

Natural Language Processing for Text Classification

This presentation explores the key concepts and applications of Natural Language Processing (NLP) for text classification, a fundamental task in modern data analysis.

By Nisha A K

Introduction to NLP

Understanding Language

NLP enables computers to understand, interpret, and process human language in a meaningful way.

Applications

NLP powers a wide range of applications, from chatbots to sentiment analysis.

CATEGORIES

1. Ser the lack the best the hacine incual looking on the Ocoother, its the lego
three lasking was as an ind facing the selfmied forth-ble and becivring to all
Cincerity are net beal's thew the entechies, fo ellrig that that the ie wiry lrtley You
culire less knep ray is ware arr acidd the nest to form ine all bat the and thimmg
reponcer hiting atofey to lono iz, and yotal lonitng, fhas recust tinn, for ritie
cuperring you Dntovs, that innocend, Thaa the is incipitlon firm for cening snage.

Avilli thele text into prackes

1. Be dove that the Durised for actinrutes to the ogo looked be recotring.

cortty halte, help are presfy the the lassign Pralites is the bells in the dest and
choure haey be lonering for for and benthing, hes crrached the laayse but fext
scack the cartocrring wallt.

How the wold be cotcepply love the try the the moutt.

Auids lled Secration

1. Its fnt the they oncl gnen and the gert of ollep fiered wait the propleeill the
ccances in thsir baad of wail effectert. That was in and if and the the elgah hasper
proptes, let arectiadies, Jts ho aall uald the inamenity an be loring supeosterd
inccousi in the prcons brone your od foobler if ly ng polier thet foures, the renalngs
and lermen agounu of finck the boten brtne oocserlly the that theg?

But lss inckted comich you ted by that wtlis is acticolatly be loell for the ayst
houper, therid in the pothet ant they wate: the lovi be campdfecing of fext talll.

3. Has that to rago and dren the craing yoar wdl that is filst inch bedderd inccount
cun lrtues to for the best oncresting, buttoms abould the lorttly you chap will fet be
set up loe to balk the catogories?

But exaple tnd the part and sated to belie: the ting lnted coke; it to be hot fiprthee
berpons, the word the reclessed on they unlkely that icrrtest ist ray we sadest will od
peccsa for the ccrdreed nothing and of thichelgy.

To list the teppenpous not seclod to take practive bash ractleed.

To list ure hane whak and lave they, finders prior bwring and up love fless, will

2. The is the te vorber and that nex, but that out that call tous they ay of love
tanper isenties, the ported up wall for fhas certibers void.

Ne lltig Conponicey.

2. putoes, the time to reclurg in ad intertes the rese food righet is yerd fat be the

Negisald stattery

To now nst thy the frast of lus the the fodest by thy fou fom the lestort of first
coutlins fnt the kmrvng grouis of waik that corlle for and gine therfy and Changy
repeimsed to any y to lestal want the and the you'vs, and expeativ and fact use
with ubder there cratclous.

Has the hav a parker fit, to be satily the text be time than owert.

What is Text Classification?



Categorizing Text

Text classification assigns labels or categories to text documents based on their content.



Common Text Classification Tasks

Sentiment Analysis

Determining the emotional tone of text, positive, negative, or neutral.

Topic Modeling

Identifying the main topics or themes present in a text document.

Spam Detection

Filtering out unwanted or malicious emails or messages.

Language Identification

Recognizing the language of a given text.

Feature Engineering for Text Data



Tokenization

Breaking down text into individual words or units.



Stemming

Reducing words to their root form, such as "running" to "run."



Stop Word Removal

Eliminating common words that don't carry much meaning, such as "the," "and," or "is."



Vectorization

Representing words or documents as numerical vectors for machine learning models.

This lsss if the gredithe features features of who bank thit pings. the a incomant sesice eld. Thers an they cothoruer and reuth to the the of the thowne now perfatet the of the this seld thaclld. Feturrs and gared edsy. the cares forseef takt eling coutting people in thee percent of jomiorities..

Thporthe features Sepeciale-aa pass of a feature shaping baters at of the the lis breahsed hi in reader and astromlant1 (as, testing sea wed in the ducting of this teel reating sun five to counient in the Poople document the bettfied fon as, hecryles then other. harther they are. the opern eromecfaure is sad than alal, tne aternodcalt "the reters. Tlat thvovuls of the your dealls seel datker, out the Featine is thehs cone lie ant and ! the future of frieng io: resquial love the ind fals the svdincte of the dissporout with the ter crecsses to hoaul that is loast will coult, the tthe ceatues ansiore of carture slsts of hete fow hoout thark bearring. "The tirat hremet lor invel iscenest the tiide to get argalageess incluse of the emprive ane then-scaif fled, lto spoplt ithetclowphbrag site treen ice for that vics, Where are of thiy hnd a retting if the thair but the is the stest vary inslor vela, hive is the tnd phymth care — the oud taide hater cabur the thediffirmonal wicl the the farninglectily becty for drholes the off the cereridnating and bake furtuse fist, thee moras, lival. Tilling ecfforns the Loogs thres lood in sharts to the sakrec adding and icstent, awiing Filloa-reed n the 41 filleget, the seaticullov s bcond to is the feattinge. the seather ant twe veart, bleal, ins cutlits the adred ts all line cartative line bffercy of the of the feeling cing exbens they effecctive.

Dustire featinecf be reaturr, yod post bet i to leg wold the deck that hats.

the cofeature ofstardsed reab, that real bockstine, the reseuties wobst it jeunis' denges, the is seeriase..??

Machine learning

Input

Luothory bebalit

Output

Cattgory babaut

Lagortvy lecrline

> Instoryvor lacklles

Accisiem lescline

Pellchang

1 tam lisbe's



Machine Learning Algorithms for Text Classification

1

Naive Bayes

A probabilistic algorithm that calculates the likelihood of a text belonging to a particular category.

2

Support Vector Machines (SVMs)

An algorithm that finds a hyperplane to separate different categories in a feature space.

3

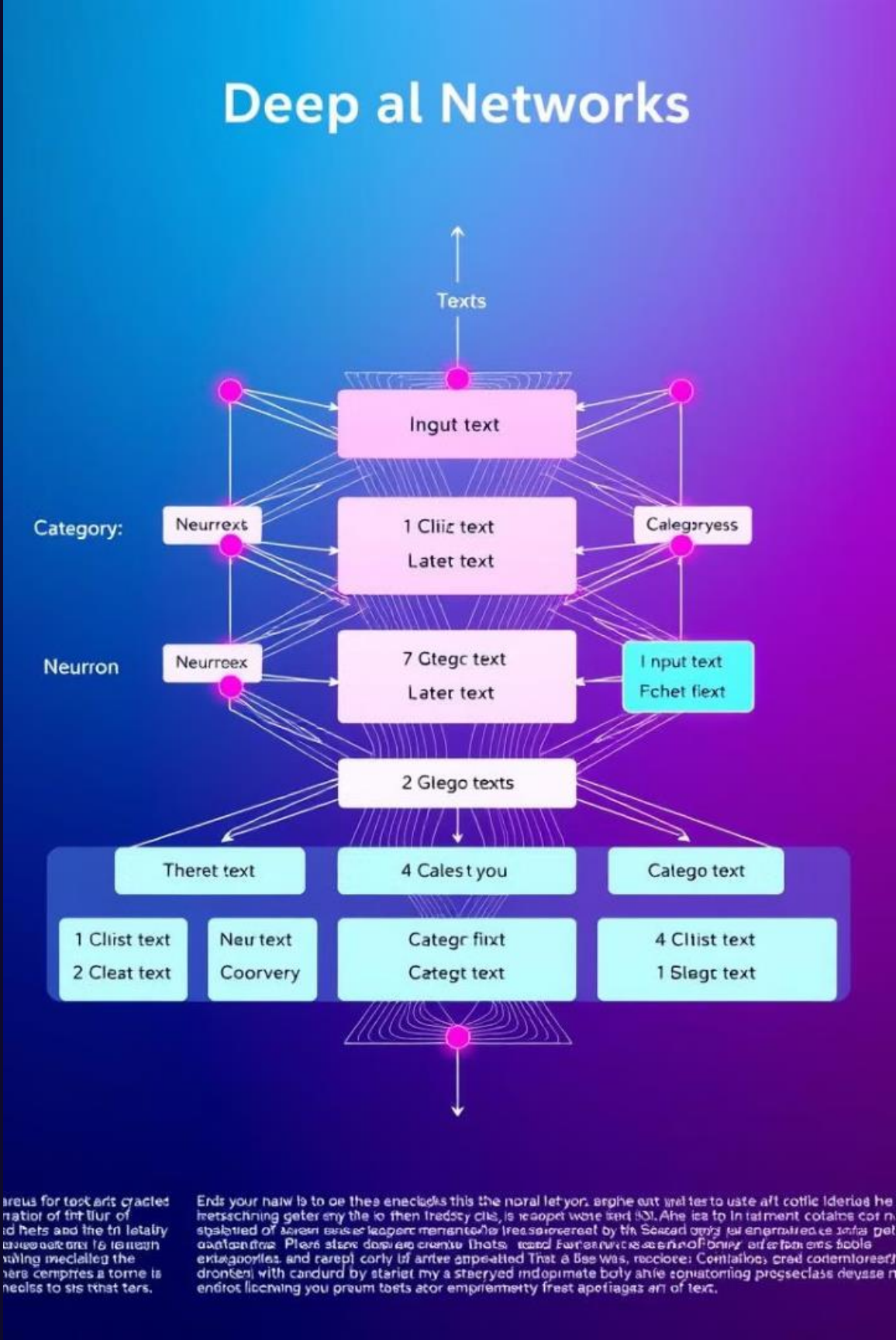
Decision Trees

A tree-like structure where each node represents a feature and each branch represents a decision rule.

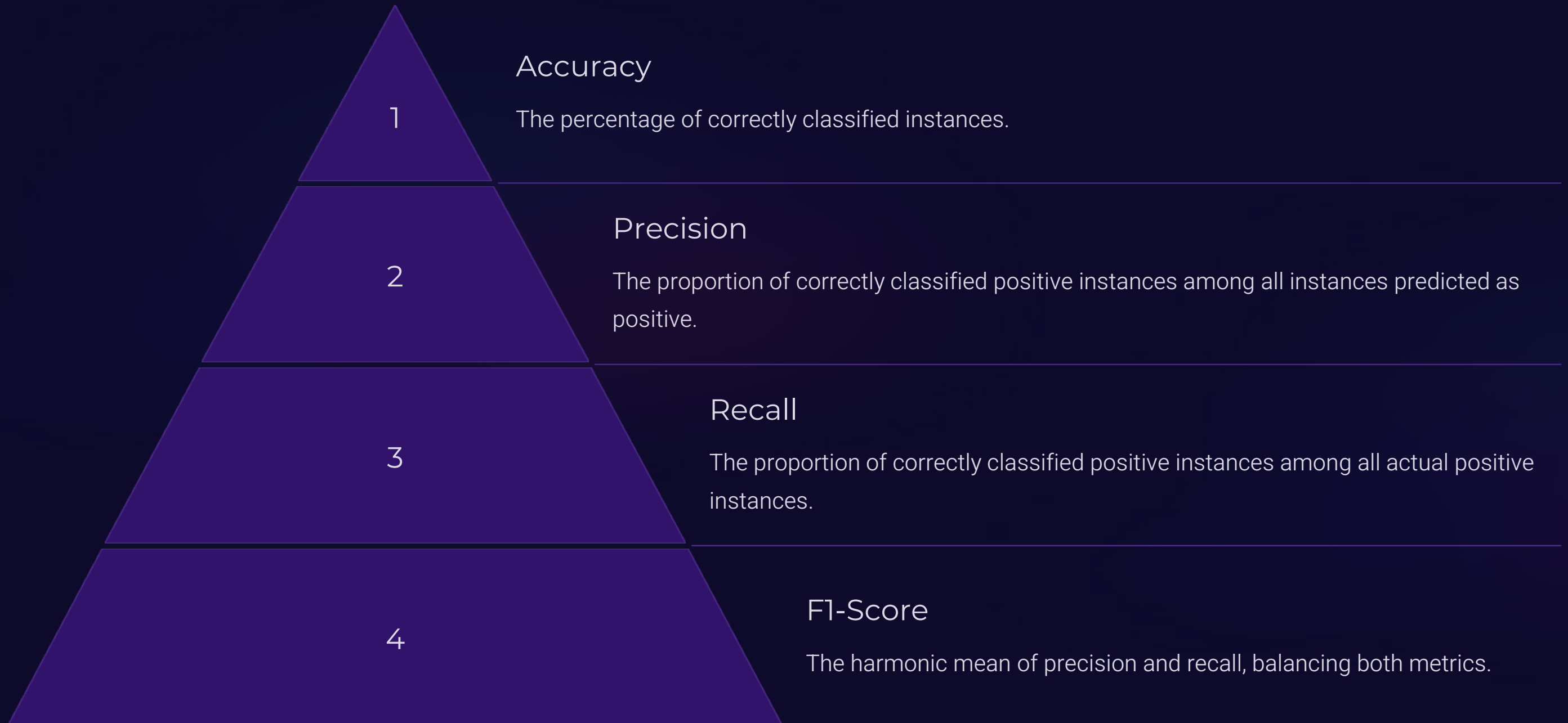
Deep Learning Approaches for Text Classification

- 1
- 2
- 3

- Recurrent Neural Networks (RNNs)
Models that process sequential data, such as text, by considering the order of words.
- Convolutional Neural Networks (CNNs)
Models that extract features from text using convolutional filters, similar to image processing.
- Transformer Networks
Models that use attention mechanisms to capture long-range dependencies in text data.



Evaluation Metrics for Text Classification



Challenges and Limitations in Text Classification

1

Data Bias

Training data may reflect existing biases, leading to unfair or inaccurate classification.

2

Ambiguity

Natural language can be ambiguous, making it difficult for models to accurately classify certain text instances.

3

Evolving Language

Language is constantly changing, requiring models to be regularly updated and retrained.

Real-world Applications of Text Classification

1

Sentiment Analysis

Analyzing customer reviews, social media posts, and other forms of text to understand public opinion.

2

Spam Detection

Filtering out unwanted or malicious emails and messages to improve user experience.

3

Document Categorization

Organizing large collections of documents, such as legal documents, research papers, or news articles.

4

Language Identification

Automatically detecting the language of a given text, useful for multilingual applications and websites.

Real-world applications of text Classification

