

Ghana Space Science & Technology Institute, GAEC

Classification in Deep Learning: A Beginner's Guide

Theophilus Ansah-Narh, Ph.D.

theophilus.ansah-narh@gaec.gov.gh

July 26, 2023





Outline

- 1 Big Data
- 2 Machine Learning
- 3 Deep Learning
- 4 Performance Metrics
- 5 DEMO
- 6 Remarks

Big Data

The Erra of Big Data I



Facets and Elements of Big Data. Image credit: Dzone website

- **Big data** is an amount of data that is enormous in volume and is constantly expanding rapidly.
- No typical data management systems can effectively store or analyze this data because of its magnitude and complexity.
- fundamental characteristics of big data are listed below

Volume

- Big Data is a vast *volume* of data generated from many sources daily, such as business processes, machines, social media platforms, networks, human interactions, and many more.
- Industry trends predict a significant increase in data volume over the next few years.
- Usually measured in gigabytes (GB), terabytes (TB), zettabytes (ZB), and yottabytes (YB)
- Nonetheless, Big data generally refers to datasets with a high volume of the order of magnitude of exabytes ($10^{18}\text{B} = 10^9\text{GB} = 10^6\text{TB} = 1\text{EB}$) and greater (Jelic *et al.* 2019).

The Erra of Big Data III

DAILY
ACTIVE USERS

1.96 Billion

DATA GENERATED
EVERY DAY

500+ TERABYTES



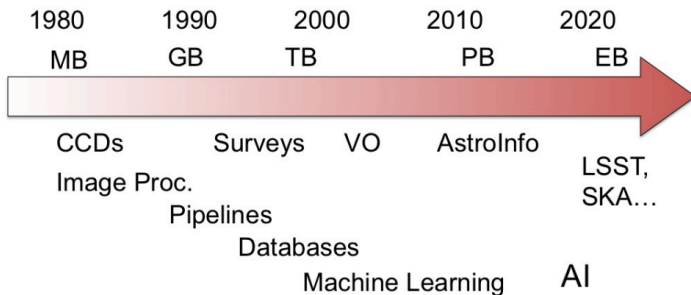
10M+ GROUPS

300M+ STORY UPDATES

Image credit: Analytics Vidhya

The Evolving Data-Rich Astronomy

An example of a “Big Data” science driven by the advances in computing/information technology



Key challenges: data heterogeneity and complexity

Image credit: Djorgovski, 2019

The Erra of Big Data V

Global data volume

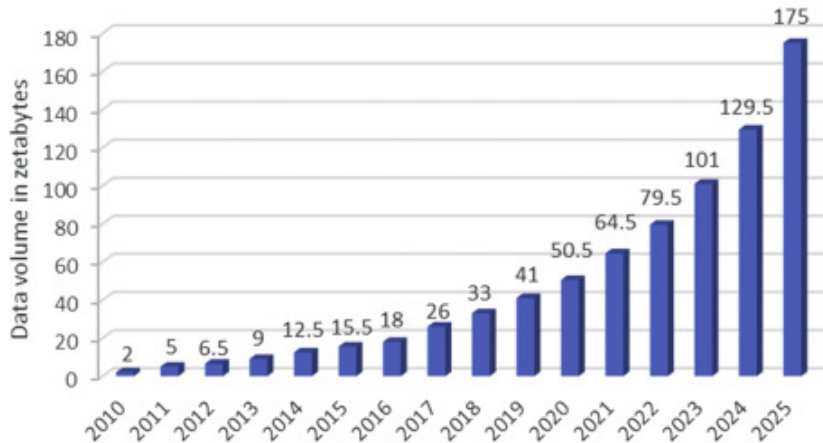


Image credit: Lei & Kong, 2020

Variety

- In the past, data is only collected from **databases** and **sheets**.
- These days the data will come in array forms, that are **PDFs**, **Emails**, **audios**, **SM posts**, **photos**, **videos**, etc.
- Big Data can be **structured**, **unstructured**, and **semi-structured** that are being collected from different sources.
 - **Structured data:** In Structured schema, along with all the required columns. It is in a tabular form. Structured Data is stored in the relational database management system.
 - **Semi-structured:** In Semi-structured, the schema is not appropriately defined, e.g., JSON, XML, CSV, TSV, and email. OLTP (Online Transaction Processing) systems are built to work with semi-structured data. It is stored in relations, i.e., tables.
 - **Unstructured Data:** All the unstructured files, log files, audio files, video files, e-mails, word processing, and image files are included in the unstructured data.

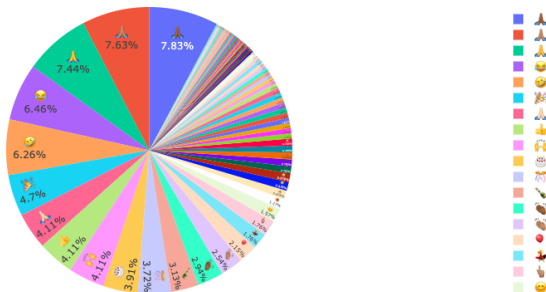
Veracity

- The accuracy of your findings can be severely harmed by poor data reliability.
- Making it one of the most crucial big data qualities
- There's a need to calibrate your data since most of the data you encounter is unstructured.

```
In [13]: print('Group wise Stats')
          print("Messages:", total_messages)
          print('Media:', media_messages)
          print('Emojis:', emojis)
          print('Links:', links)
```

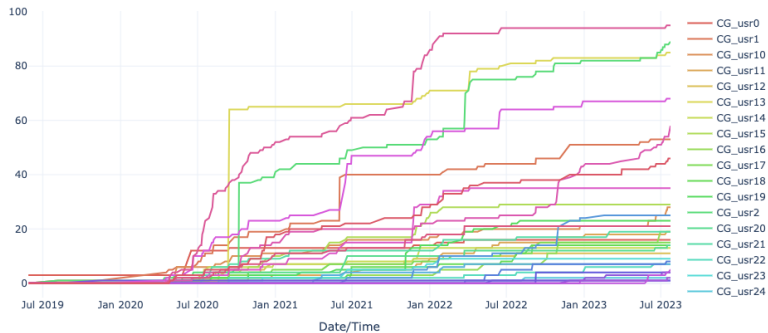
```
Group wise Stats
Messages: 845
Media: 182
Emojis: 511
Links: 188
```

Emoji Distribution



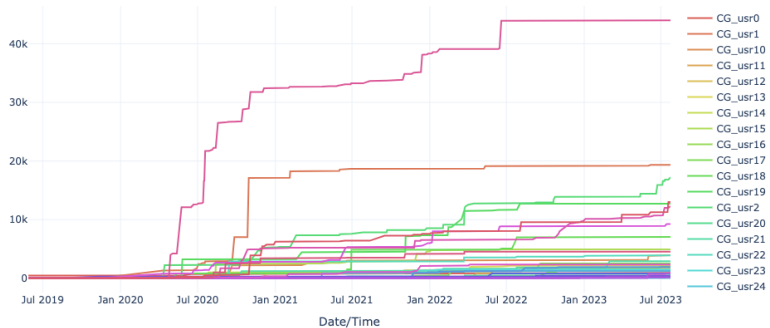
The Erra of Big Data X

User interventions count (cumulative)



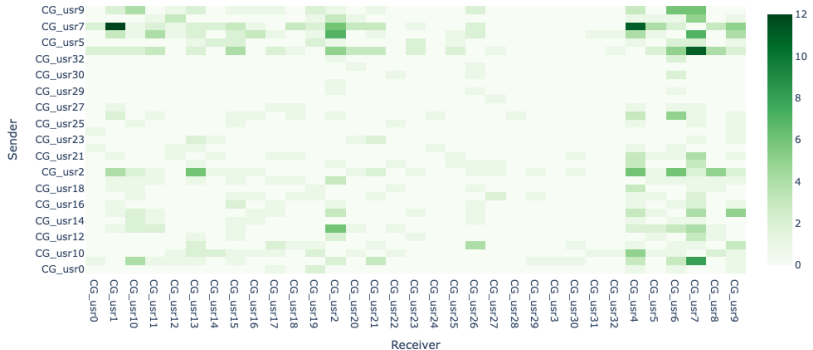
The Erra of Big Data XI

Count of sent characters (cumulative)



The Erra of Big Data XII

Response matrix



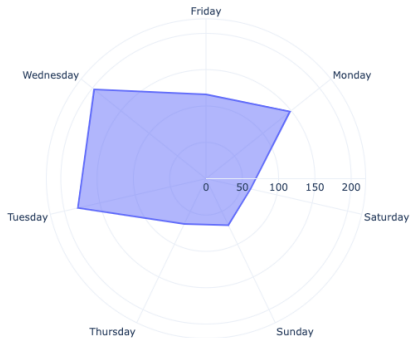




The Erra of Big Data XV



The Erra of Big Data XVI





The Erra of Big Data XVII

Value

- On this data set, analysis and pattern recognition are performed.
- The results of the method may be used to determine the value of the data.
- Making it one of the most crucial big data qualities.

Machine Learning

Machine Learning Process



Image credit: mapendo site

Machine Learning II

Traditional modeling:



Machine Learning:

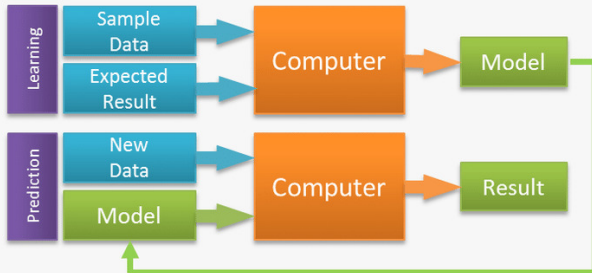


Image credit: Mehra & Hasanuzzaman, (2020)



TYPES OF MACHINE LEARNING

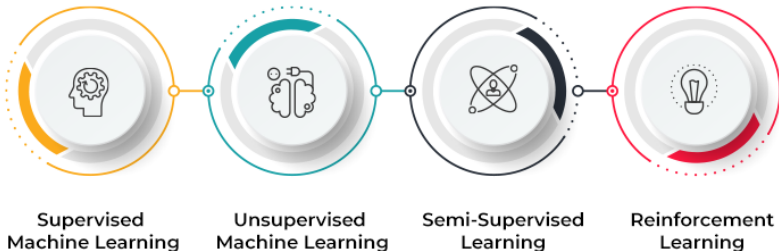


Image credit: spiceworks site

Artificial Intelligence

The theory and development of computer systems able to perform tasks normally requiring human intelligence

Machine Learning

Gives computers "the ability to learn without being explicitly programmed"

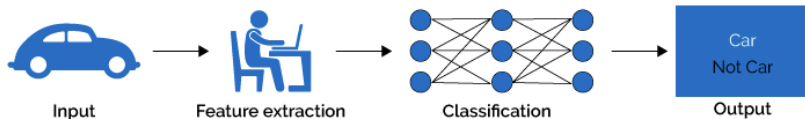
Deep Learning

Machine learning algorithms with brain-like logical structure of algorithms called artificial neural networks

LEVITY

Deep Learning

Machine Learning



Deep Learning

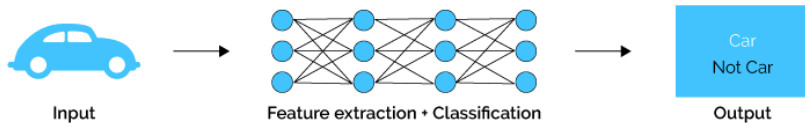
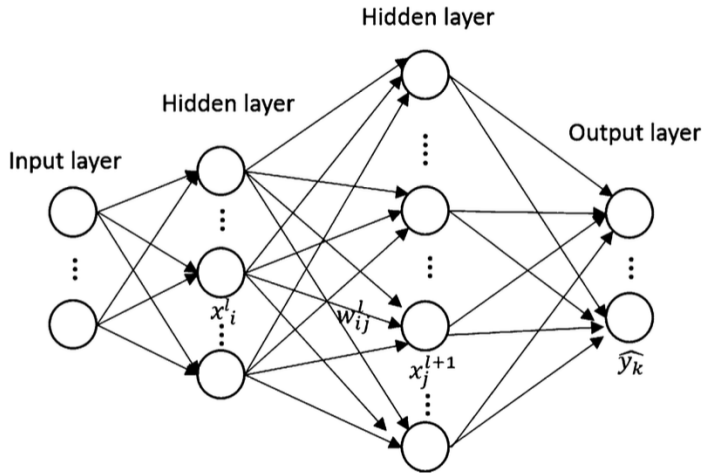


Image credit: Odi & Nguyen, (2018)

Deep Learning II



Schematic of a feed-forward neural network

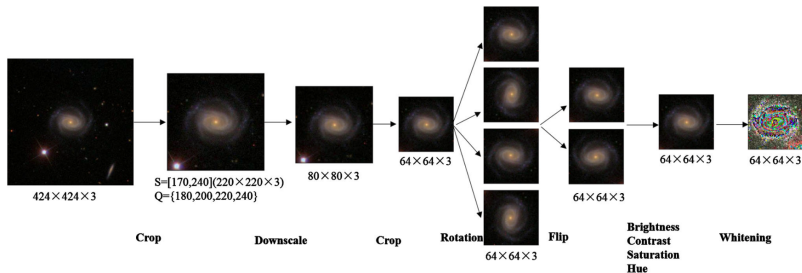


Image credit: Zhu et al , (2019)

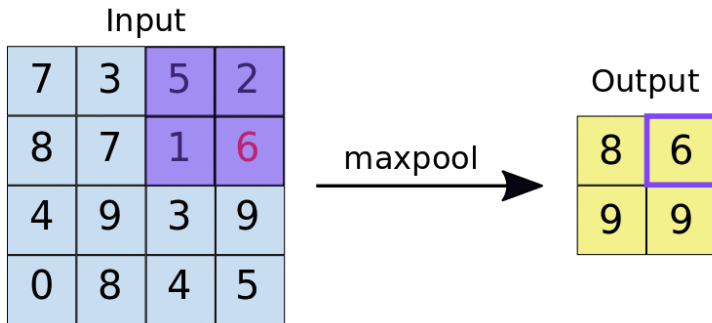
1	1	1	0	0
0	1	1	1	0
0	0	1 _{x1}	1 _{x0}	1 _{x1}
0	0	1 _{x0}	1 _{x1}	0 _{x0}
0	1	1 _{x1}	0 _{x0}	0 _{x1}

Image

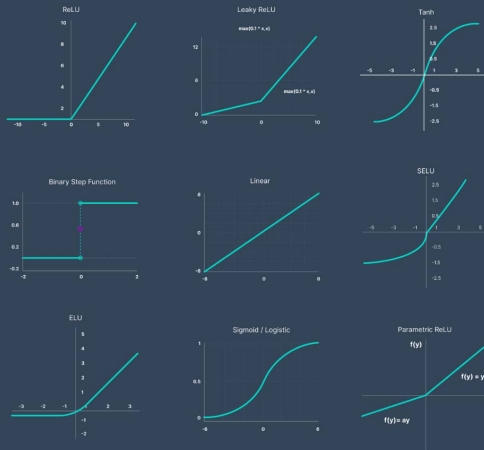
4	3	4
2	4	3
2	3	4

Convolved
Feature

Image credit: Medium



Deep Learning VI



Deep Learning VII

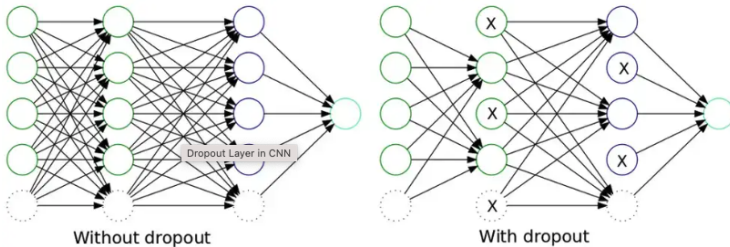


Image credit: Dot Net

Performance Metrics

Performance Metrics Classification I

		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

Performance Metrics Classification II

		Predicted	
		0	1
Actual	0	TN	FP
	1	FN	TP

Performance Metrics Classification III

Metric	Formula	Evaluation focus
Accuracy	$ACC = \frac{TP + TN}{TP + TN + FP + FN}$	Overall effectiveness of a classifier
Precision	$PRC = \frac{TP}{TP + FP}$	Class agreement of the data labels with the positive labels given by the classifier
Sensitivity	$SNS = \frac{TP}{TP + FN}$	Effectiveness of a classifier to identify positive labels. Also called true positive rate (TPR)
Specificity	$SPC = \frac{TN}{TN + FP}$	How effectively a classifier identifies negative labels. Also called true negative rate (TNR)
F_1 score	$F_1 = 2 \frac{PRC \cdot SNS}{PRC + SNS}$	Combination of precision (PRC) and sensitivity (SNS) in a single metric
Geometric mean	$GM = \sqrt{SNS \cdot SPC}$	Combination of sensitivity (SNS) and specificity (SPC) in a single metric
Area under (ROC) curve	$AUC = \int_0^1 SNS \cdot dSPC$	Combined metric based on the receiver operating characteristic (ROC) space (<i>Powers, 2011</i>)

DEMO

Remarks

Workflow of ML

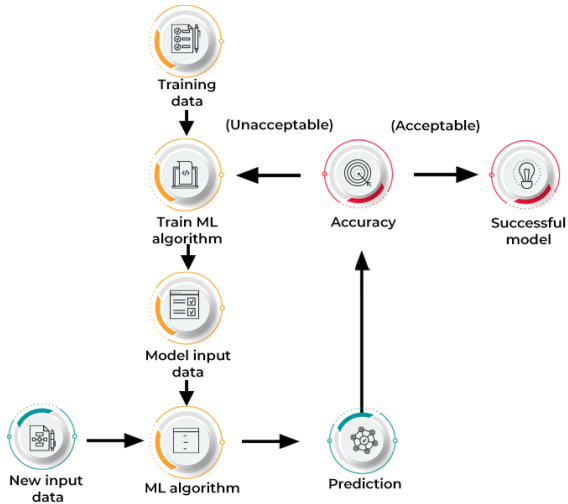


Image credit: spiceworks site

That's All Folks

