



# COSMOS (Cluster 6, Week 3)

Ian Lin, Leonard Collomb

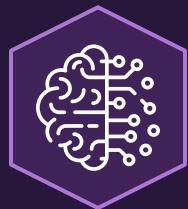
# What We Learned



## Neural Networks

Systems inspired by the human brain; output has an input, weight, and bias; backpropagation, loss functions, gradient descent, hyperparameters, data augmentation

\*



## NLP

Field of ML where machines learn to understand human languages; ASR (Whisper), TTS (Fastspeech 2), Text Generation (GPT); env. impact, hallucinations

\*



## Hearing

Outer, middle, inner ear; loudness perception (90+, 130+ dB); humans are more sensitive to lower frequencies; we tend to associate frequency and intensity

# Project Overview

## Research Q's

How can NLP (Natural Language Processing) be used to identify the targets of hate speech in social media?

Which model would be optimal for these types of text classification?

## Methods

Used TF-IDF Vectorization to tokenize hate speech dataset with 136k rows from Hugging Face

Trained, then compared the performances of 23 Scikit-Learn models

## Near Future...

Tune hyperparameters to finalize most effective models

Further error analysis

Compare most effective models with existing methods

# Results

Model	Accuracy	F1 Score	Average
ETC	0.978497207	0.9773853882	0.9779412976
RFC	0.9780767614	0.9769541621	0.9775154617
DTC_SINGLE	0.9767553607	0.9762658597	0.9765106102
LOGREG_CV	0.9766352333	0.9761432026	0.976389218
SVC	0.9769355517	0.9757831513	0.9763593515
LINSVC	0.974352814	0.9734115049	0.9738821595
ETC_SINGLE	0.9717100126	0.97114699	0.9714285013
RIDGE	0.9678058742	0.9655955792	0.9667007267
PERC	0.9666046009	0.9664367242	0.9665206625
VC	0.9625803352	0.960084781	0.9613325581
SVM	0.9572346687	0.954651305	0.9559429869
CCCV	0.9569944141	0.9548422532	0.9559183336
LR	0.9562135864	0.9533245286	0.9547690575
OVO	0.9563937774	0.9526216066	0.954507692
OVR	0.9563937774	0.9526216066	0.954507692
GBC	0.950147156	0.9438884366	0.9470177963
ABC	0.9481049913	0.9439320835	0.9460185374
KNN	0.9417382425	0.9406206615	0.941179452
DTC	0.9417983062	0.9326339054	0.9372161058
MNB	0.9342302841	0.9210917443	0.9276610142
NC	0.9165115022	0.9182445062	0.9173780042
BNB	0.8914048892	0.9038206582	0.8976127737
OCC	0.8980719563	0.8498447448	0.8739583505

~94% Accuracy Overall!

Confusion Matrix: Sexuality

