

→ Why? ⇒ LLM is also Text Handling

→ What? ⇒ Understanding Text

↳ ML ~ Logistic Reg, KNN, Naive Bayes

Random Forest

→ Numeric Data ~ Nominal & Ordinal

→ Lexical Analysis

LLM: 40B tokens

→ Sentences, Tokens, Corpus

→ How? ⇒ 1) Problems

2) Handling the text data (Regex)

3) Probabilistic Models
(Understanding Grammar)

→ Issues w/ Text:

→ Lexical Ambiguity:

Lexical - Text

Ambiguity - Multiple

meanings.

... : 1111 1111 .11 42112 4211

Ambiguity - multiple meanings.

eg: Will Will will Will's Will

Modals Name Verb Name Noun

Verb of Person of Person

eg: Buffalo Bully

Animal City in US

eg: Rose rose to put rose roses on her rows of roses

Noun Verb Adj. Thinness Alignment Flower

Sea-food

→ Structural Ambiguity:

→ The man saw the boy w/ the binoculars.

→ Flying planes can be dangerous

→ Language Imprecision & Vagueness:

Qualitative

V U I

→ It is wrong here.

Qualitative

→ Hospitals are sued by 7 foot doctors.

Quantitative

→ Italian Printing found by tree.

→ King ~ Spanish

Como se llame?

What is your name?

¿Cual tu es nombre?

→ Pervasive Ambiguity:

→ I made her duck.

→ I called duck for her.

→ I asked duck belonging to her.

→ I used my wand to make her a duck.

→ I created a duck she runs from the egg.

→ I made her duck during shooting.

→ Idioms:

... in court ...

- Ball in your court -
- Burning the midnight oil.
- Neologisms: (Slangs)
 - Retweet.
 - Google (Signal) X
- New Sense of words
 - ↳ That's sick dude!