

Introduction to Probability and Statistics



Ansar Shahzadi

School of Electrical Engineering & Computer Science
National University of Science and Technology(NUST)

Objectives



Course Logistics



Evaluation Criteria



What is probability and statistics?



Why study statistics?



Some Applications of Probability and Statistics



Branches of Statistics



Define Population, Sample, Parameter,
& Statistic



Data types

Course Logistics

- Course Schedule: Monday , Tuesday, Wednesday, Thursday and Friday
- Location: SEECs

| Time | Monday | Tuesday | Wednesday | Thursday | Friday |
|---------|------------------|------------------|------------------|------------------|------------------|
| 9-950 | BEE-12A(Room-14) | BEE-12C(Room-13) | | BEE-12C(Room-13) | BEE-12A(Room-10) |
| 10-1050 | BEE-12C(Room-13) | BEE-12D(Room-1) | | BEE-12D(Room-12) | |
| 11-1150 | | | BEE-12A(Room-14) | | BEE-12D(Room-13) |
| 12-1250 | | BEE-12B(Room-12) | BEE-12B(Room-7) | | BEE-12B(Room-14) |

Course Logistics

Office Hours

- – Wednesday 10am – 11am
- – Otherwise by appointment

Office Location

- – Room# A-303 SEECs Faculty block
- – Phone# 051-9085 2361
- – Mobile# 03345306117
- – Email:
ansar.shahzadi@seecs.edu.pk

Evaluation Criteria



Assignments

10 %



Quizzes

10%



Mid Term

30 %



End Semester Exam

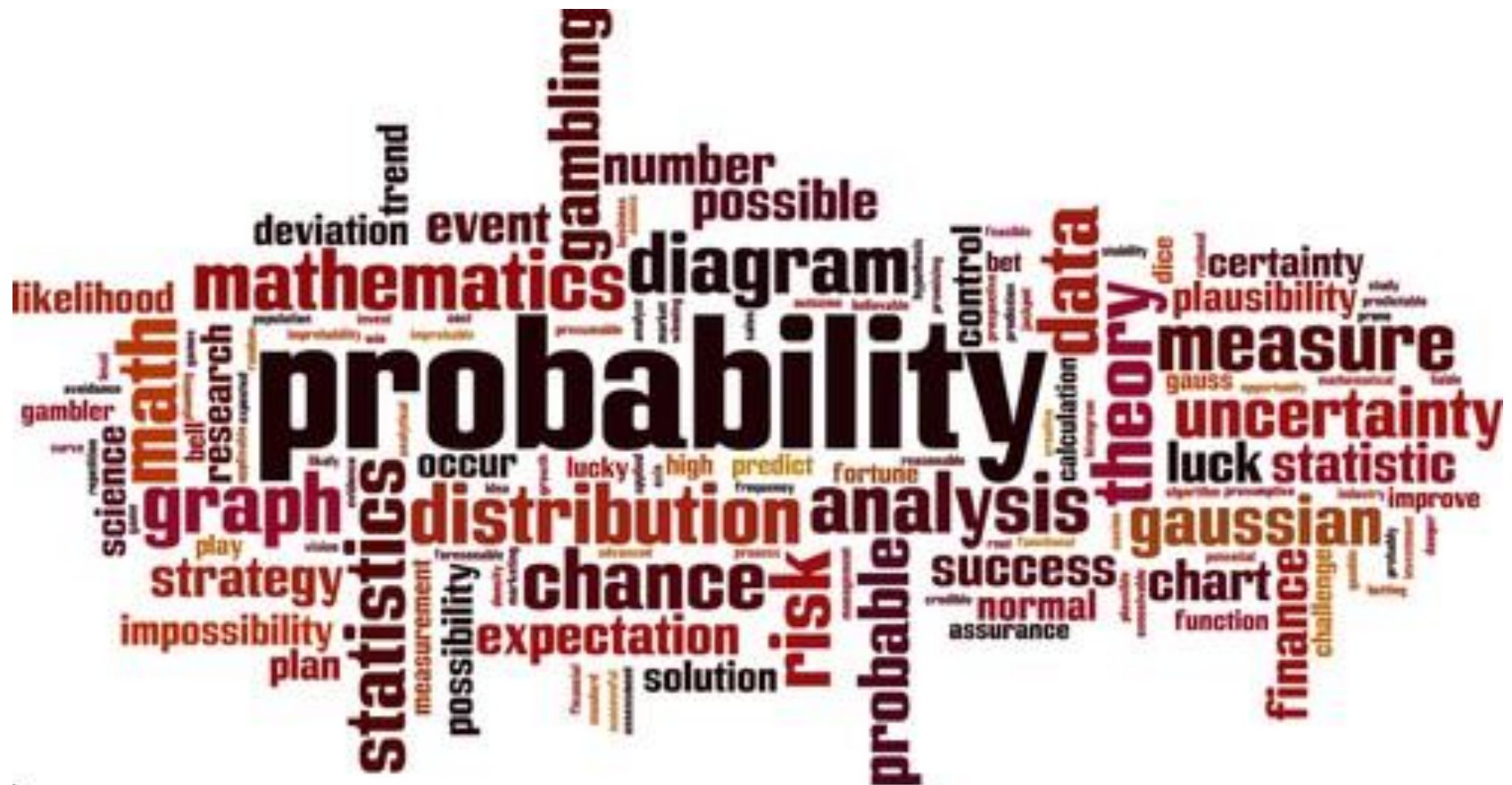
50 %



Why study statistics?



Probability is the Science of Uncertainty

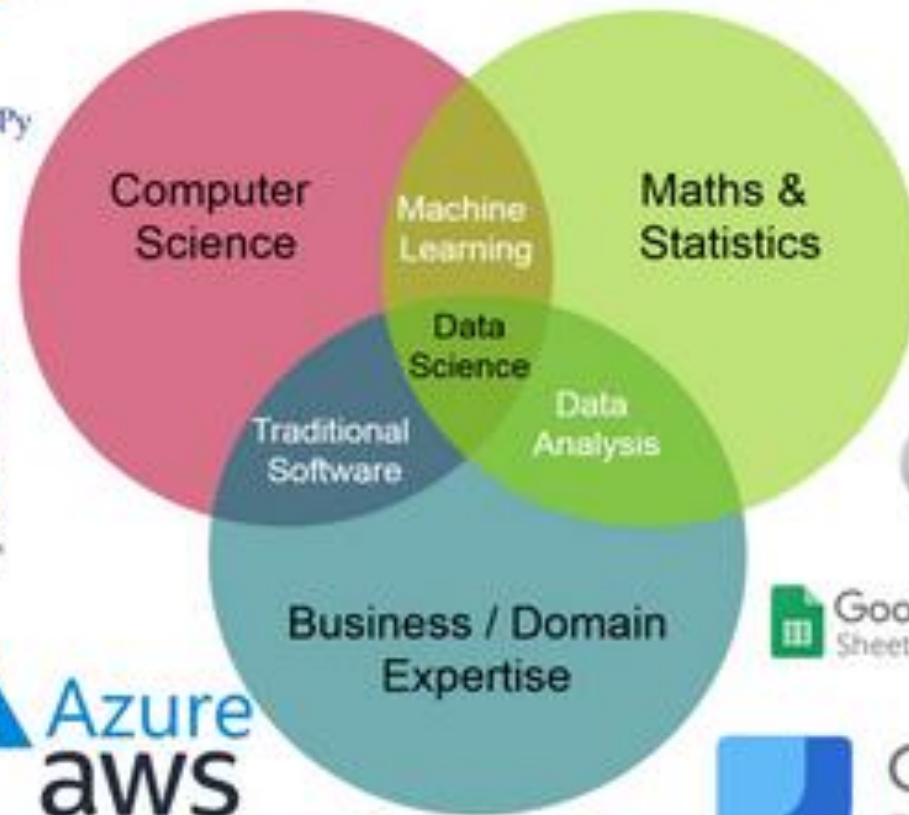


shutterstock.com · 1382524664

A close-up, blue-tinted photograph of a document. In the upper right, a silver pen with a black nib is positioned over a line graph. The graph features a solid line and a dotted line, both showing an upward trend. The number '5' is visible on the left side of the graph, and '2,47' is visible on the right side. The background is a light blue, textured surface.

Some Applications of Probability and Statistics

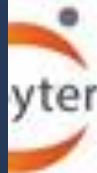
Data Science

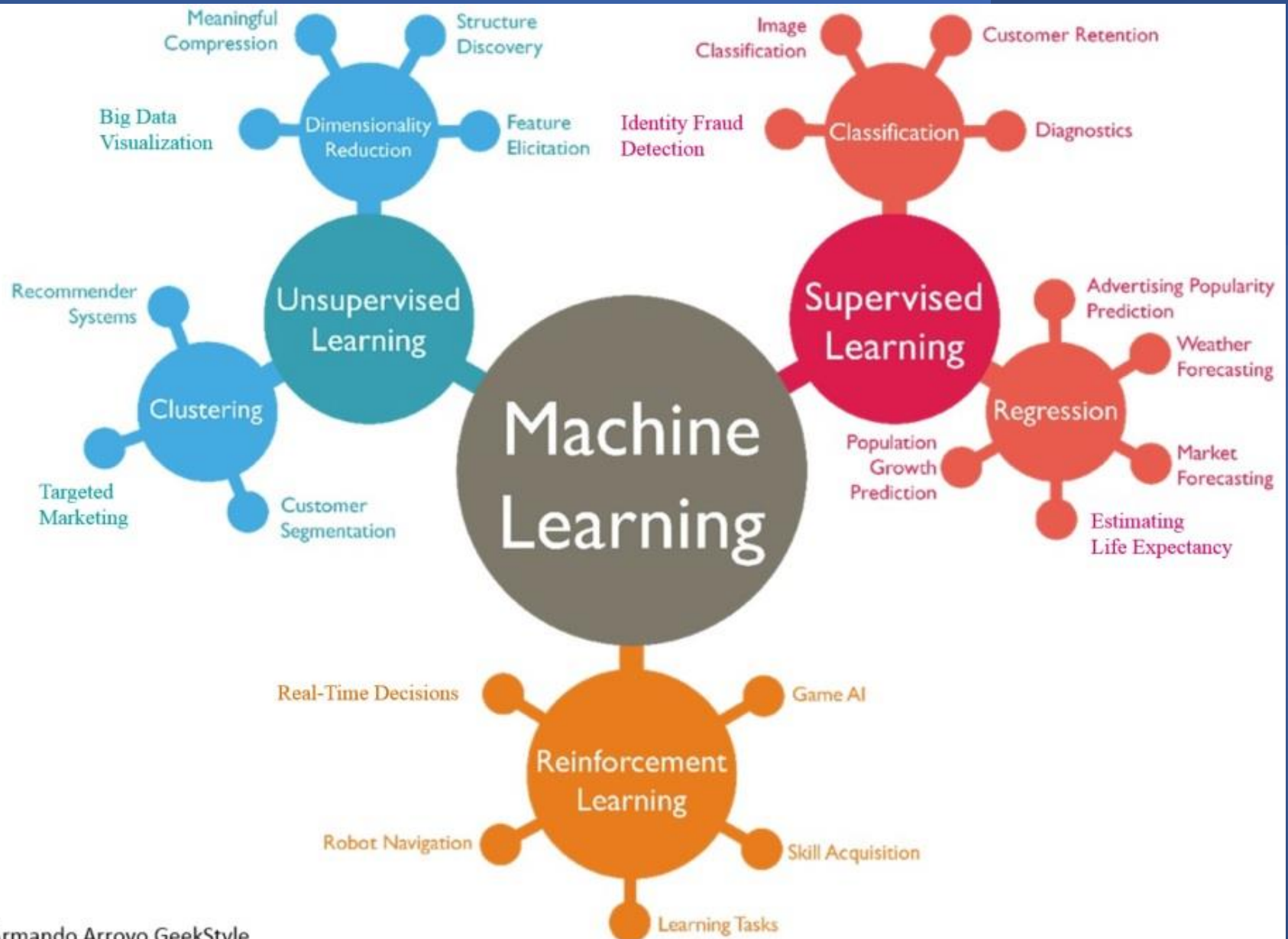


$$\begin{matrix} \sqrt{} & = & \neq \\ + & - & \div & \times \\ \% & \pi & \sigma \end{matrix}$$



Statistics





3 Key Artificial Intelligence Statistics You Should Know

1 Which processes rely heavily on AI?

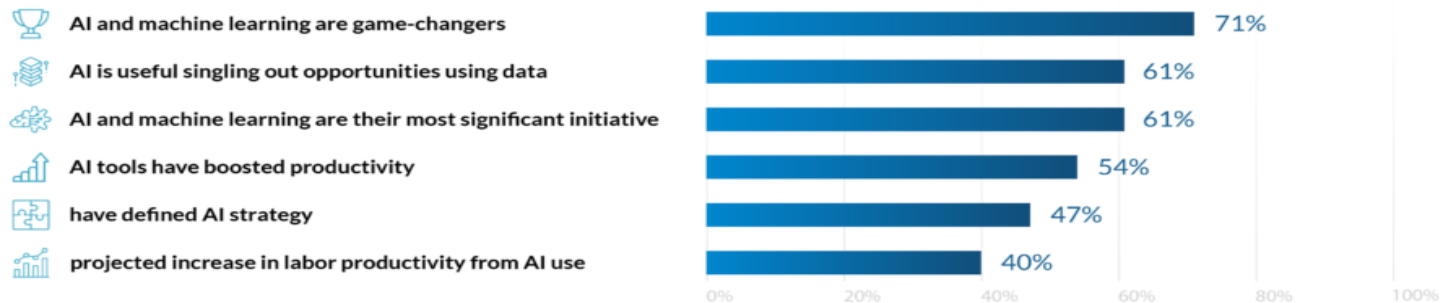
Source: Statista

Global AI ranked by revenue of use cases:



2 How do organizations and leaders perceive AI?

Sources: MemSQL, Accenture, narrativesscience.com, cmo.com, pwc.com

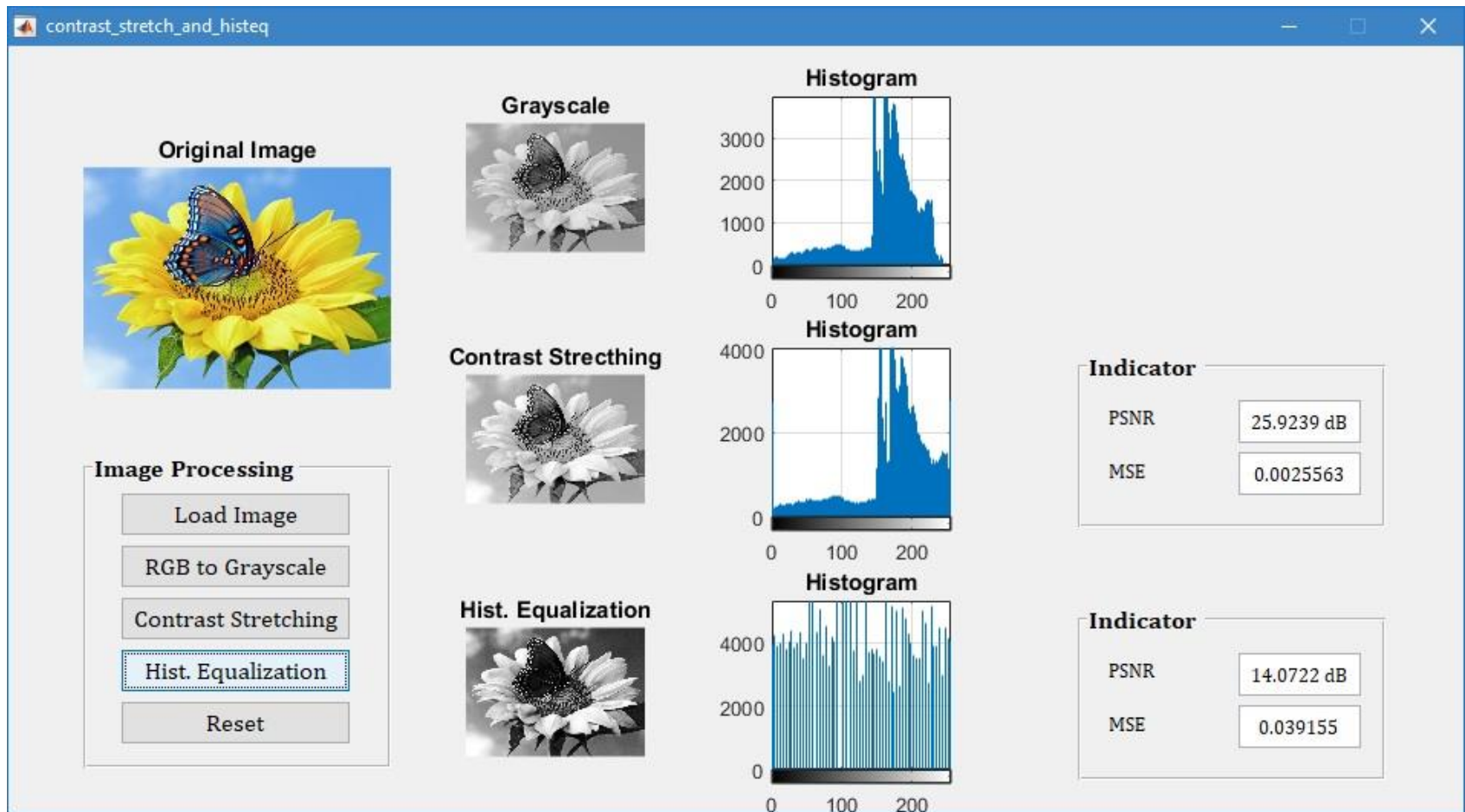


3 What are the top benefits of AI adoption?

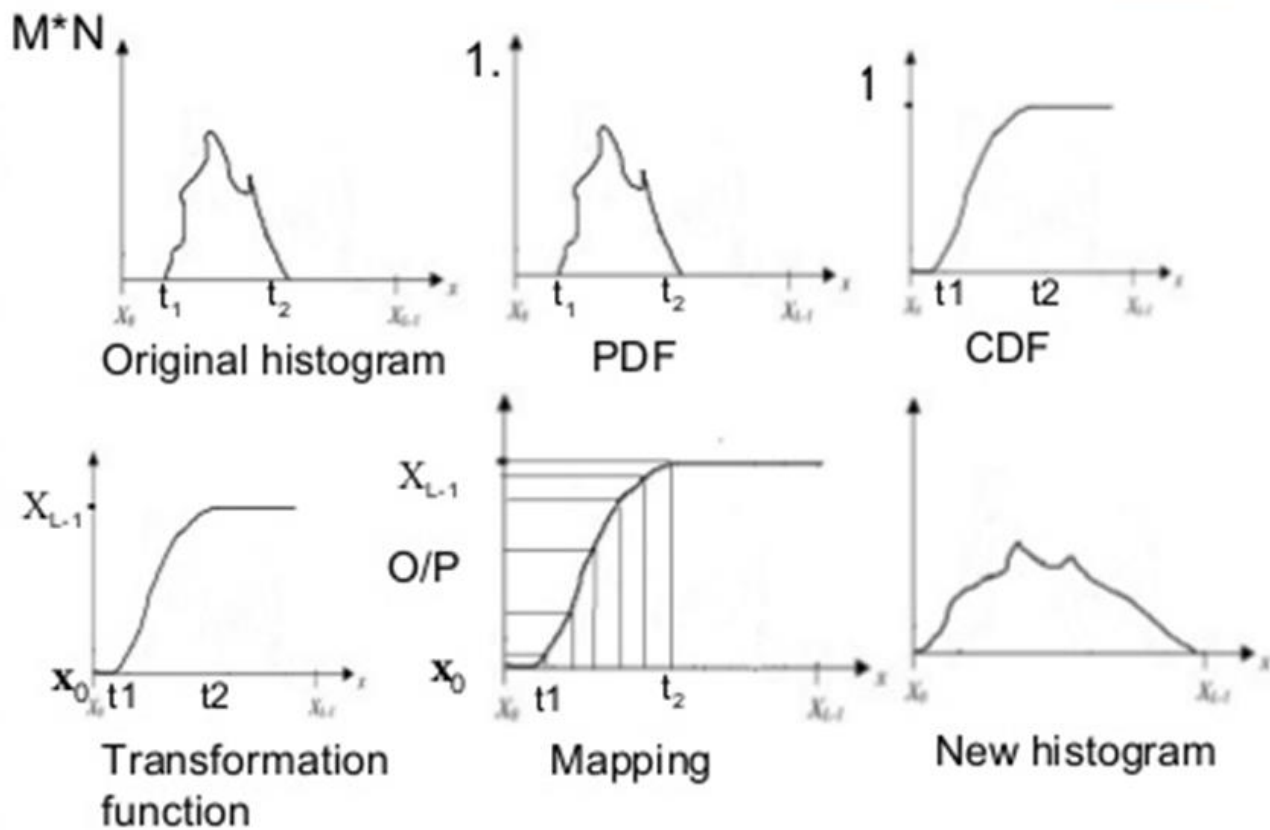
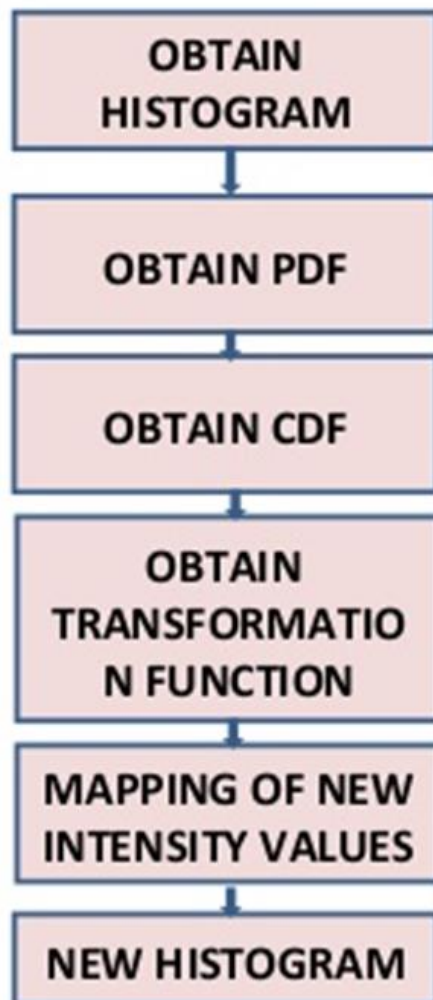
Source: The Economist, Forbes, pwc.com, hbr.org



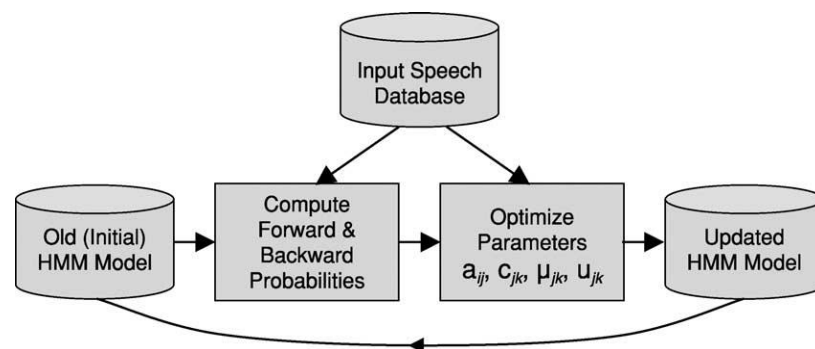
Application of Histogram in Digital image Processing



GLOBAL HISTOGRAM EQUALIZATION



Statistical Techniques Used in Speech Recognition



Naïve Bayes (NB) Classifier

- ▶ **"Bayes"** because the class c^* attached to an example \mathbf{x} is determined by the Bayes' Theorem

$$c^* = h_{\text{Bayes}}(\mathbf{x}) = \arg \max_{j=1..m} P(c_j) P(\mathbf{x} | c_j)$$

when the attribute space is high dimensional direct estimation is hard unless we introduce some assumptions



- ▶ **"Naïve"** because of its very naïve **independence assumption**:

all the attributes are conditionally independent given the class

$$P(\mathbf{x} | c_j) = \prod_{i=1}^n P(X_i = x_i | c_j)$$

$P(\mathbf{x} | c_j)$ can be decomposed into a product of n terms, one term for each attribute

**NB
Classification
Rule**

$$c^* = h_{\text{NB}}(\mathbf{x}) = \arg \max_{j=1..m} P(c_j) \prod_{i=1}^n P(X_i = x_i | c_j)$$

Applications of Naïve Bayes Classifier



Document
Categorization



Email
Spam Filtering



Sentiment
Analysis

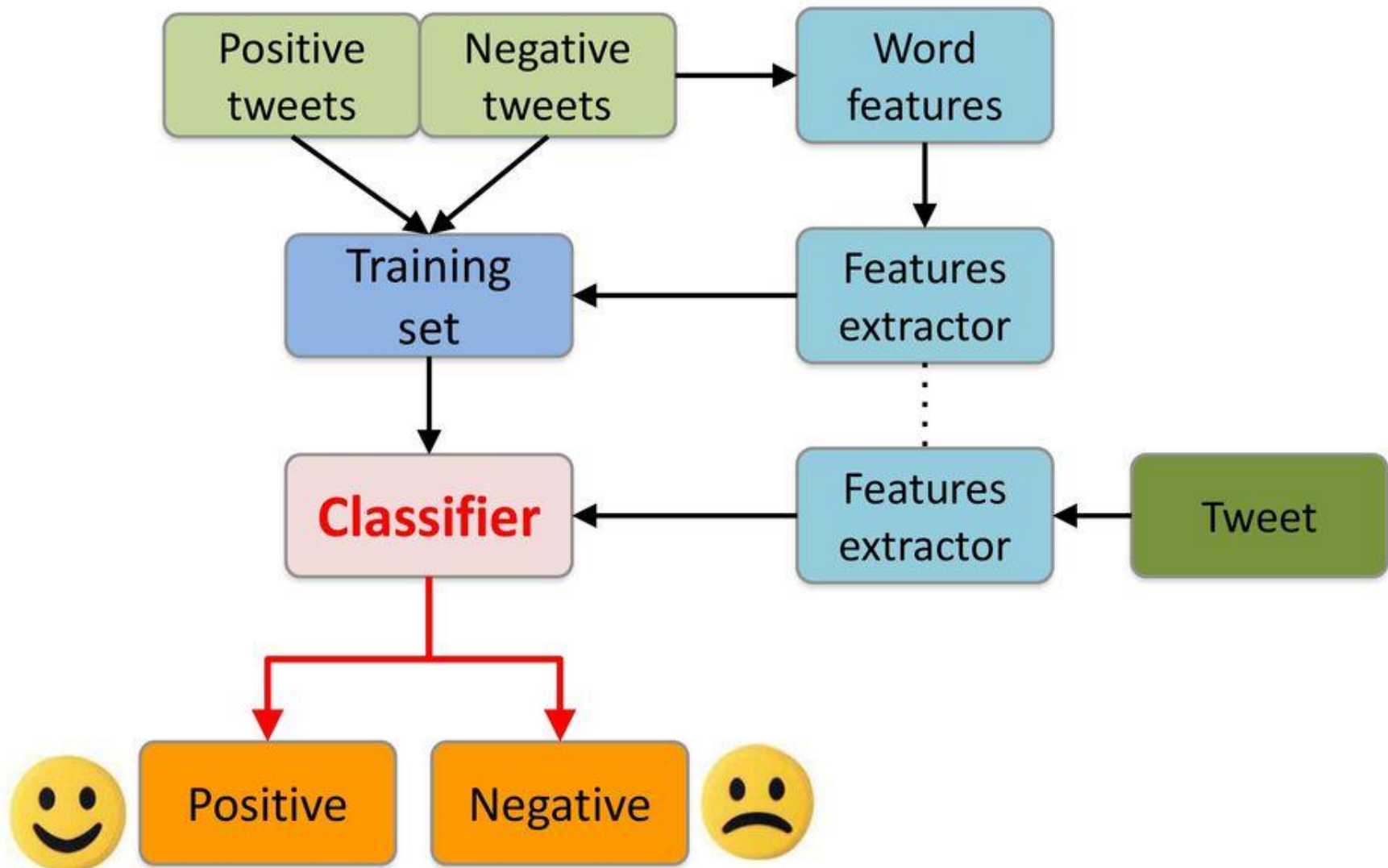
Sentiment Analysis

Sentiment Analysis

What is Sentiment Analysis?

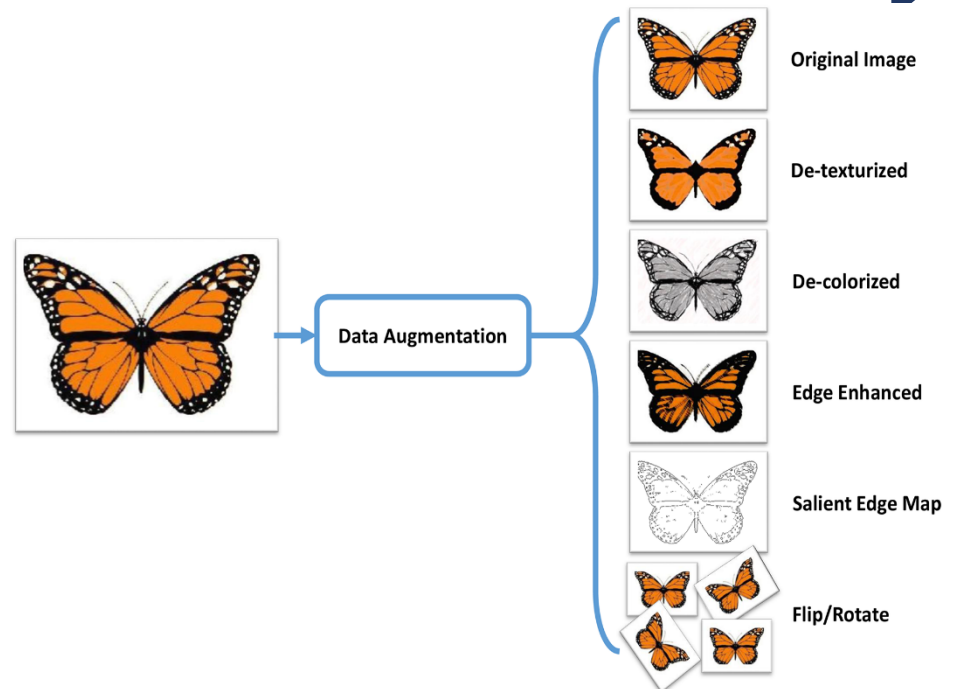


Sentiment Analysis Architecture



Data augmentation

Data augmentation is an effective method to expand the training data by applying transformations and deformations to the labeled data, resulting in new samples as additional training data. A key attribute of the data augmentation is that the labels remain unchanged after applying those transformations.



Data Augmentation

Data – What If We Don't Have Enough Data?

• Data Augmentation for Effective Training Set Expansion

- In many cases, data augmentation techniques used in natural images does not semantically make sense in medical image
(flips, rotations, scale shifts, color shifts)
- Physically-plausible deformations or morphological transform can be used in limited cases.
- More augmentation choices for texture classification problems.

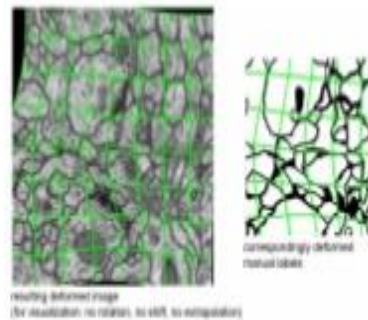
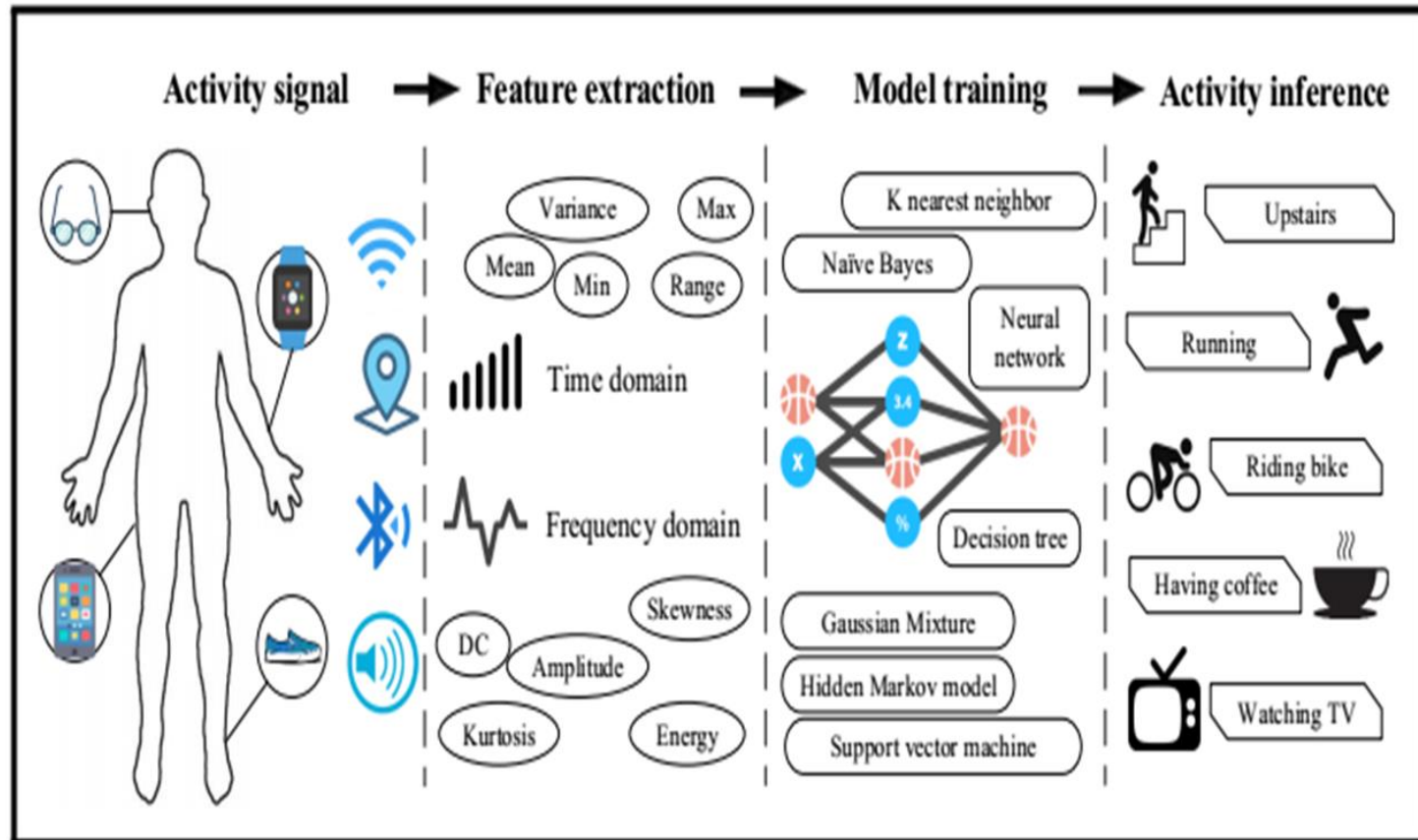


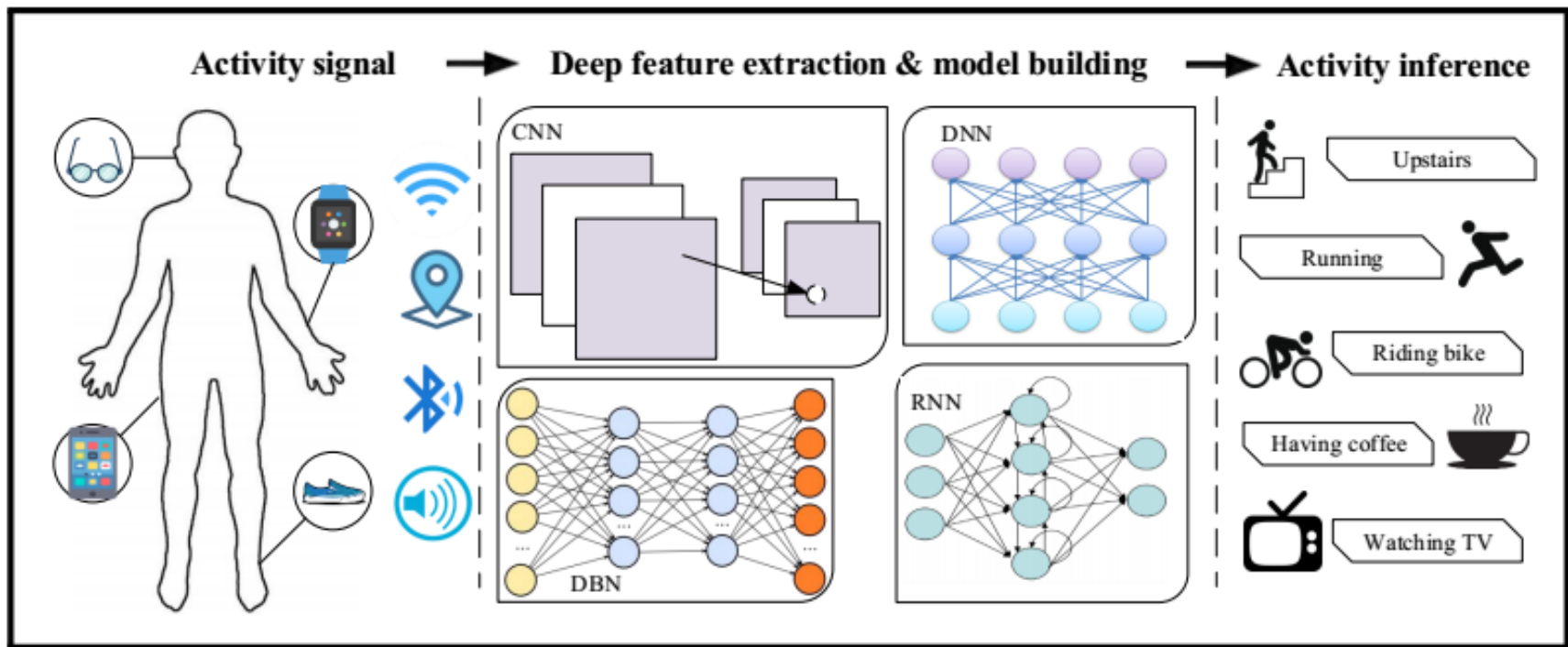
Figure 3. We generate several random thin-plate-spline deformations in 2D in order to generate slight variations that are physically plausible in our training data. Some examples are shown here.

H. R. Roth et. al., MICCAI, 2015

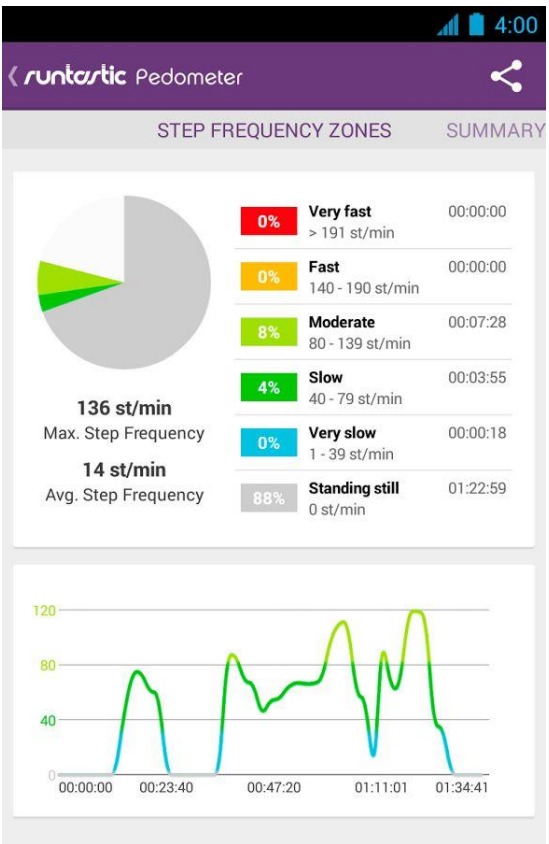
Activity Tracker (Old Approach)



Activity Tracker (New Approach)

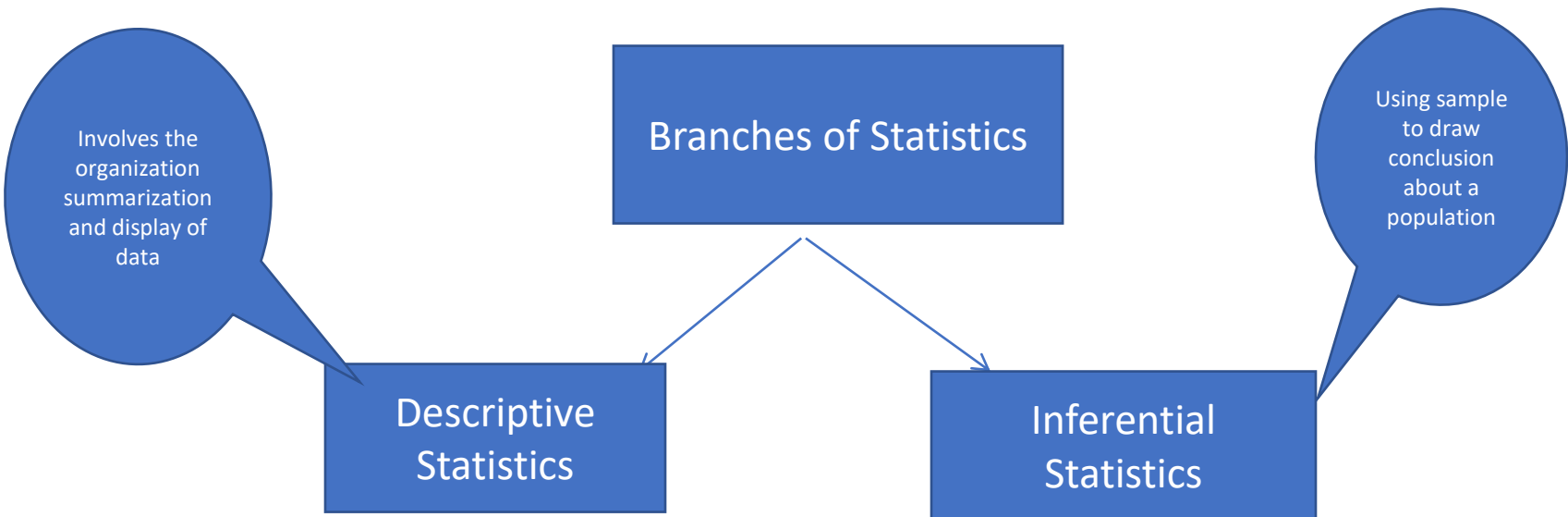


Pedometer



Statistics

Statistics concerns data; their collection, analysis, and interpretation





Definitions

- Population: The entire group of individuals is called the population
- Parameter: A numerical measurement describing some characteristic of a Population
- Sample: Representative part of population
- Statistic: A numerical measurement describing some characteristic of a sample

Data Types

- Quantitative: Numbers representing counts or measurements
 - Discrete: Counts
 - Continuous: Measurements
- Qualitative: Quality, Attribute



Levels of Measurements

- Nominal
 - categories only
- Ordinal
 - categories with Some order
- Interval
 - differences but no natural starting point
- Ratio
 - differences and a natural starting point