Designing intelligent conversation : A Chatbot

One Day WORKSHOP

Prof. S. A. Desai, ET Dept., YCCE
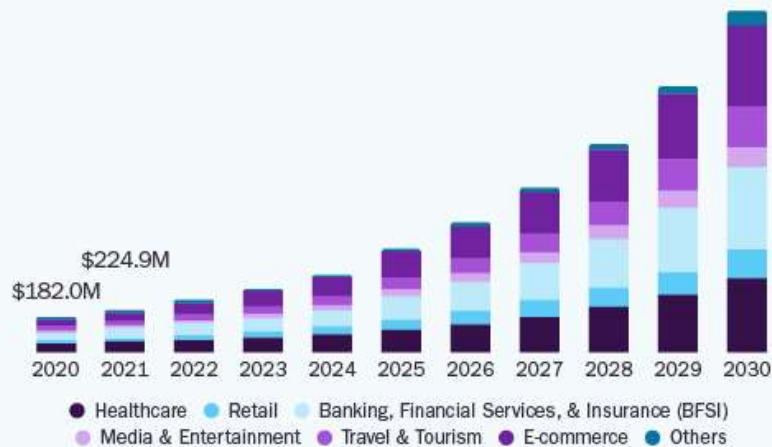
Date: 19 Nov 2022

Global Chatbot market

In 2021 – market size was valued at USD 525.7 million.

From 2022 to 2030 – expected to expand at a compound annual growth rate (CAGR) of 25.7%.

North America Chatbot Market
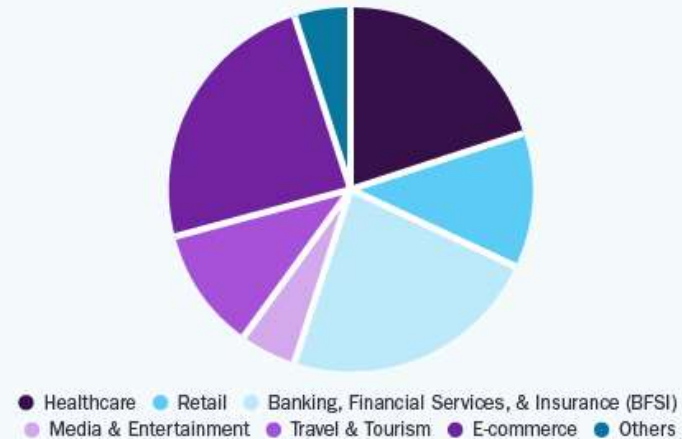size, by vertical, 2020 - 2030 (USD Million)

GRAND VIEW RESEARCH

$224.9M
$182.0M

2020 2021 2022 2023 2024 2025 2026 2027 2028 2029 2030

● Healthcare ● Retail ● Banking, Financial Services, & Insurance (BFSI)
● Media & Entertainment ● Travel & Tourism ● E-commerce ● Others

26.9%
U.S. Market CAGR,
2022 - 2028

Source:
www.grandviewresearch.com

Global Chatbot Market
share, by vertical, 2021 (%)

GRAND VIEW RESEARCH

● Healthcare ● Retail ● Banking, Financial Services, & Insurance (BFSI)
● Media & Entertainment ● Travel & Tourism ● E-commerce ● Others

$525.7M
Global Market Size,
2021

Source:
www.grandviewresearch.com

Industries working in this domain

01 Coforge

02 TATA / TATA ELXSI

03 Persistent

04 happiest minds — The Mindful IT Company

Industries working in this domain

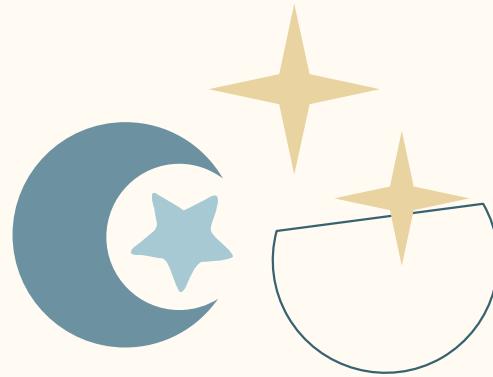|	
Haptik	Maruti Techlabs
Matellio	Quytech
Yugasa software Labs	Trigma
Hidden Brains	Teplar Solutions
Squareboat	Talentica

Job Descriptions

- Chatbot Developer
- Technical lead
- Python Developer
- Testing Engineering
- Data Scientist
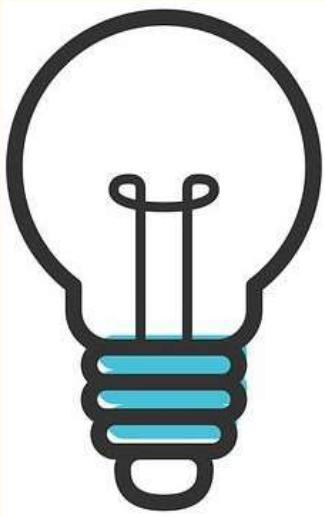
Salary : ₹ 2.5 to ₹ 14.0 Lakhs

And we humans have our own languages ...called Natural Language

If we have to ask questions to Computer . . like

- What is the next train to Delhi after 9pm?

- What genes are associated with Diabetes?

- What are people liking about our new product?

- How much time I have before my next meeting?

We have to convert it to Computer's language

Wouldn't it be nice if computers can also use natural language . . .

That's the goal of NLP

Natural Language Processing (NLP)

Helps computers to communicate with humans in their own language.

Why is NLP hard?

He approached many <u>banks</u> for the loan.

Delhi is situated on the <u>banks</u> of Yamuna river.

So how do we build logic to do NLP?

- It's very hard for us humans to build the rules . . (to code)

- Hence, we commonly use <u>Machine Learning</u> for NLP

NLP presents a huge opportunity …

as most organization have humongous
amount of textual data but struggling to get
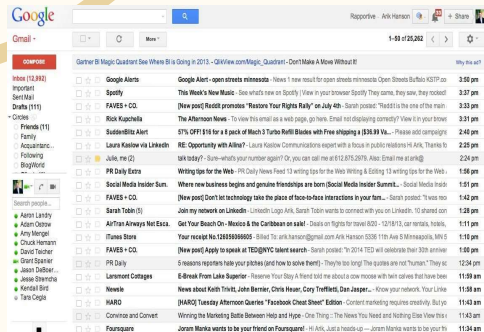value out of it.

Healthcare

Banking & Insurance

Fashion

Manufacturing

Travel

Social Media

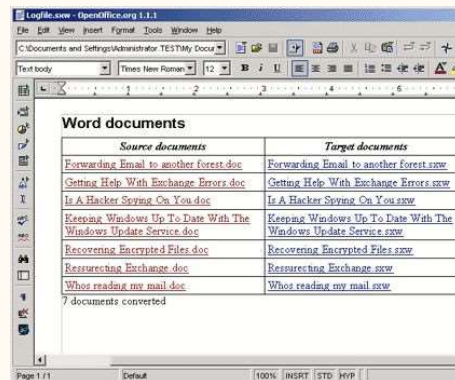Retail

NLP is useful for every industry …

Emails

Social Media Updates

Chat interactions

Office documents

podcast

...other audio data

Some examples of Language Data

Applications of NLP

Google Home, Alexa etc

Digital assistants

Applications of NLP

Help us write better

Improved communication

Applications of NLP
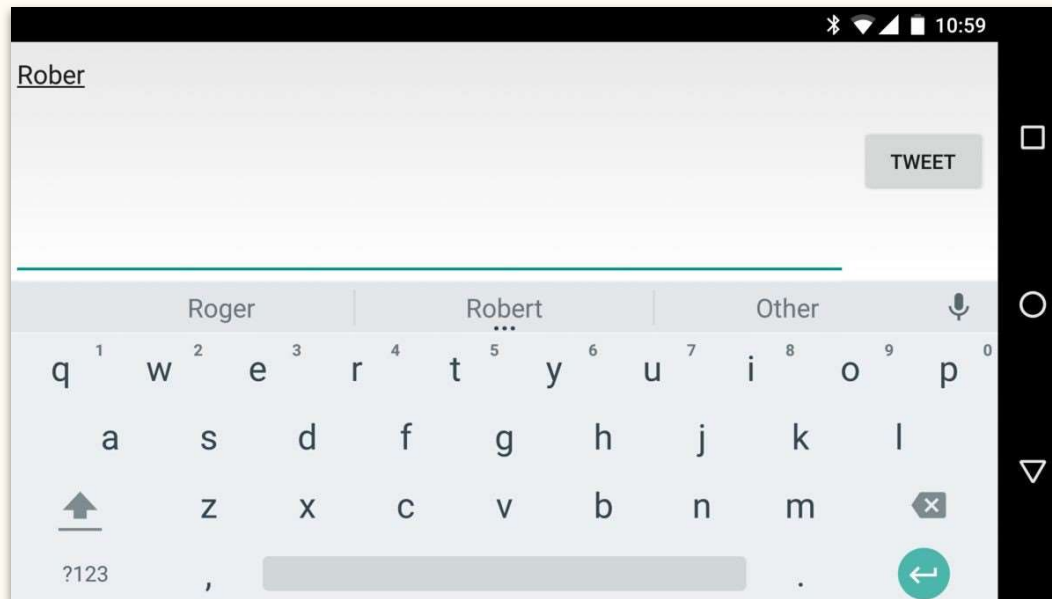

Tendulkar playing cricket

Describe a picture
(Image Captioning)

Applications of NLP

Mobile phone keyboard

Predict next word(s) as we type in...

Applications of NLP

Search suggestions

Applications of NLP

Chatbots

Make support available in 24x7 format

Applications of NLP

Sentiment analysis
Do people like our products or Services?

Building NLP
Models

	CRIM	ZN	INDUS	CHAS	NOX	RM	AGE	DIS	RAD	TAX	PTRATIO	B	LSTAT
0	0.00632	18.0	2.31	0.0	0.538	6.575	65.2	4.0900	1.0	296.0	15.3	396.90	4.98
1	0.02731	0.0	7.07	0.0	0.469	6.421	78.9	4.9671	2.0	242.0	17.8	396.90	9.14
2	0.02729	0.0	7.07	0.0	0.469	7.185	61.1	4.9671	2.0	242.0	17.8	392.83	4.03
3	0.03237	0.0	2.18	0.0	0.458	6.998	45.8	6.0622	3.0	222.0	18.7	394.63	2.94
4	0.06905	0.0	2.18	0.0	0.458	7.147	54.2	6.0622	3.0	222.0	18.7	396.90	5.33
5	0.02985	0.0	2.18	0.0	0.458	6.430	58.7	6.0622	3.0	222.0	18.7	394.12	5.21
6	0.08829	12.5	7.87	0.0	0.524	6.012	66.6	5.5605	5.0	311.0	15.2	395.60	12.43
7	0.14455	12.5	7.87	0.0	0.524	6.172	96.1	5.9505	5.0	311.0	15.2	396.90	19.15
8	0.21124	12.5	7.87	0.0	0.524	5.631	100.0	6.0821	5.0	311.0	15.2	386.63	29.93
9	0.17004	12.5	7.87	0.0	0.524	6.004	85.9	6.5921	5.0	311.0	15.2	386.71	17.10

Boston Housing Price dataset

Structured Data

1. has features (columns)
2. All records have same features
3. Features maintain order across examples

Hunger Games (2012)

4 of 4 people found the following review helpful

★★★★★ **Recent Convert**, September 18, 2012

By **dharmadude** - See all my reviews

This review is from: Barefoot Running - The Movie: How to Run Light and Free by Getting in Touch with the Earth (NTSC/US Version) (DVD)

As yet, have not viewed film in its entirety but was moved to post a brief note, because thus far, I absolutely LOVE this DVD!! The videography is stunning, the setting on Maui is gorgeous (of course) and the material is very well presented. As a relative new-comer to this barefoot running "thing", I was hoping for some solid, fundamental instruction, as well as inspiration to continue on my fitness path.
I was not disappointed. Having tired of the ever-present aches, sprains & other maladies associated with "normal" distance running, my ethusiasm for running has only recently returned, thanks to the barefoot approach. (sometimes "cheat" with miniamlist shoes) As a result of watching a good portion of this eloquently produced film, I am now fully convinced that I will be a barefoot runner for the duration. Was also quite impressed by the authors, who are a husband & wife team, I think. They exude a truly genuine quality & are clearly passionate about the work they are doing. Not to mention that they appear to be in excellent condition. Guess they practice what they preach.
Now excuse me while I get back to watching the video.

★☆☆☆☆

By Ryan Galaska on February 16, 2016

Verified Purchase

Hunger isn't a game.

0 of 1 people found this review helpful

Textual Data

1. Do not have features like in tabular data.
2. Examples usually have different size.

What could be the features of textual Data?

4 of 4 people found the following review helpful

★★★★★ **Recent Convert**, September 18, 2012

By **dharmadude** - See all my reviews

This review is from: Barefoot Running - The Movie: How to Run Light and Free by Getting in Touch with the Earth (NTSC/US Version) (DVD)

As yet, have not viewed film in its entirety but was moved to post a brief note, because thus far, I absolutely LOVE this DVD!! The videography is stunning, the setting on Maui is gorgeous (of course) and the material is very well presented. As a relative new-comer to this barefoot running "thing", I was hoping for some solid, fundamental instruction, as well as inspiration to continue on my fitness path. I was not disappointed. Having tired of the ever-present aches, sprains & other maladies associated with "normal" distance running, my ethusiasm for running has only recently returned, thanks to the barefoot approach. (sometimes "cheat" with miniamlist shoes) As a result of watching a good portion of this eloquently produced film, I am now fully convinced that I will be a barefoot runner for the duration. Was also quite impressed by the authors, who are a husband & wife team, I think. They exude a truly genuine quality & are clearly passionate about the work they are doing. Not to mention that they appear to be in excellent condition. Guess they practice what they preach.
Now excuse me while I get back to watching the video.

Features in textual Data

1. Words?

2. Characters?

3. Combination of words (n-grams)?

4. Sentences?

How do we find features of text data?

1. Breakdown text into individual tokens (e.g words)
2. Find the unique tokens
3. Unique tokens form the vocabulary of our dataset

Document #1

He is a good boy. She is also good.

"He", "is", "a", "good", "boy". "She", "is" "also" "good"
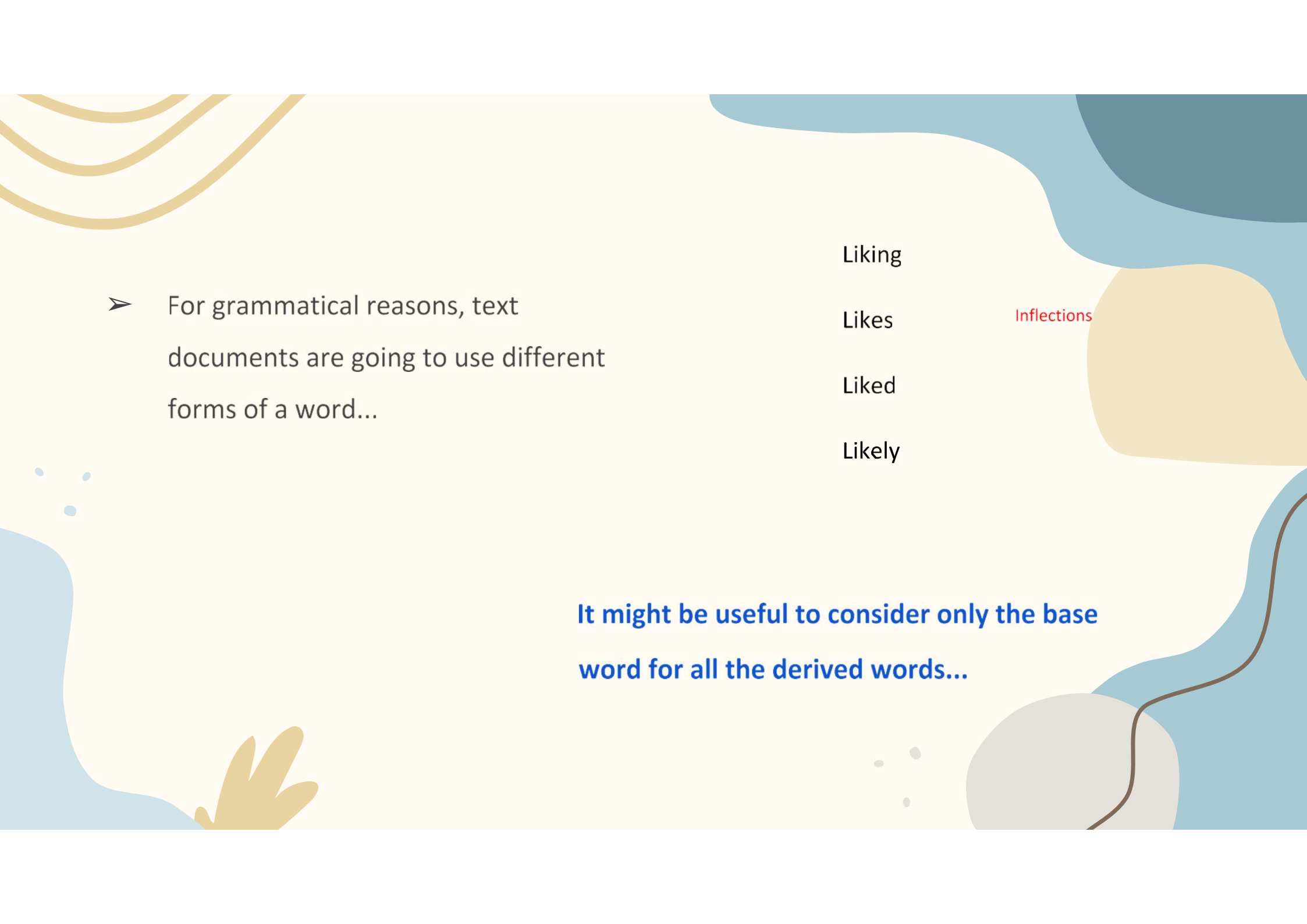
Document #2

Radhika is a good person.

"Radhika" "is" "a" "good" "person".

Tokenization

Breaking text into tokens e.g words, chars etc

➢ For grammatical reasons, text documents are going to use different forms of a word...
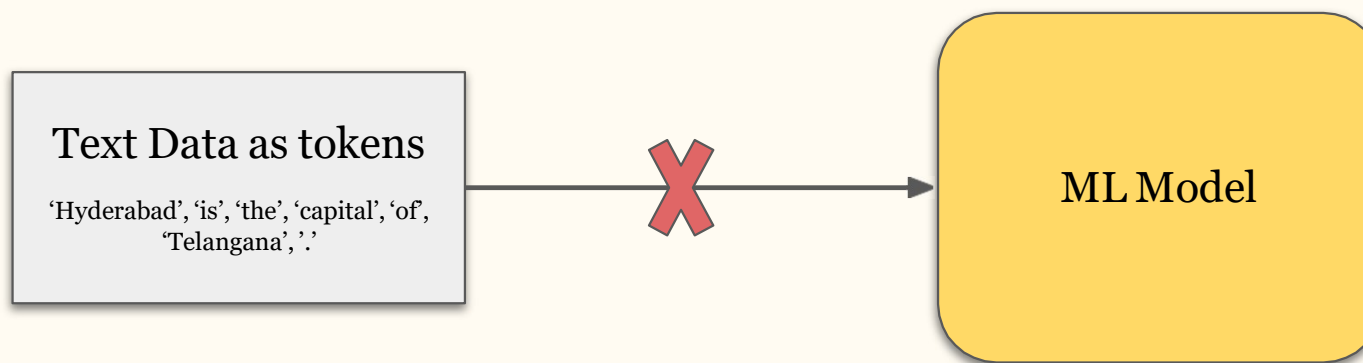
Liking

Likes

Liked

Likely

Inflections

It might be useful to consider only the base word for all the derived words...
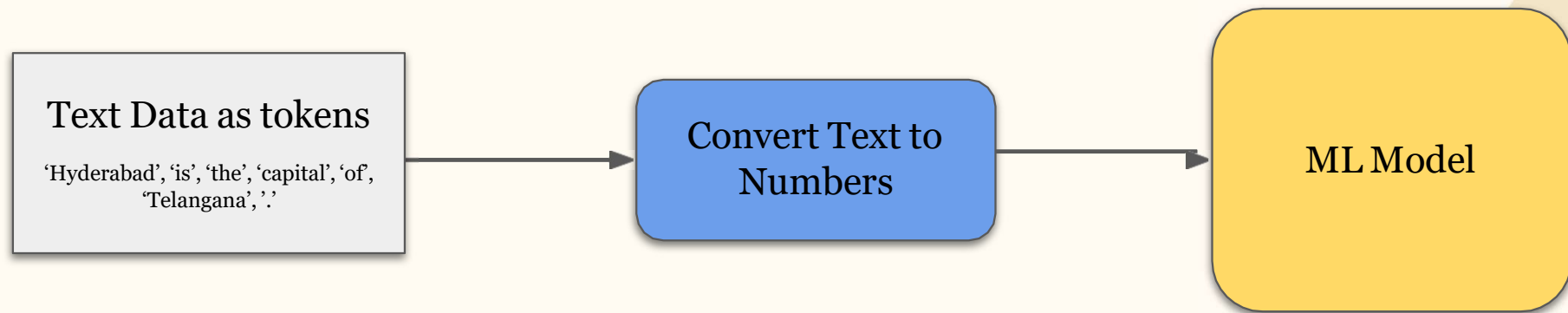
How to change derived words into Base or **Root** word?

Two popular approaches …

1. Lemmatization
2. Stemming

Text Data as tokens

'Hyderabad', 'is', 'the', 'capital', 'of', 'Telangana', '.'

ML Model

Math works with numbers

Text Data as tokens

'Hyderabad', 'is', 'the', 'capital', 'of', 'Telangana', '.'

Convert Text to Numbers

ML Model

Vectorization

Converting each token into number(s)

How to convert individual documents into numbers using index and vocabulary?

We have multiple approaches …

1. Bag of words
2. CountVectorizer
3. TF-IDF Vectorizer

Bag of Words

Words

1. A simple feature extraction approach in NLP
2. Ignores grammar / structure
3. Represents each document by measuring presence of vocabulary words

Document #1

He is a good boy. She is also good.

Document #2

Radhika is a good person.

Vocabulary

a, also, boy, good, He, is, person, She, Radhika

	a	also	boy	good	He	Is	person	She	Radhika
Index	0	1	2	3	4	5	6	7	8
Document# 1									

Check if each word in Vocabulary appears in Document #1

Document #1

He is a good boy. She is also good.

Document #2

Radhika is a good person.

Vocabulary

a, also, boy, good, He, is, person, She, Radhika

	a	also	boy	good	He	Is	person	She	Radhika
Index	0	1	2	3	4	5	6	7	8
Document #1	1	1	1	1	1	1	0	1	0

Indicate '1' if word was found, otherwise indicate '0'

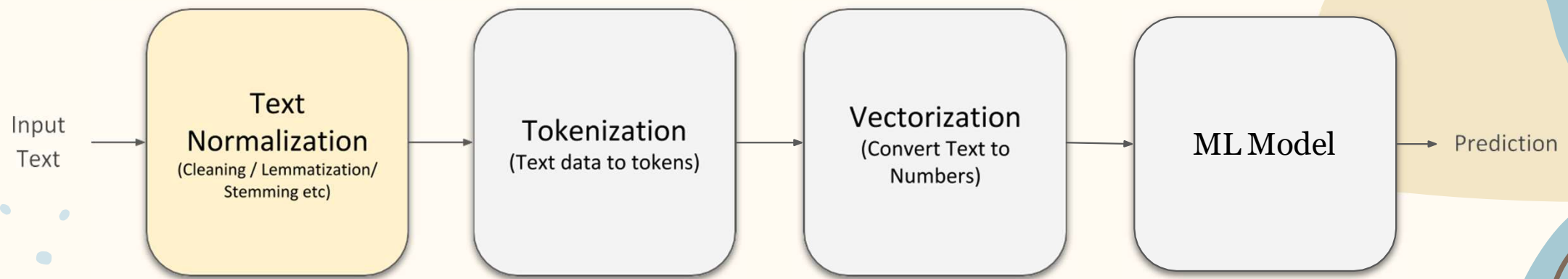Document #1

He is a good boy. She is also good.

Document #2

Radhika is a good person.

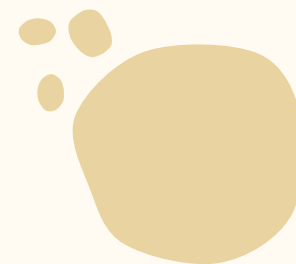Vocabulary

a, also, boy, good, He, is, person, She, Radhika

	a	also	boy	good	He	Is	person	She	Radhika
Index	0	1	2	3	4	5	6	7	8
Document #1	1	1	1	1	1	1	0	1	0
Document #2	1	0	0	1	0	0	1	0	1

Our data is now in table format and ready for ML model :)

Input Text → **Text Normalization** (Cleaning / Lemmatization/ Stemming etc) → **Tokenization** (Text data to tokens) → **Vectorization** (Convert Text to Numbers) → **ML Model** → Prediction

NLP Pipeline

Thank You