Welcome Here!!

NATURAL LANGUAGE PROCESSING (NLP)

INTRODUCTION TO AI, ML, DL AND NLP

DSN LEKKI-AJAH
BY ABEREJO HABEEBLAH O.
X (TWITTER): @HABEREJO

Day 1

Tuesday, 4th March, 2025

Why are we here

To equip learners with skills on how Artificial Intelligence Models are built, learn Machine Learning, Deep Learning and focus on Natural Language Processing (NLP).

We should build some project right?

Yes!!

If you are in with me,

We should build some project.

Classes Structure

Every Tuesday throughout the month of March.

4 classes in total.

Each class for about 1:55 minutes

Days are NOT likely to change, although, my schedule can be OUCH, I wil reachout earlier before. Understand me bikoooo

Who should be here

Anyone willing to learn something new

You aspire to work as a Data Scientist

You aspire to work as a Machine Learning Engr.

You are a researcher with interest in AI

which of these are you?????

Who should be here

Have some experience handling data

Understand basic python programming language

Have some idea about ML and Deep learning

Eager to learn something new

You ticked any of these boxes???
Let's Gooooooo

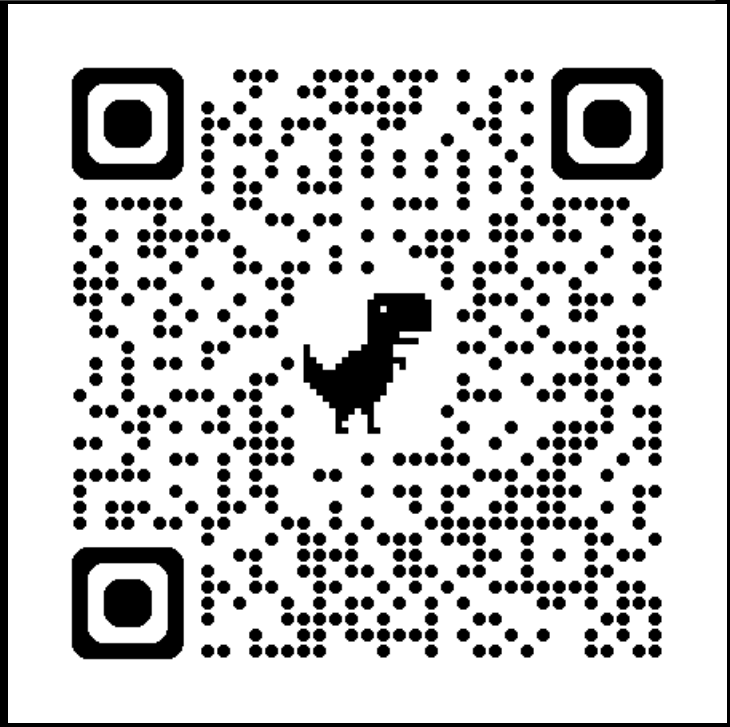
Where Do We Begin

- Understanding Machine Learning
- Machine Learning common algorithms (aka Shallow Algorithms)
- Explain Deep Learning
- Deep Learning vs Machine Learning

Don't worry, I understand you bruh,
I will briefly discuss ML and
Deep Learning before moving to NLP proper.

Who I am?

- A Machine Learning Engineer
- A young boy eager to master AI



Scan to visit my portfolio
or
https://bheez.netlify.app

Course Structure

- Understanding Machine Learning
- Machine Learning common algorithms (aka Shallow Algorithms)
- Explain Deep Learning
- Deep Learning vs Machine Learning
-

Don't worry, I understand you bruh,
I will briefly discuss ML and
Deep Learning before moving to NLP proper.

What is Artificial Intelligence (AI)

Artificial Intelligence is the ability of machines to mimic human intelligence. It enables computers and systems to learn from data, solve problem, and make decisions just like humans do.

AI is when machines start acting like they are smarter than humans.

What is Artificial Intelligence (AI)

Let's relate like this again:

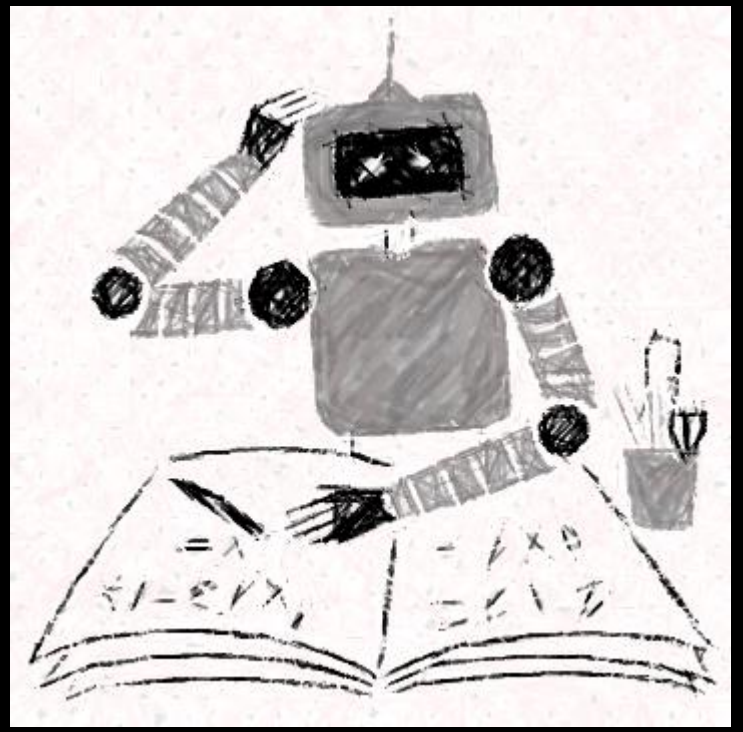
AI is like teaching your computer to do the thinking for you, only that it doesn't demand coffee breaks or complain about Mondays!!

hahaha😀😀😀😀

AI is when machines start acting like they are smarter than humans.

What is Machine Learning (ML)

Machine learning is turning things (data) into numbers and finding patterns in those numbers.



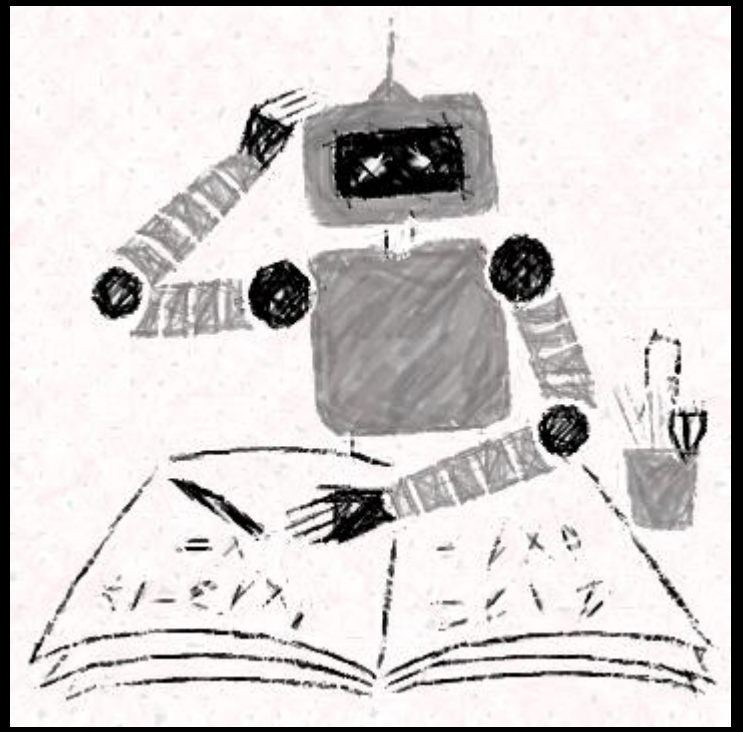
Got It???

Let's Goooooo!!!

What is Machine Learning (ML)

Think of it as teaching a computer how to recognize patterns without explicitly programming it.

It uses some complex algorithms to find patterns.



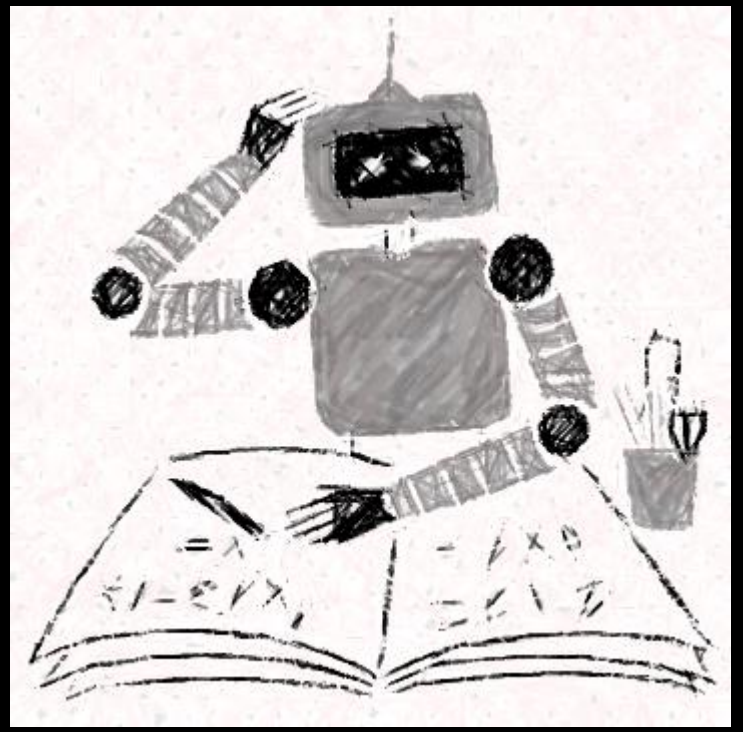
Got It???

Let's Goooooo!!!

What is Machine Learning (ML)

Find pattern? How?

Imagine teaching a child to identify fruits. You show them a banana, an apple, and an orange, telling them the names. After seeing many examples, they can recognize a new banana without needing you to tell them. That's what ML does—it learns from past examples and generalizes.



Got It???

Let's Goooooo!!!

What is Deep Learning (DL)

A deeper version of ML. It works like our brain with many layers of learning.

Instead of just learning basic rules, it finds complex patterns automatically by using lots of examples.

Got It???

Let's Goooooo!!!

What is Deep Learning (DL)

It uses multi-layered neural networks (i.e. Deep Neural Networks) to automatically extract features from data.

It is especially effective for tasks like image recognition, speech processing, and NLP.

Got It???

Let's Goooooo!!!

What is Deep Learning (DL)

Neural networks? How?

Imagine teaching a robot to recognize handwriting. Traditional ML would require manually programming rules for different letters. But Deep Learning would just look at thousands of handwritten examples and figure out the differences itself—like how humans learn handwriting by seeing and practicing.

Got It???

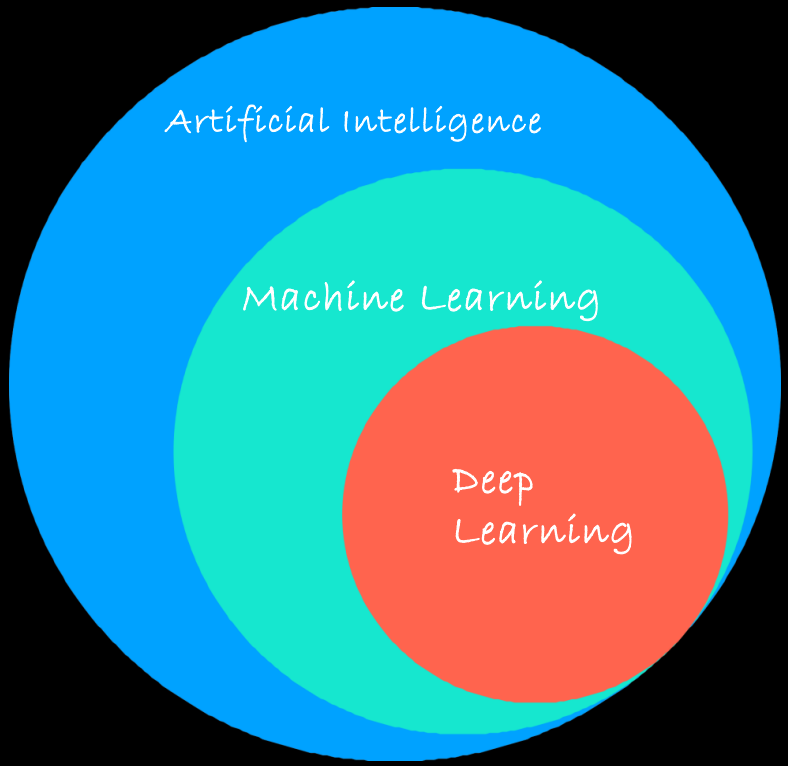
Let's Goooooo!!!

AI vs ML vs Deep Learning (DL)

AI is like a **robotic assistant** that can follow instructions,

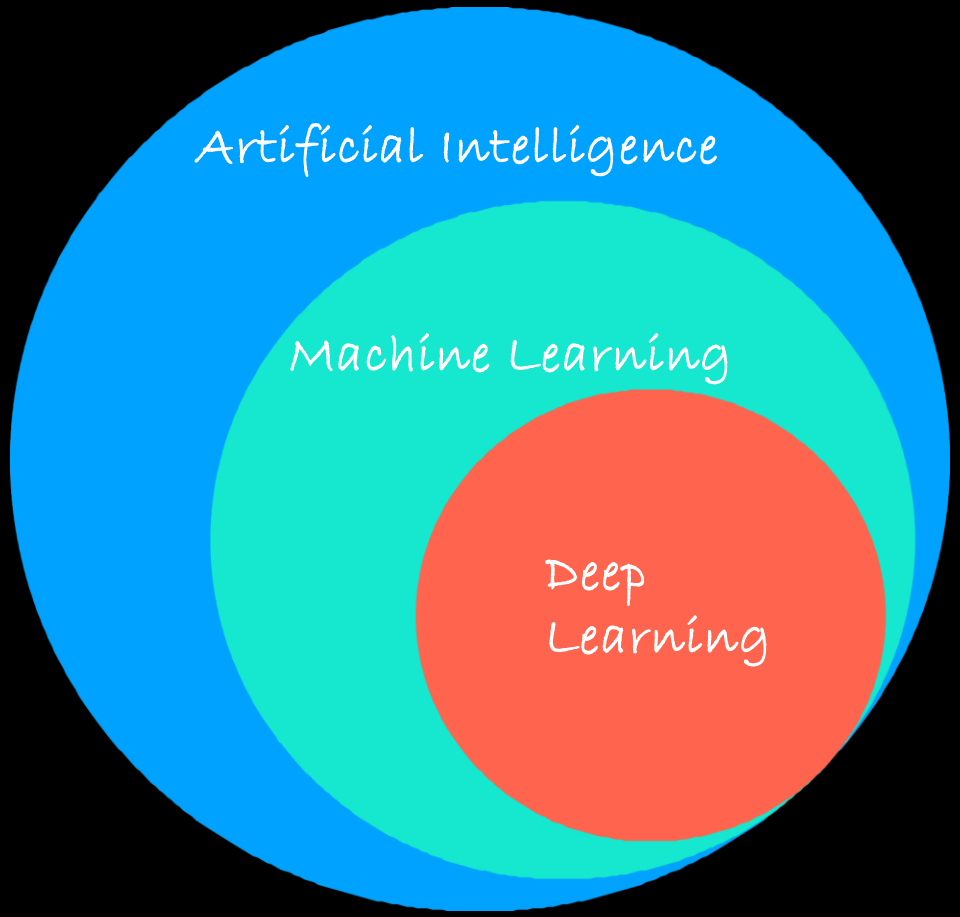
ML is when it learns from experience,

DL is when it learns so deeply that it can recognize faces, voices, and even create art! 🎨🤖



Where do we group NLP??

AI vs ML vs Deep Learning (DL)

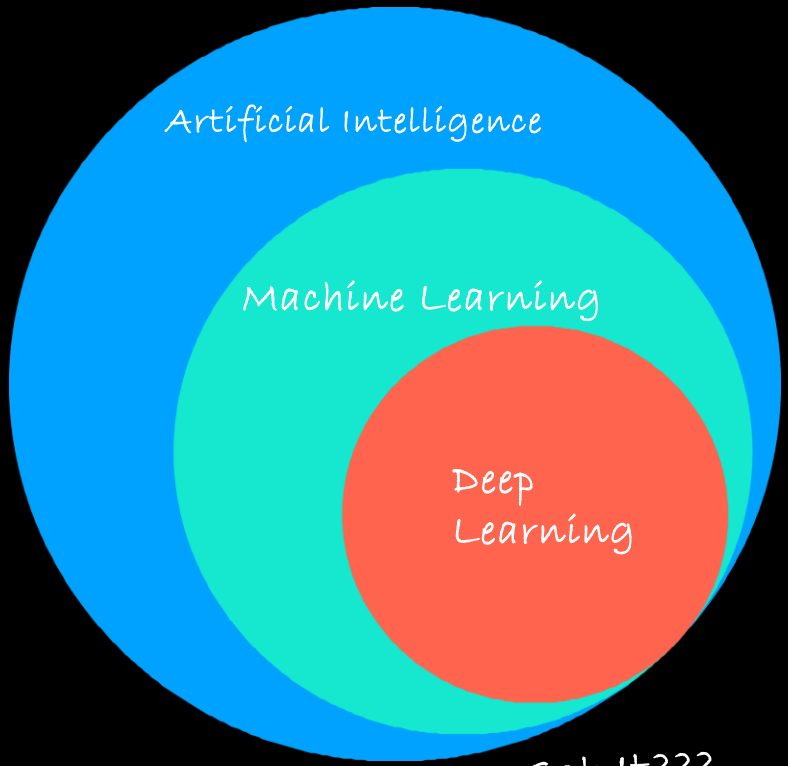


NLP is a sub of Deep Learning..

AI vs ML vs Deep Learning (DL)

Let's relate with this:

AI is like a kitchen assistant following recipes, ML is like a cook who tweaks dishes based on feedback, and DL is like a top chef who invents new recipes just by understanding flavors! 🍳🤖🔥

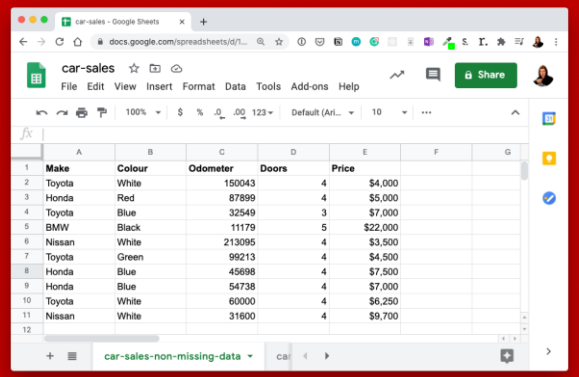


Got It???

Let's Goooooo!!!

ML vs DL

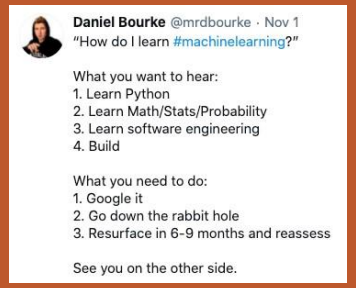
Machine Learning

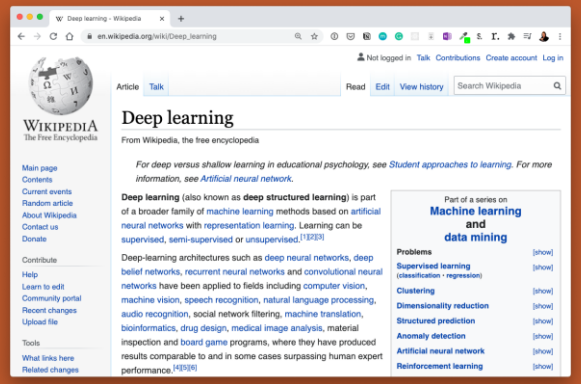


Make	Colour	Odometer	Doors	Price
Toyota	White	150043	4	$4,000
Honda	Red	87899	4	$5,000
Toyota	Blue	32549	3	$7,000
BMW	Black	11179	5	$22,000
Nissan	White	213095	4	$3,500
Toyota	Green	99213	4	$4,500
Honda	Blue	45698	4	$7,500
Honda	Blue	54738	4	$7,000
Toyota	White	60000	4	$6,250
Nissan	White	31600	4	$9,700

Structured data

Deep Leanring





Unstructured Data

What is Natural Language Processing (NLP)

NLP helps computers understand and talk like humans.

It enables computers to read, interpret, and generate human-like text.

It allows your phone to predict your next word, understand voice commands, and translate languages.

NLP combines ML and Deep Learning to process and understand human language.

Includes tasks like stemming, tokenization, NER, sentiment analysis.

NER – named entity recognition

Here is a scenario; Imagine chatting with a smart assistant (like ChatGPT or Siri). You say, "What's the weather today?", and it understands your words, fetches present weather data, and responds. That's NLP in action.

Natural Language Processing (NLP) in Industry

Voice Assistants 🎙️ **:** Siri, Alexa, and Google Assistant understand and respond to your voice commands.

Chatbots 💬**:** Websites use NLP-powered bots to answer your questions instantly.

Autocorrect & Predictive Text 📱 **:** Your phone suggests next words while typing to speed things up, Also can correct spellings when you typed wrongly.

Search Engines 🔍**:** To understands what you mean, even if you type a messy query.

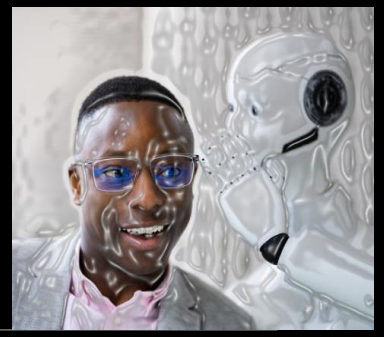
Translation Apps 🌍**:** Google Translate helps you understand different languages.

Spam Filters 📧**:** Your email sorting out spam messages automatically.

Sentiment Analysis 😀😡**:** Companies analyze tweets and reviews to see what people feel about their brand.

All this industry examples uses NLP in operation. Can I get from you??

NLP use cases



"What's the topic of this text?" (text classification)

"Does this text contain abuse?" (content filtering)

"Does this text sound positive or negative?" (sentiment analysis)

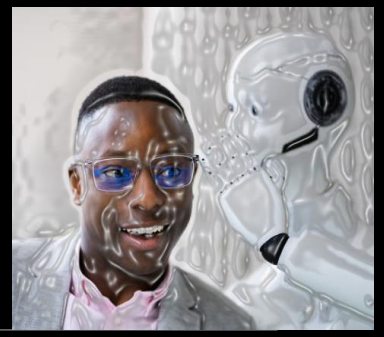
"What should be the next word in this incomplete sentence?" (language modeling)

"How would you say this in German?" (translation)

"How would you summarize this article in one paragraph?" (summarization)

Got The Point Now????

Brief History of Natural Language Processing (NLP)



2011: George Dahl and Microsoft Research trained a deep neural network to recognize words from speech recordings—a major deep learning breakthrough in NLP.

2012: AlexNet revolutionized deep learning in image recognition, leading to increased interest in machine vision.

2015: Deep learning models started competing in NLP tasks like machine translation, matching traditional ML accuracy with less effort and lower computational cost.

2016-2017: Deep learning models in NLP became more efficient and began outperforming traditional ML approaches in accuracy.

Impact: Enabled real-time translation on mobile devices without requiring an internet connection, making NLP more accessible and practical.

All this industry examples uses NLP in operation.

What's Next 😀 😀

Maybe I don't know also

After all of the theoretical explanation, can we begin some hands-on or we stop here today?

What's Next

Maybe I don't know also

If Yes, 

Let's get our Machine setup for hands-on.

Setting up our PC

We need install some tools if we are to work locally (i.e without internet)

Since we are online, we may not delve into this page in details

If you will be working locally, ensure to install the following

1. Python (ensure to add to path when installing)
2. Jupyter Lab (I like to use command - pip install jupyterlab – in my CMD/Terminal)
3. Tensorflow (use the command - pip install tensorflow – into your CMD/Terminal)

Some frameworks and libraries we will use as we move

If working online,
visit google colab and let's start

What's Next 😀 😀

Maybe I don't know also

After setup, 

Let's begin proper proper

The END

DSN LEKKI-AJAH

BY ABEREJO HABEEBLAH O.

X: @HABEREJO

Day 1

NATURAL LANGUAGE PROCESSING (NLP)

TEXT PREPROCESSING AND TOKENIZATION

DSN LEKKI-AJAH

BY ABEREJO HABEEBLAH O.

X: @HABEREJO

Day 2

How computer sees language

Computers see text as a sequence of character,

Computers don't understand words or meaning like humans do

They see all English words as binary language (i.e. 1s and 0s)

So, now, we need to help computer understand our language to what it can fully understand

What do we do to help computers understand the text? Let's GO

Binary Language == 1 and 0

Understand Human Language

There are arguably 3 steps to help computer understand human texts,

1. Text Preprocessing
2. Tokenization
3. Numerical Representation (Word Embedding)

Now, Lets discuss each further!!

Binary Language == 1 and 0

Step 1: Text Preprocessing

Why Preprocess Text?

Imagine you have a room full of clothes, toys, and papers scattered everywhere. Finding what you need is hard, right?

Raw text data is similar. It's full of "**noise**" – things that make it difficult for computers to understand the meaning.

Got It???

Let's Goooooo!!!

Step 1: Text Preprocessing

At this stage, lets understand preprocessing of text,

Proper Text preprocessing is important to make sure we are only feeding the right processed information into our algorithm and irrelevant text are gone.

e.g: The Quick! bRoWn foX juMPS.

Got It???

Let's Goooooo!!!

Step 1: Text Preprocessing

Text preprocessing cleans and prepares text data for NLP tasks. It addresses issues like inconsistencies (uppercase/lowercase), irrelevant information (stop words), and variations in word forms (running vs. run). This makes it easier for NLP models to learn.

e.g: The Quick! bRoWn foX juMPS.

Got It???

Let's Goooooo!!!

Step 1: Text Preprocessing

We prepare text by:

1. Lowercasing
2. Removing Punctuation Marks
3. Removing Stopwords (i.e words like A, of, in, etc)
4. Stemming / Lemmatization

e.g: The Quick! bRoWn foX juMPS.

NB:
There is no specific way to prepare text, it is much dependent on what type of NLP project we are building!

Step 1: Text Preprocessing

1. Lowercasing: To humans, A capitalized word at the beginning of a sentence (e.g., She) has the same meaning as when it's used later in a sentence (she).

By converting all characters in a corpus to lowercase, we disregard any use of capitalization.

Computers treat "The" and "the" as different words. Lowercasing makes everything consistent.

The Quick! bRoWn foX juMPS.

the quick! brown fox jumps.

NB:
Corpus in this regard means collection of writings.

Step 1: Text Preprocessing

Though, in a larger corpus that has many more examples of individual uses of words, the word, **general (an adjective meaning "widespread")** have different meaning to the word **General (a noun meaning the commander of an army).** One option is not to lowercase everything, or try an option called **Part-Of-Speech (POS) Tagging**

The Quick! bRoWn foX juMPS.

the quick! brown fox jumps.

NB:
Corpus in this regard means collection of writings.

Step 1: Text Preprocessing

2. Removing Punctuation Marks: Punctuation marks generally don't add much value to a natural language model and so are **often** removed.

Removing punctuation would not be an advantage in all cases. Consider, for example, if you were building a question-answering algorithm, which could use question marks to help it identify questions.

The Quick! bRoWn foX juMPS.

the quick brown fox jumps

Depending on the NLP project, some steps are compulsory

Step 1: Text Preprocessing

3. Removing Stopwords: Stopwords are frequently occurring words that tend to contain relatively little distinctive meaning, such as **the**, **at**, **which**, **and**, **of** etc.

There is no universal consensus on the precise list of stop words, but depending on your application it may be sensible to ensure that certain words are (or aren't!) considered to be stop words.

The Quick! bRoWn foX juMPS.

quick brown fox jumps

Depending on the NLP project, some steps are compulsory

Step 1: Text Preprocessing

For example, when building a model to **classify movie reviews** as **positive** or **negative**. Some lists of stop words include negations like **didn't, isn't, and wouldn't** that might be critical for our model to identify the sentiment of a movie review, so these words probably shouldn't be removed.

Negations may be helpful as stop words for some classifiers but probably not for a sentiment classifier

So be careful with this one. In many instances, it will be best to remove only a limited number of stop words.

The Quick! bRoWn foX juMPS.

quick brown fox jumps

Depending on the NLP project, some steps are compulsory

Step 1: Text Preprocessing

4. Stemming / Lemmatization:

Stemming: Stemming is the truncation of words down to their stem, **i.e their root form.**

For example, the words **house**, **housed** and **housing** both have the stem **hous**.

With smaller datasets in particular, stemming can be productive because it pools words with similar meanings into a single token.

The Quick! bRoWn foX juMPS.

quick brown fox jumps

Depending on the NLP project, some steps are compulsory

Step 1: Text Preprocessing

There will be more examples of this stemmed token's context, enabling techniques like **word2vec** or **GloVe** to more accurately identify an appropriate location for the token in word-vector space

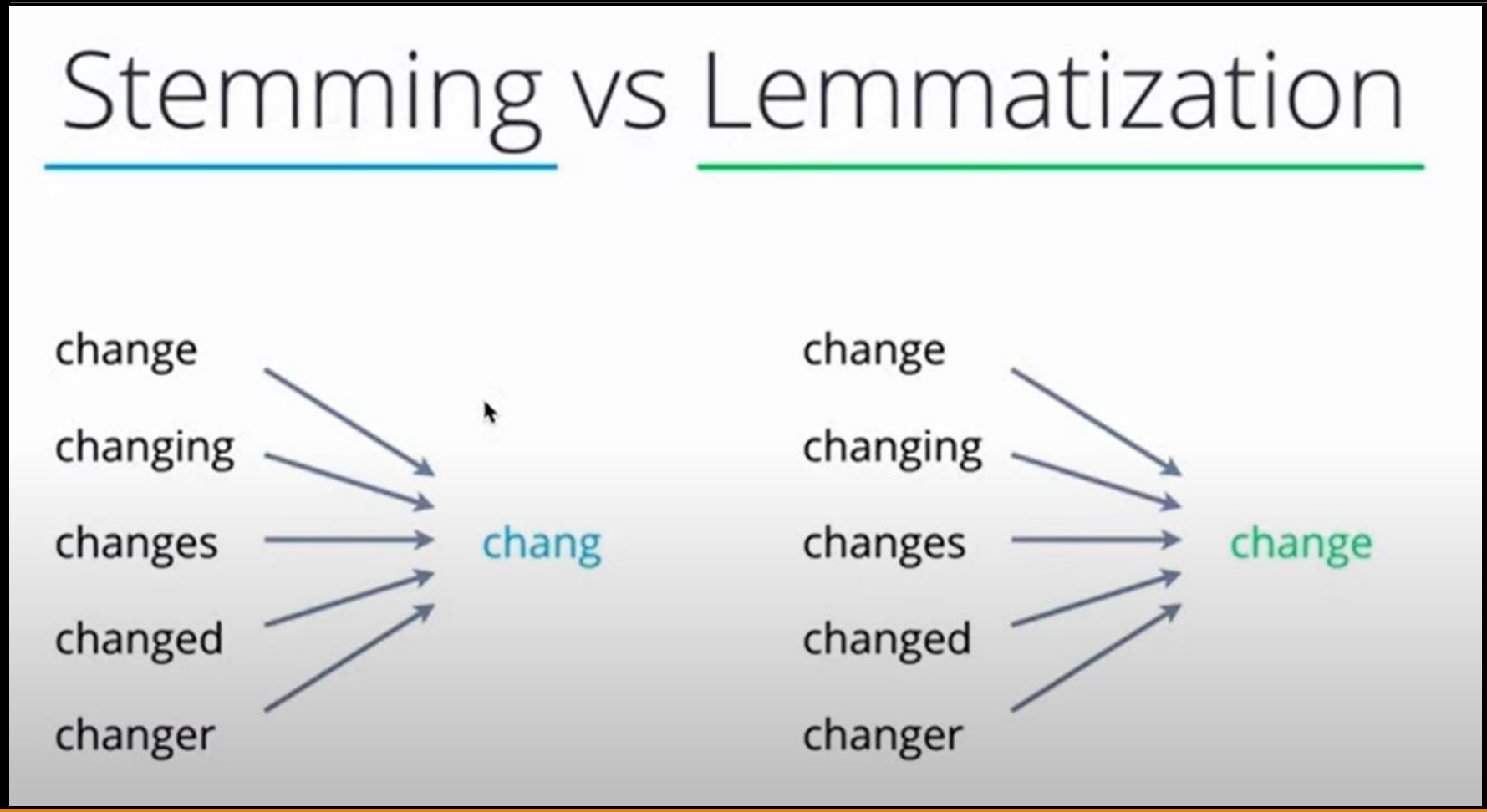
To stem words, you can use the Porter algorithm5 provided by **nltk**. To do this, you create an instance of a **PorterStemmer()** object and then add its stem() method

The Quick! bRoWn foX juMPS.

quick brown fox jumps

NLTK == Python Natural Language Toolkit. It's a package

Step 1: Text Preprocessing



From this comparison, Stemming returns to its root form. Stemming is quite harsh

While

Lemmatization is quite flexible and return it to a common word.

Choose based on the kind of words you have and your project.

Step 1: Text Preprocessing

🧐 Difference Between Stemming and Lemmatization

Both **stemming and lemmatization** reduce words to their **root form**, but they work differently!

Feature	Stemming 🪓	Lemmatization 📖
How It Works	Chops off word endings based on rules (heuristics).	Uses a dictionary to find the actual base form of a word.
Speed	Fast but less accurate.	Slower but more accurate.
Output	May not return a real English word.	Always returns a real word.
Use Case	Search engines, keyword extraction.	NLP tasks requiring correct meaning, like chatbots and sentiment analysis.
Example 1	"running" → "run"	"running" → "run"
Example 2	"happiness" → "happi" ❌	"happiness" → "happiness" ✅
Example 3	"better" → "better" ❌	"better" → "good" ✅

Stemming follows simple rules to cut off suffixes like -ing, -ed, or -ly.

Lemmatization uses linguistic knowledge to find the true root form of a word.

Step 1: Text Preprocessing

0	I absolutely LOVE this product!! ❤️ It's super efficient and really worth the money. Definitely recommend! 👍	i absolutely love this product!! ❤️ it's super efficient and really worth the money. definitely recommend! 👍	i absolutely love this product its super efficient and really worth the money definitely recommend	absolutely love product super efficient really worth money definitely recommend	absolut love product super effici realli worth money definit recommend	absolutely love product super efficient really worth money definitely recommend
1	Worst purchase ever... 😡 Waste of money. DO NOT BUY!! Full of issues.	worst purchase ever... 😡 waste of money. do not buy!! full of issues.	worst purchase ever waste of money do not buy full of issues	worst purchase ever waste money buy full issues	worst purchas ever wast money buy full issu	worst purchase ever waste money buy full issue
2	This product does what it says, but nothing special. 🤷 It's okay for the price, I guess.	this product does what it says, but nothing special. 🤷 it's okay for the price, i guess.	this product does what it says but nothing special its okay for the price i guess	product says nothing special okay price guess	product say noth special okay price guess	product say nothing special okay price guess
3	AMAZING quality and fast shipping!!! 🚀🔥 #satisfied #fastdelivery	amazing quality and fast shipping!!! 🚀🔥 #satisfied #fastdelivery	amazing quality and fast shipping satisfied fastdelivery	amazing quality fast shipping satisfied fastdelivery	amaz qualiti fast ship satisfi fastdeliveri	amazing quality fast shipping satisfied fastdelivery
4	Terrible! Had high expectations, but it broke in a week. Really disappointed. 😞	terrible! had high expectations, but it broke in a week. really disappointed. 😞	terrible had high expectations but it broke in a week really disappointed	terrible high expectations broke week really disappointed	terribl high expect broke week realli disappoint	terrible high expectation broke week really disappointed
5	This phone is great 📱, but the battery drains too fast. 🔋😕	this phone is great 📱, but the battery drains too fast. 🔋😕	this phone is great but the battery drains too fast	phone great battery drains fast	phone great batteri drain fast	phone great battery drain fast
6	I love how easy it is to use! 😍 Definitely a game-changer.	i love how easy it is to use! 😍 definitely a game-changer.	i love how easy it is to use definitely a gamechanger	love easy use definitely gamechanger	love easi use definit gamechang	love easy use definitely gamechanger
7	Do not buy this laptop! 👎 It crashes every 10 minutes. So frustrating! 😡	do not buy this laptop! 👎 it crashes every 10 minutes. so frustrating! 😡	do not buy this laptop it crashes every 10 minutes so frustrating	buy laptop crashes every 10 minutes frustrating	buy laptop crash everi 10 minut frustrat	buy laptop crash every 10 minute frustrating
8	The camera quality is excellent! 📸 Love the night mode. 🌙✨	the camera quality is excellent! 📸 love the night mode. 🌙✨	the camera quality is excellent love the night mode	camera quality excellent love night mode	camera qualiti excel love night mode	camera quality excellent love night mode
9	Meh... the product is just average. 😐 I expected more for this price.	meh... the product is just average. 😐 i expected more for this price.	meh the product is just average i expected more for this price	meh product average expected price	meh product averag expect price	meh product average expected price
10	Great customer service! 🙌 They replaced my faulty item within 24 hours.	great customer service! 🙌 they replaced my faulty item within 24 hours.	great customer service they replaced my faulty item within 24 hours	great customer service replaced faulty item within 24 hours	great custom servic replac faulti item within 24 hour	great customer service replaced faulty item within 24 hour

Love it Right????, I love it too

Some Discussion

Text -> Turn into numbers -> build a model -> train the model to find patterns -> use patterns (make predictions)

That's the flow

📌 Everything in a tip

You've successfully completed text preprocessing on customer feedback data. We've covered:

- ✔ Lowercasing – Ensuring consistency in text.
- ✔ Removing Punctuation & Emojis – Cleaning unnecessary characters.
- ✔ Stopword Removal – Keeping only meaningful words.
- ✔ Stemming & Lemmatization – Reducing words to their root form for better NLP.

🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥🔥

Let's move on

Text -> turn into numbers -> build a model -> train the model to find patterns -> use patterns (make predictions)

If you understand everything up to this point, identify yourself as a Machine Learning Engineer specializing in Natural Language Processing!

😀 😀 😀 😀

QUESTIONS AND ANSWER

The END Day 1

DSN LEKKI-AJAH
BY ABEREJO HABEEBLAH O.
X: @HABEREJO

Next Class schedule:
Tuesday, 11th March, by 5:00pm.

🧠 Time for quick Brain Teaser🔍

We've cleaned, transformed, and optimized text like real NLP pros! Now, let's see how well you followed along.🧐

Question: Given this raw text:👉

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

If we apply the following preprocessing steps, what will be the output at each stage?

1️⃣ Lowercasing: What does the text look like after converting to lowercase?

🧠 Time for quick Brain Teaser🔍

1️⃣ Lowercasing: What does the text look like after converting to lowercase?

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

✅ "i'm loving this nlp session!!! 😍🔥 it's amazing how we can clean text properly. but... is it always the best approach?"

🧠 Time for quick Brain Teaser🔍

Question: Given this raw text:👉

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

If we apply the following preprocessing steps, what will be the output at each stage?

2️⃣ Removing Punctuation & Emojis: What’s left after getting rid of #,!, 😍🔥 , and other punctuation marks?

🧠 Time for quick Brain Teaser🔍

2️⃣ Removing Punctuation & Emojis: What's left after getting rid of #,!, 😍🔥, and other punctuation marks?

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?"

✅ "im loving this nlp session its amazing how we can clean text properly but is it always the best approach"

🧠 Time for quick Brain Teaser🔍

Question: Given this raw text:👉

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

If we apply the following preprocessing steps, what will be the output at each stage?

3️⃣ Stopword Removal: Which words will be removed, and what remains?

🧠 Time for quick Brain Teaser🔍

3️⃣ Stopword Removal: Which words will be removed, and what remains?

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

✅ "loving nlp session amazing clean text properly always best approach"

🧠 Time for quick Brain Teaser🔍

Question: Given this raw text:👉

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

If we apply the following preprocessing steps, what will be the output at each stage?

4️⃣ Stemming & Lemmatization: If we use:

Porter Stemmer, how would words like "loving" and "amazing" change?

Lemmatization, what would be the base form of "loving" and "clean"?

🧠 Time for quick Brain Teaser🔍

4️⃣ Stemming & Lemmatization: If we use:

Porter Stemmer, how would words like "loving" and "amazing" change?

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

loving" → "love“
"amazing" → "amaz“
"clean" → "clean" (remains the same)
✅ "love nlp session amaz clean text proper always best approach"

🧠 Time for quick Brain Teaser🔍

4️⃣ Stemming & Lemmatization: If we use: Lemmatization (Using WordNet Lemmatizer), what would be the base form of "loving" and "clean"?

"I'm LOVING this NLP session!!! 😍🔥 It's AMAZING how we can clean text properly. But... is it always the best approach?“

"loving" → "love“
"amazing" → "amazing" (unchanged)
"clean" → "clean“
✅ "love nlp session amazing clean text properly always best approach"

🧠 🚀 Bonus Question: 🔍

We removed **stopwords** to clean up our text. But can you think of a situation where removing **stopwords** might actually hurt the meaning or usefulness of our data? 🤔

What can be done in such case???

💡 Hint: Think about cases where small words like "not," "is," or "to" are important in understanding the sentence.

Open-ended answers are welcomed!!!

🧠 🚀 Bonus Question: 🔍

Answer: ✅ Yes! There are several cases where stopwords play a crucial role:

1️⃣ **Sentiment Analysis** – Words like "not" and "very" can change sentiment.
Example: "The movie is not bad" (removing "not" changes the meaning).

2️⃣ **Question Answering –** Stopwords help retain context in queries.
Example: "What is the capital of France?" makes more sense than just "capital France?''

3️⃣ **Machine Translation** – Removing stopwords might lead to incorrect translations.
Example: Translating "I am going to the market" without "am" and "to" could alter the meaning.

So, while removing stopwords often helps clean data, it's important to consider the task before blindly dropping them! 🚀

🧠 Try This

💡 Try This! Modify the dataset with your own text and observe how preprocessing affects different types of input.

Generate into a CSV File and share before the end of the program.

Let's keep building! 🔥🔥🔥

Let's move on

Text -> turn into numbers -> build a model -> train the model to find patterns -> use patterns (make predictions)

If you understand everything up to this point, identify yourself as a Machine Learning Engineer specializing in Natural Language Processing!

😀 😀 😀 😀

The END Day 1

DSN LEKKI-AJAH
BY ABEREJO HABEEBLAH O.
X: @HABEREJO

Next Class schedule:
Tuesday, 11th March, by 5:00pm.