and carry them home like a mask or a carving, but on the degree to which he is able to clarify what goes on in such places (...) This raises some serious problems of verification, all right--or, if "verification" is too strong a word for so soft a science (I, myself, would prefer "appraisal"), of how you can tell a better account from a worse one. But that is precisely the virtue of it. ***If ethnography is thick description*** *and ethnographers those who are doing the describing,* ***then the determining question for any given example of it (...) is whether it sorts winks from twitches and real winks from mimicked ones.***



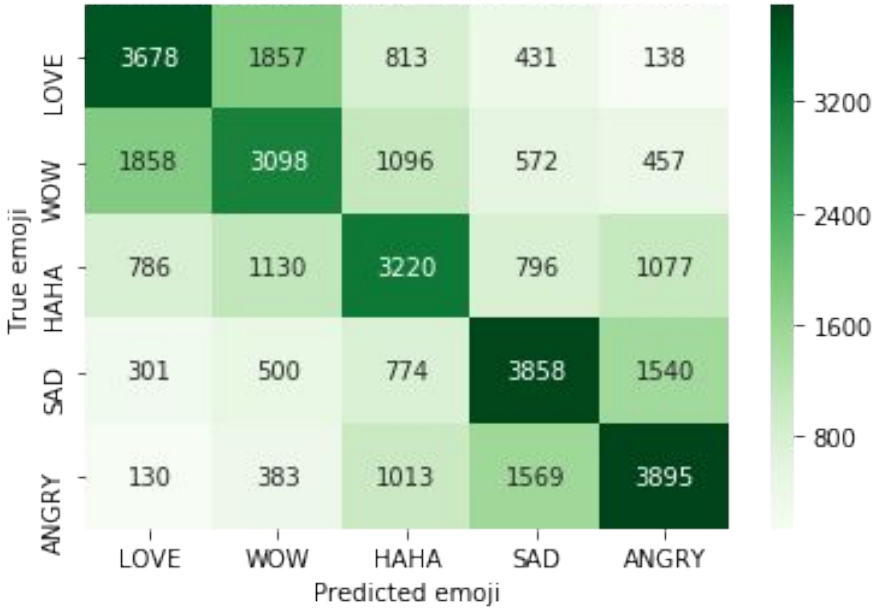
A Turing test for thick machines

HUMANS OF THE TANT-LAB



Accuracy
52%

NEURAL NETWORK TRAINED ON 140K COMMENTS + REACTIONS



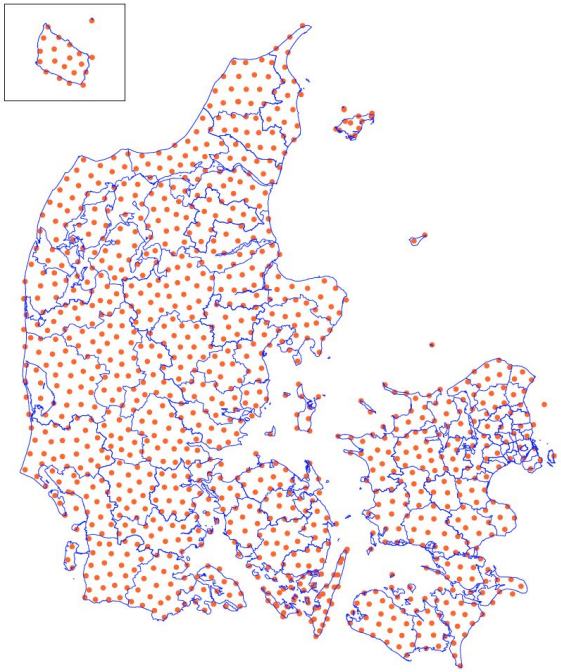
Accuracy
51%

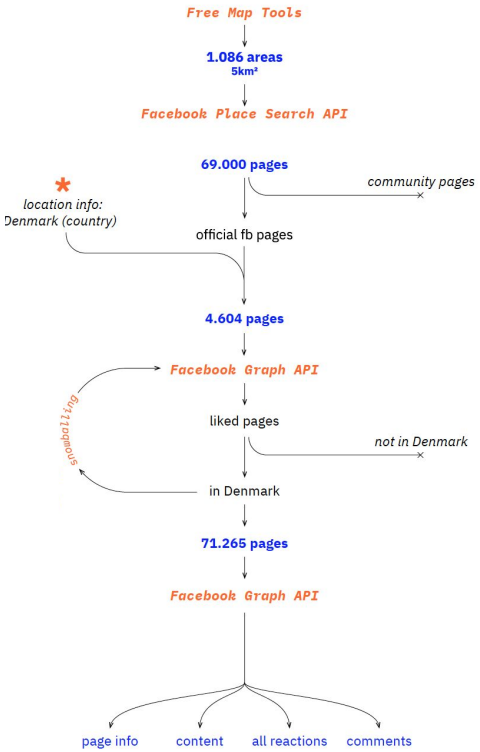
The Atlas of Danish Facebook Culture

A reaction to the 2018 Facebook APIcalypse

What is Danish Facebook?

- 25.000.000 posts by 71.265 Danish pages in the period 2012-2018
- 128.000.000 comments
- 700.000.000 like reactions
- 23.500.000 emoji reactions





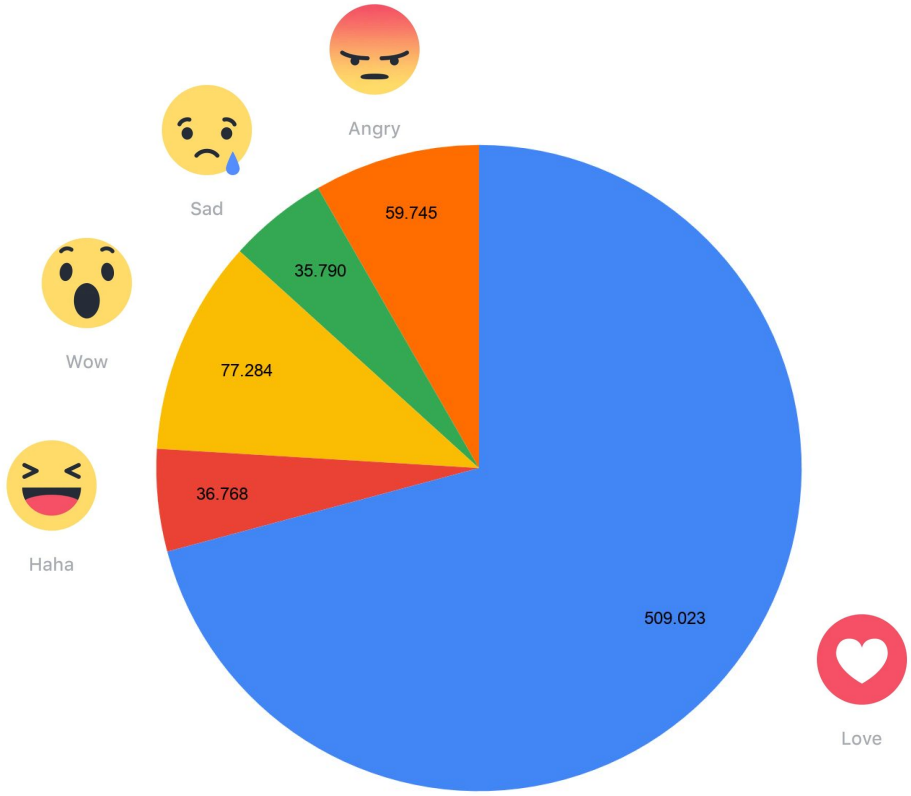
The Atlas of Danish Facebook Culture

Building a subset for The Thick Machine - 1

Input: The Atlas of Danish Facebook Culture

Filter: Pairs of emoji reactions and first comments on made on the same posts by the same users. Length of post text > 50 characters AND comment text > 50 characters

Output: 700.000 reaction/comment pairs

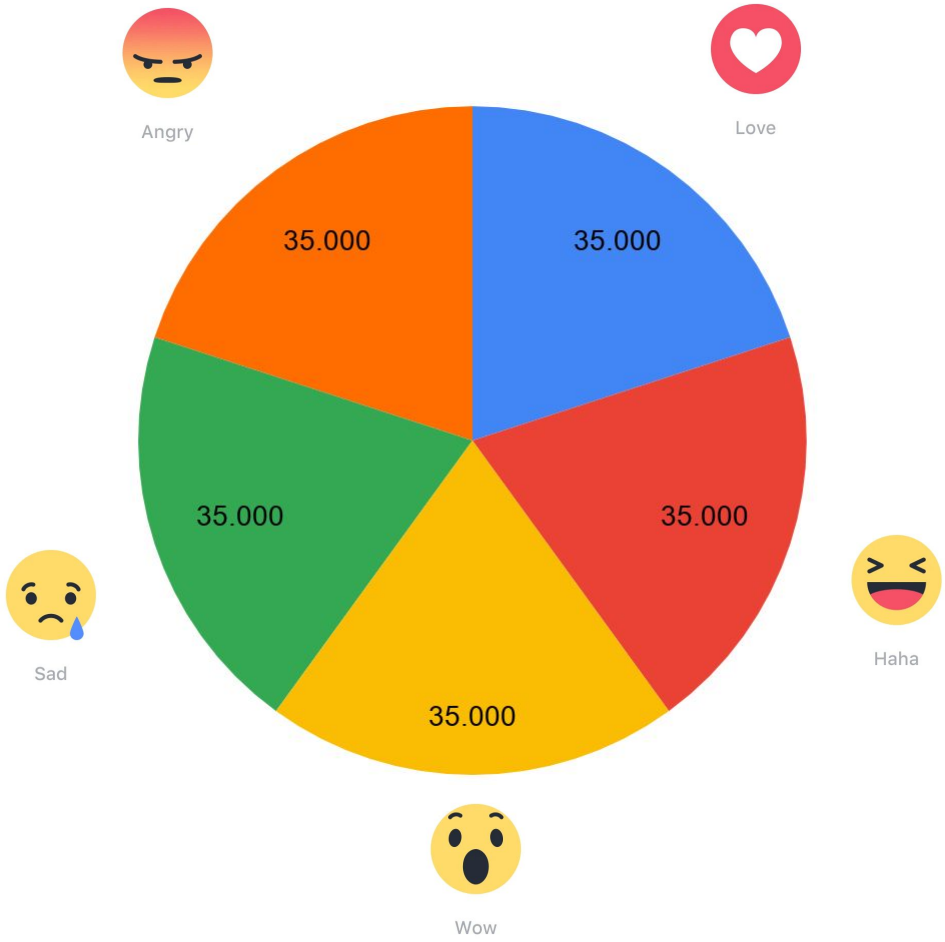

Angry
Sad
Wow
Haha
Love
59.745
35.790
77.284
36.768
509.023

Building a subset for The Thick Machine - 2

Input: 714.000 emoji reaction/comment pairs

Filter: Balance the amount of reactions to get an even distribution

Output: 175.000 emoji reaction/comment pairs (35.000 for each reaction type)



Under the hood

sklearn.neural_network.MLPClassifier

class sklearn.neural_network.**MLPClassifier**(*hidden_layer_sizes=(100,), activation='relu', solver='adam', alpha=0.0001, batch_size='auto', learning_rate='constant', learning_rate_init=0.001, power_t=0.5, max_iter=200, shuffle=True, random_state=None, tol=0.0001, verbose=False, warm_start=False, momentum=0.9, nesterovs_momentum=True, early_stopping=False, validation_fraction=0.1, beta_1=0.9, beta_2=0.999, epsilon=1e-08, n_iter_no_change=10, max_fun=15000*)

[source]

Multi-layer Perceptron classifier.

This model optimizes the log-loss function using LBFGS or stochastic gradient descent.

New in version 0.18.

Parameters:

hidden_layer_sizes : *tuple, length = n_layers - 2, default=(100,)*
The ith element represents the number of neurons in the ith hidden layer.

activation : *{'identity', 'logistic', 'tanh', 'relu'}, default='relu'*
Activation function for the hidden layer.

- 'identity', no-op activation, useful to implement linear bottleneck, returns f(x) = x
- 'logistic', the logistic sigmoid function, returns f(x) = 1 / (1 + exp(-x)).
- 'tanh', the hyperbolic tan function, returns f(x) = tanh(x).
- 'relu', the rectified linear unit function, returns f(x) = max(0, x)

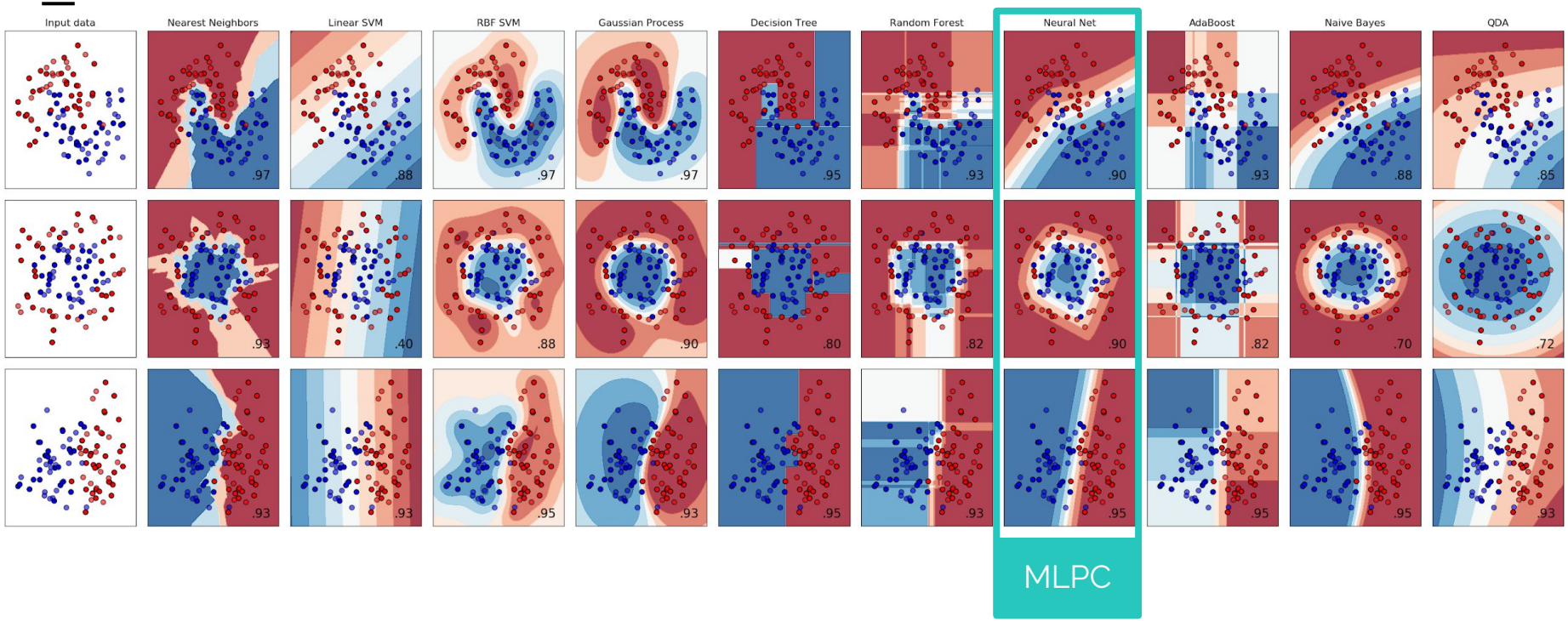
solver : *{'lbfgs', 'sgd', 'adam'}, default='adam'*
The solver for weight optimization.

- 'lbfgs' is an optimizer in the family of quasi-Newton methods.
- 'sgd' refers to stochastic gradient descent.
- 'adam' refers to a stochastic gradient-based optimizer proposed by Kingma, Diederik, and Jimmy Ba

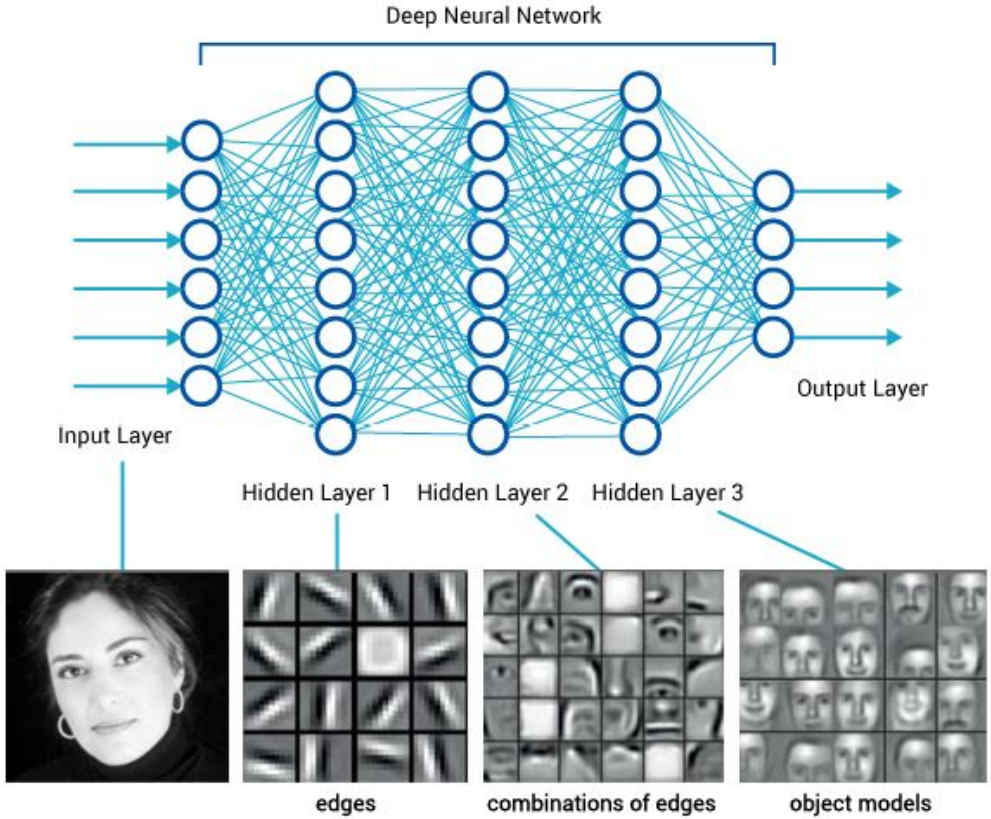
https://analyticsindiamag.com/a-beginners-guide-to-scikit-learns-mlpclassifier/

https://scikit-learn.org/stable/modules/generated/sklearn.neural_network.MLPClassifier.html

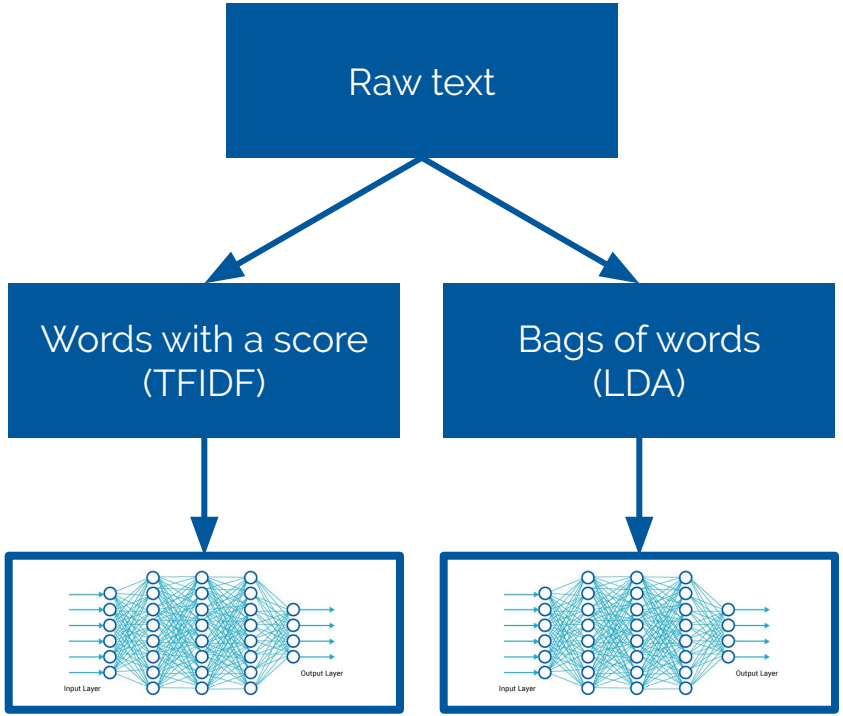
We use SciKit Learn's
MLPClassifier, aka
"Multi-layer Perceptron classifier"



Different classifiers have different behaviors



*MLPClassifier is a **neural network** with multiple (hidden) layers.*



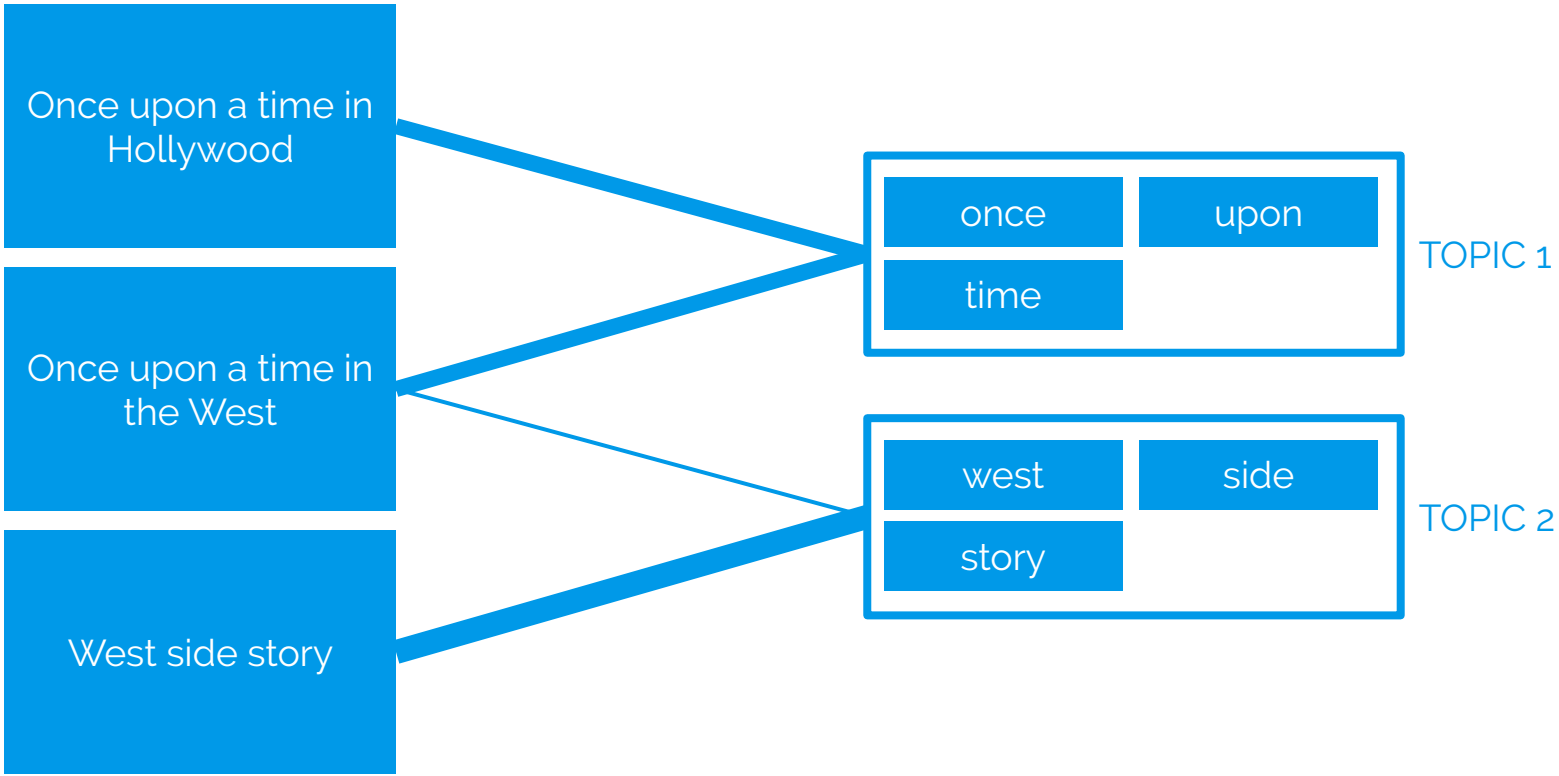
"In machine learning, ... ***feature extraction starts from an initial set of measured data and builds derived values (features) intended to be informative and non-redundant****, facilitating the subsequent learning and generalization steps, and in some cases leading to better human interpretations." - Wikipedia*

Vectorization. Turning text into features (think: columns with numbers).

$$\frac{Occurrences_of_the_word_in_doc}{\log(Doc_containing_word\ /\ All_doc)}$$

	once	upon	time	west	side
Once upon a time in Hollywood	1/log(⅔)	1/log(⅔)	1/log(⅔)	0	0
Once upon a time in the West	1/log(⅔)	1/log(⅔)	1/log(⅔)	1/log(⅔)	0
West side story	0	0	0	1/log(⅔)	1/log(⅓)

Text Frequency x Inverse Document Frequency (TFIDF)



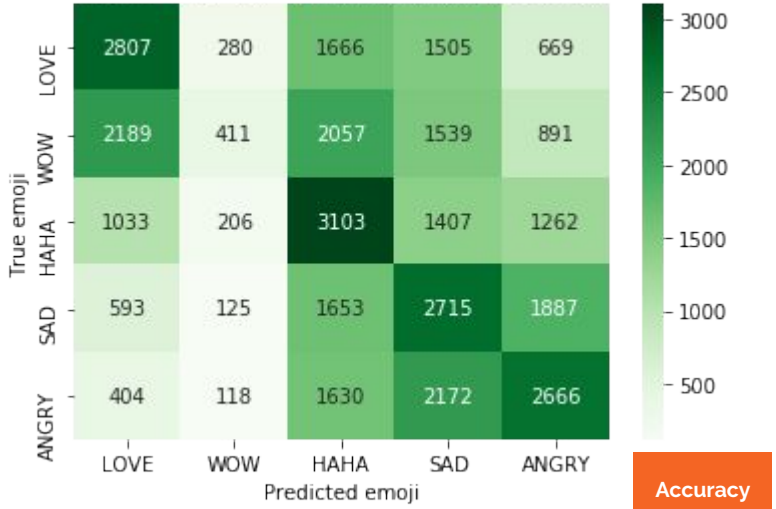
Latent Dirichlet Allocation (LDA)

Results

features \ *classifier*	*MLPC: 1 hidden layer, 100 nodes* *(Multi Layer Perceptron Classifier)*	*MLPC: 2 hidden layers, 1000 + 100 nodes* *(Multi Layer Perceptron Classifier)*
TFIDF *(Term Frequency - Inverse Document Frequency)*	Trained on 1% of dataset: **41% accurate**	Trained on 1% of dataset: **42% accurate**
	Trained on 80% of dataset: **49% accurate**	Trained on 80% of dataset: **51% accurate**
LDA *(Latent Dirichlet Allocation)*	Trained on 1% of dataset: *not done*	Trained on 1% of dataset: *not done*
	Trained on 80% of dataset: **31% accurate**	Trained on 80% of dataset: **32% accurate**

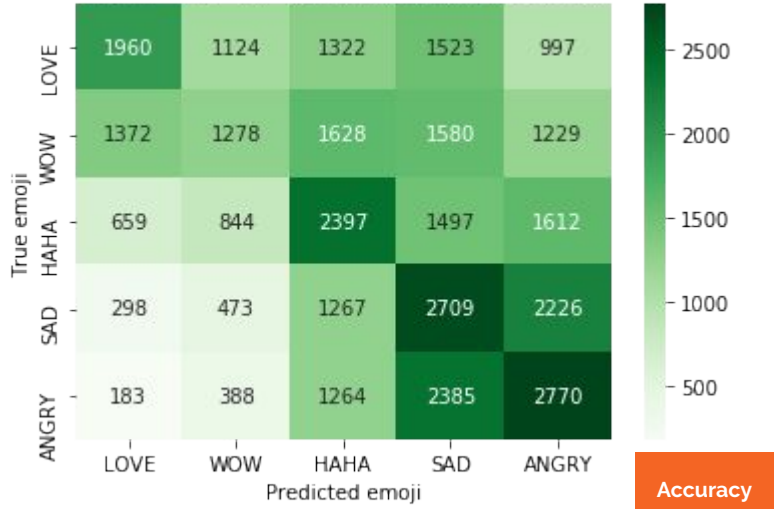
LDA feature extraction, training on 80% of dataset (140,000 comments)

MLPClassifier with 1 hidden layer (100 nodes)



Accuracy
31%

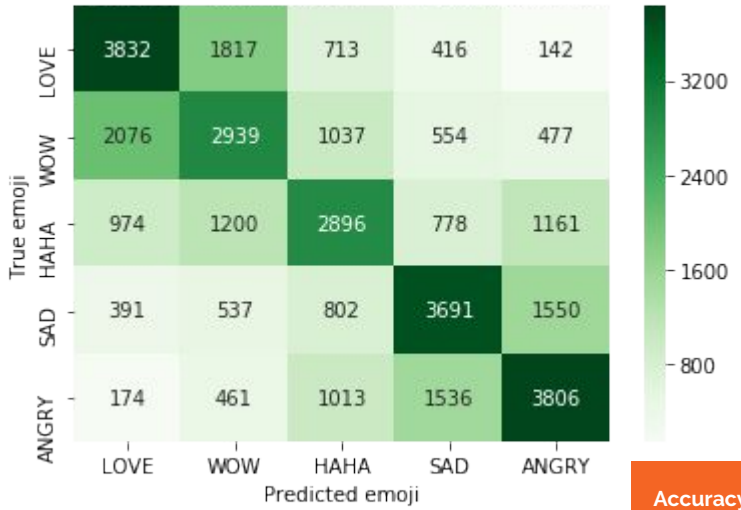
MLPClassifier with 2 hidden layers (1000 + 100 nodes)



Accuracy
32%

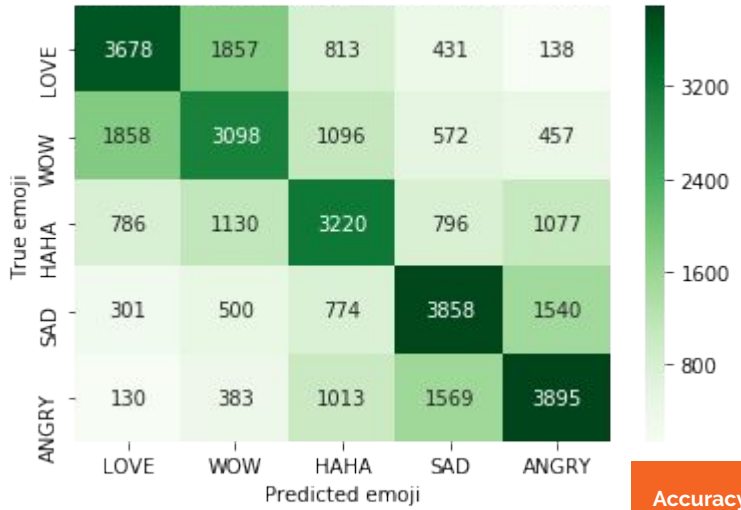
TFIDF feature extraction, training on 80% of dataset (140,000 comments)

MLPClassifier with 1 hidden layer (100 nodes)



Accuracy
49%

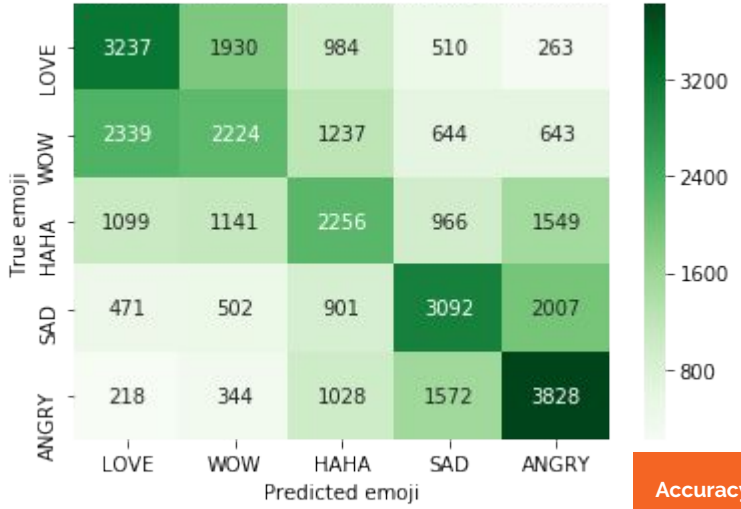
MLPClassifier with 2 hidden layers (1000 + 100 nodes)



Accuracy
51%

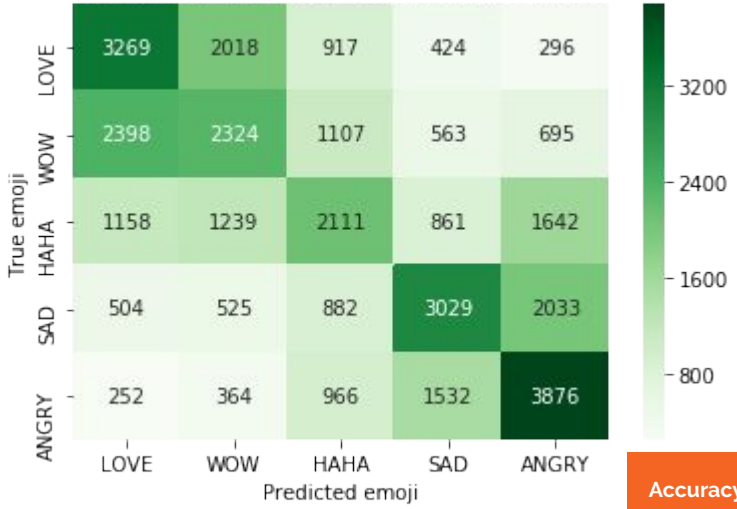
TFIDF feature extraction, training on 1% of dataset (1750 comments)

MLPClassifier with 1 hidden layer (100 nodes)



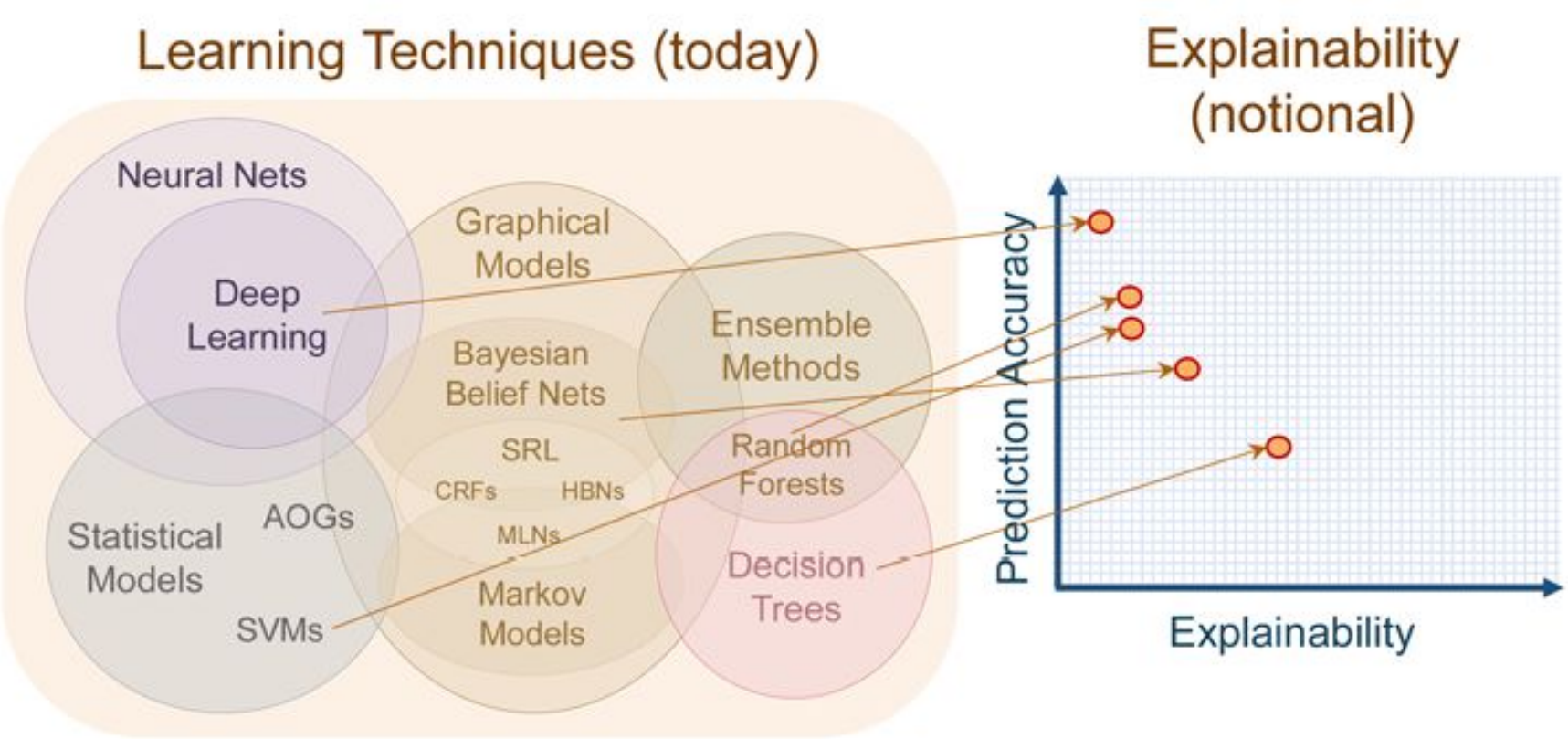
Accuracy
41%

MLPClassifier with 2 hidden layers (1000 + 100 nodes)



Accuracy
42%

Unexplainable machines?

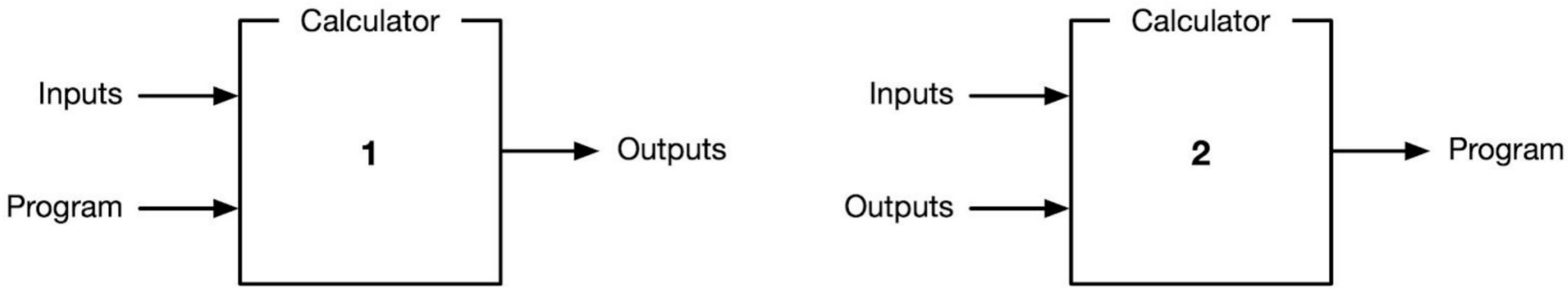


http://nautil.us/issue/40/learning/is-artificial-intelligence-permanently-inscrutable

*A trade-off between **accuracy** and **explainability***

—

Figure 1. Hypothetical-deductive machines (1) and inductive machines (2)



Cardon, D., Cointet, J. P., & Mazières, A. (2018). Neurons spike back. Réseaux, (5), 173-220.

*With **inductive machines**, the outcome is the program.*

Post-hoc interpretations

*Lipton, Z. C. (2016). **The Mythos of Model Interpretability**, (Whi). Retrieved from http://arxiv.org/abs/1606.03490*

"While post-hoc interpretations often do not elucidate precisely how a model works, they may nonetheless confer **useful information for practitioners** and end users of machine learning. Some common approaches to post-hoc interpretations include natural language explanations, **visualizations of learned representations or models**, and explanations by example *(e.g. this tumor is classified as malignant because to the model it looks a lot like these other tumors).*"

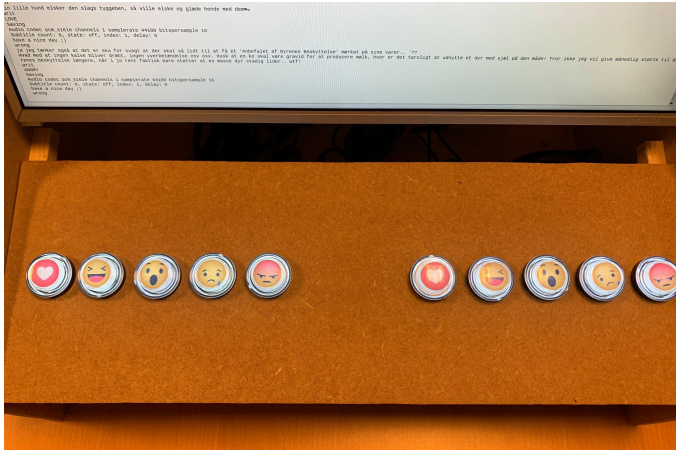
Building the Thick Machine

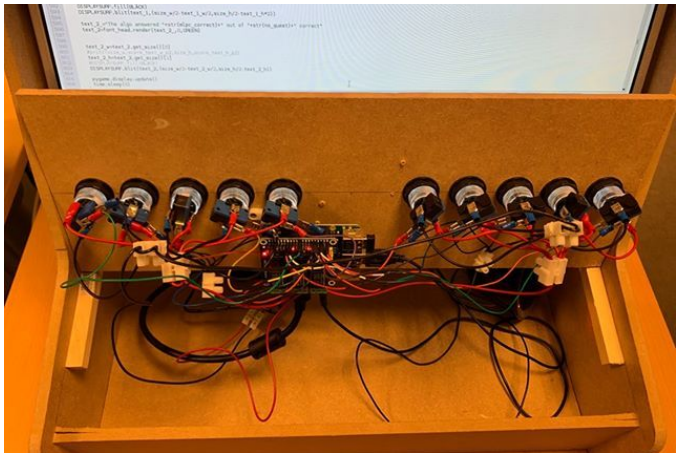
A gamified coding experience

The hardware

Parts list:

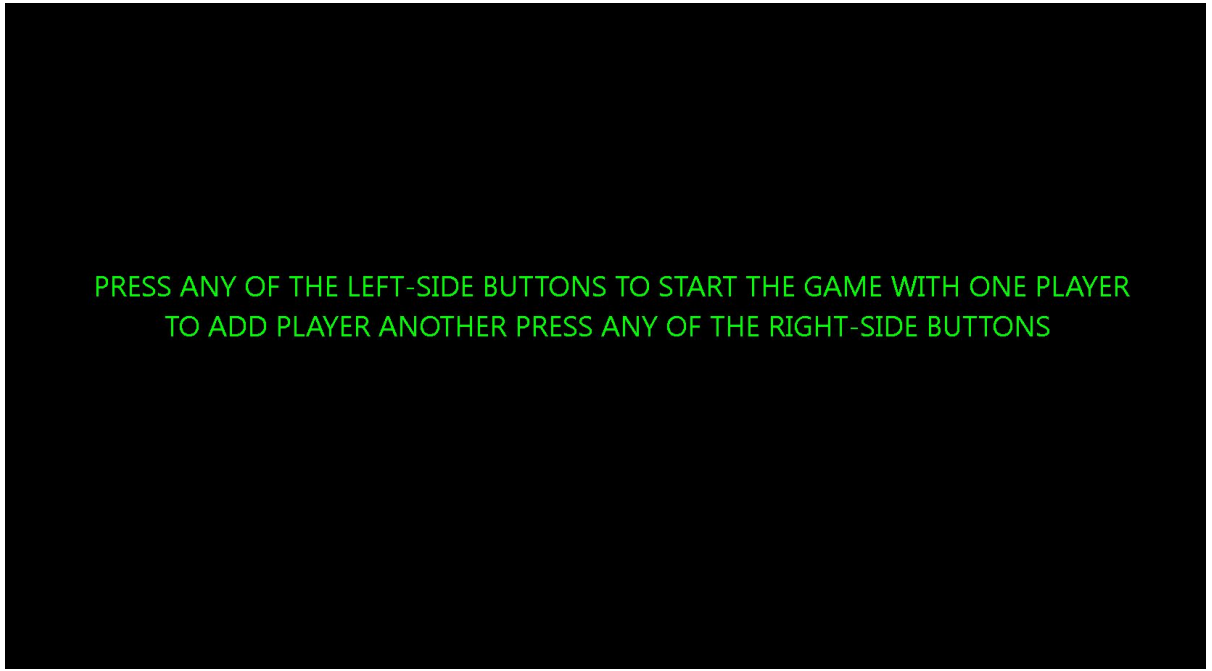
- Raspberry Pi 3B+
- Relay hat for Raspberry Pi
- 10 arcade buttons with light
- 19' computer monitor
- HDMI cable
- Old computer speakers
- Bartop arcade machine cabinet
- Wires and screws





The game

- Written in Python using Pygame
- Interactions through 2x5 digital inputs (emoji buttons)



In [*]:

```
guesses=[]
for comment in comments:
    print("Post")
    print(posts[comments.index(comment)])
    print("_____")
    print("Comment")
    print(comment)
    print("_____")
    print("Enter your guess in the input box below (L for Love, H for Haha, S for Sad, W for Wow and A for Angry)")
    guess=input()
    guesses.append(guess)
    break
```

```
Post
- Jeg har brugt 800 millioner euro til at købe Milan. Jeg tror ikke, jeg går konkurs ved at ofre 30-40 millioner. Donnarumma si
dder på tribunerne, indtil hans kontrakt udløber, har A.C. Milan's præsident Yonghong Li udtalt til Il Giornale.
_____
Comment
Kølig præsident vi har fået. Tror sommeren bliver mere underholdende end selve sæsonen 😂😂💪👊🤘
_____
Enter your guess in the input box below (L for Love, H for Haha, S for Sad, W for Wow and A for Angry)
H
```

Coding script in Jupyter Notebook

Post	Comment	Guess
"- Jeg har brugt 800 millioner euro til at købe Milan. Jeg tror ikke, jeg går konkurs ved at ofre 30-40 millioner. Donnarumma sidder på tribunerne, indtil hans kontrakt udløber, har A.C. Milan's præsident Yonghong Li udtalt til Il Giornale."	Kølig præsident vi har fået. Tror sommeren bliver mere underholdende end selve sæsonen 😂😂💪👊🤘	4

Spreadsheet coding scheme

Machines for explanation or explication?

‘Thickness’ ≠ passing for a native

Geertz, C. (1973). ***The Interpretation of Cultures,*** p.11

Variously called ethnoscience, componential analysis, or cognitive anthropology (a terminological wavering which reflects a deeper uncertainty), this school of thought holds that culture is composed of psychological structures by means of which individuals or groups of individuals guide their behavior. "A society's culture," to quote Goodenough again, this time in a passage which has become the locus classicus of the whole movement, "consists of whatever it is one has to know or believe in order to operate in a manner acceptable to its members." And ***from this view of what culture is follows a view, equally assured, of what describing it is--the writing out of systematic rules, an ethnographic algorithm, which, if followed, would make it possible so to operate, to pass (physical appearance aside) for a native****. (Geertz 1973:11)*

Predicting ❤️‍🩹 when reaction is ❤️‍🩹

Machine: PASSING FOR A NATIVE

Situation: SHALLOW

POST (translated):

..... Christmas present of the century🙏🏻🙏🏻🙏🏻

Predicting 😡 when reaction is 😡

Machine: PASSING FOR A NATIVE

Situation: SHALLOW

POST (translated):

We have filled our 12 storage facilities with road salt, and from October 1st our winter crew is ready to move out when frost and snow hit the roads. Here is a snakpeak from the salt storage. On average 15 tonnes of salt in each hall. Enough salt for approximately 1 billion egg sandwiches 😉

COMMENT (translated):

Why don't you cut the crap and make people learn to drive instead! 😡

Predicting 😡 when reaction is 😂

Machine: FAILING TO PASS FOR A NATIVE

Situation: DEEPER

POST (translated):

For the Prince, the decision not to be buried next to the Queen is the natural consequence of not receiving the same treatment as his spouse when it comes to the title and function he has always desired, says chief of communications for the Royal House, Lene Balleby.

COMMENT (translated):

I will gladly swap problems with him. Happy to refrain from a royal title in exchange for 8 mio a year. Or just the 29 mio for the sarcophagus. Then he can fight my [beep] ex and the system without getting as much as a penny from in return. Get a real problem, King Carrot.

Predicting ❤️ when reaction is 😂

Machine: FAILING TO PASS FOR A NATIVE

Situation: DEEPER

POST (translated):

It's autumn. You're out running and see this pile of fallen leaves :D ...

Are you the type that runs through or around it...?

COMMENT (translated):

Definitely through the pile. It's the most fabulous sound of autumn 😍🍂🍁

Predicting ❤️ when reaction is 😂

Machine: FAILING TO PASS FOR A NATIVE

Situation: DEEPER

POST (translated):

Do you know someone in need of cleaning out the bedroom? Many of our clients demand a special section for adult toys. (...) That is why we're opening a new initiative on the first floor of Børneloppen [flea market for families with kids]. Finally you can sell your private equipment in a discreet and efficient manner.

Toys should be unused or washed with soap. We recommend that you bring batteries since toys that can be tested sell better.

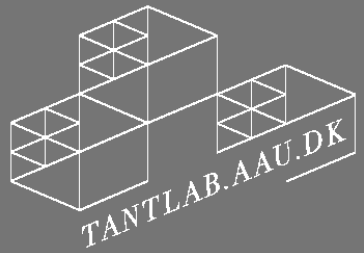
COMMENT (translated):

🤣🤣🤣😂😂 Stine [surname redacted] we have to go and see if we can make room for some new stuff 🤣🤣🤣😂😂

Repurposing algos as explication machines?

THANK YOU

EXTRAS



Predicting ❤️ when reaction is ❤️

Machine: PASSING FOR A NATIVE

Situation: SHALLOW

POST (translated):

👉🏼😊😍 C O M P E T I T I O N 😍😊👈🏼

You and your friend now have the chance to win the beatieful blouses from Molly Jo.

You will win a bordeaux and a blue one, that way you can wear different colours.

(...)

👉🏼 Tag your friend who deserves a treat 😍

(...)

COMMENT (translated):

Red size m and blue size 38 to my friend Karen [surname redacted].💕💕💕.

Predicting ❤️ when reaction is ❤️

Machine: PASSING FOR A NATIVE

Situation: SHALLOW

POST (translated):

We wish you an amazingly merry Christmas Eve with your loved ones 🎄🎆

▸▸ Question of the dayl: Who would you spol with an extra present 🎁🎅

Answer the question and participate in the competition for 2 x 600.- gift cards to sackit.dk (total value of 1200.-) 😍🎵💃

COMMENT (translated):

My lovely husband Poul [surname redacted] who supports me everyday after I injured my back and now need help to do so many things. If anyone really deserves to be spoiled it is him ❤️ Crossing everything I've got and hoping this will be my lucky time 👍🙏🍀🍀🍀❤️

Predicting 😡 when reaction is 😂

Machine: FAILING TO PASS FOR A NATIVE

Situation: DEEPER

POST (translated):

This summer the teachers won't even get to see the test assignments for their own students. The evaluation is left entirely to the assessors. What will this mean?

COMMENT (translated):

But - why? Perhaps there will also come a day where we are no longer allowed to teach our own students?

Predicting when reaction is

Machine: PASSING FOR A NATIVE

Situation: SHALLOW

POST (translated):

Smart: Text a live chicken to a poor family!

Text CHICKEN to 1911 (20 kr), and donate a chicken. This way you help them help themselves.

Do write a comment in the thread below if you choose to send a chicken! Happy easter!

COMMENT (translated):

How many “chickens” end up as administration costs or fat salaries for management?

Predicting 😡 when reaction is 😡

Machine: PASSING FOR A NATIVE

Situation: SHALLOW

POST (translated):

Then it f..... Happened again again 😤😤 burglary at the club house 👊👊👊.

Damit I hope you throw up from all the beer/soda you stole.

Could you for fuck sake find something else to do than ruining it for others 😤😤😤

COMMENT (translated):

Arghhh...god damn it how annoying
😡😡😡😡😡👊🏻👊🏻👊🏻👊🏻👊🏻👊🏻👊🏻👊🏻👊🏻💥💥
💥💥💥💥💥💥💥💥💥💥

Predicting 😢 when reaction is 😂

Machine: FAILING TO PASS FOR A NATIVE

Situation: DEEPER

POST (translated):

UPDATE: The driver who was involved in the car crash has now passed away.

COMMENT (translated):

Andreas [surname redacted]] 😦so it was pretty bad after all