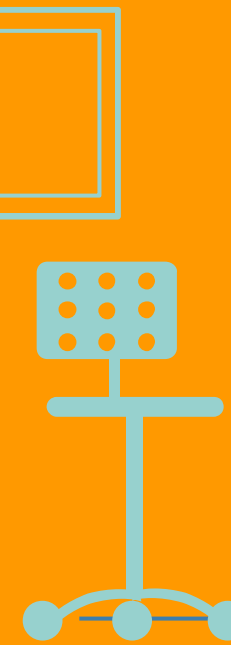
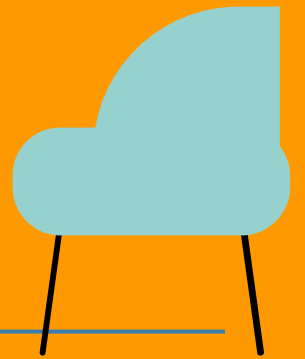
"In the world of machine learning, the more you feed your model, the smarter it becomes. It's like giving your computer a crash course in common sense!"





ArthurKakande

Follow

1 follower · 8 following

Block or Report

Overview | Repositories 22 | Projects | Packages | Stars 1

ArthurKakande / README.md

Hi there 👋

I'm Arthur Kakande, i enjoy 🔬 turning data into insights and building intelligent solutions! 💡. Welcome to my GitHub profile!

About Me

- 🔭 I'm currently working on The Africa Knowledge Graph
- 🌟 I'm interested in AI and Intelligent Systems
- 💬 Ask me about Information Retreival, Semantic Technologies, Expert Systems, Natural Language Undertstanding, ML, Distributed AI, etc.
- 📫 How to reach me: https://twitter.com/arthurkakande

My Skills

- Programming languages: R, Python, Java, SQL
- Frameworks and libraries: Streamlit, R Shiny, Tensorflow, Langchain
- Tools and technologies: Protege, GraphDB
- Databases: PostgreSQL
- Other skills: Statistical Analysis

Get in Touch

Feel free to reach out to me if you have any questions, ideas, or just want to say hello! I'm always excited to connect with fellow developers and enthusiasts.

Let's collaborate, learn, and create amazing things together!

Machine Learning

In learning we don't give the computer instructions on how to perform a task, rather we give it data or information and let it learn some patterns to be able to perform a task on its own.

Supervised Learning;
Given a data set of input and output pairs, learn a function that maps inputs to outputs.

Classification;
A task in supervised learning that deals with mapping an input to a discrete category.


₹

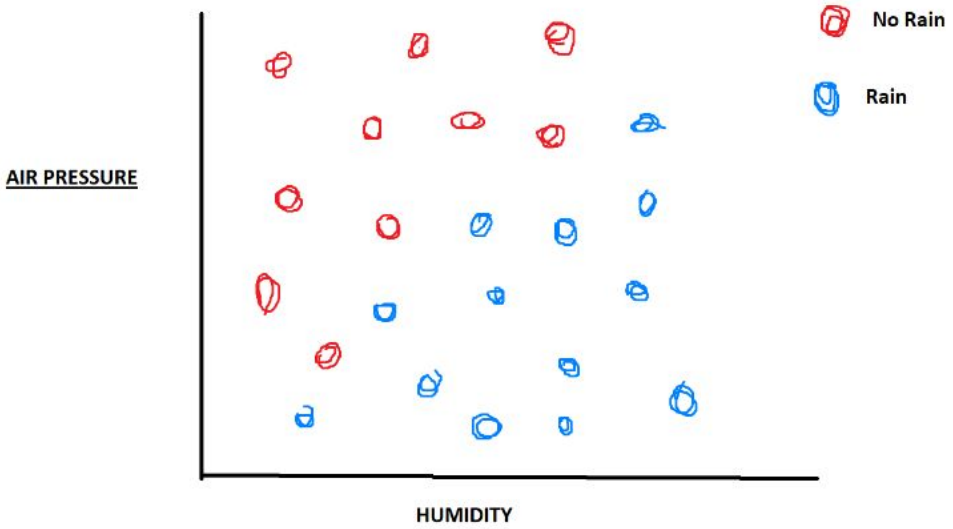

DEBT

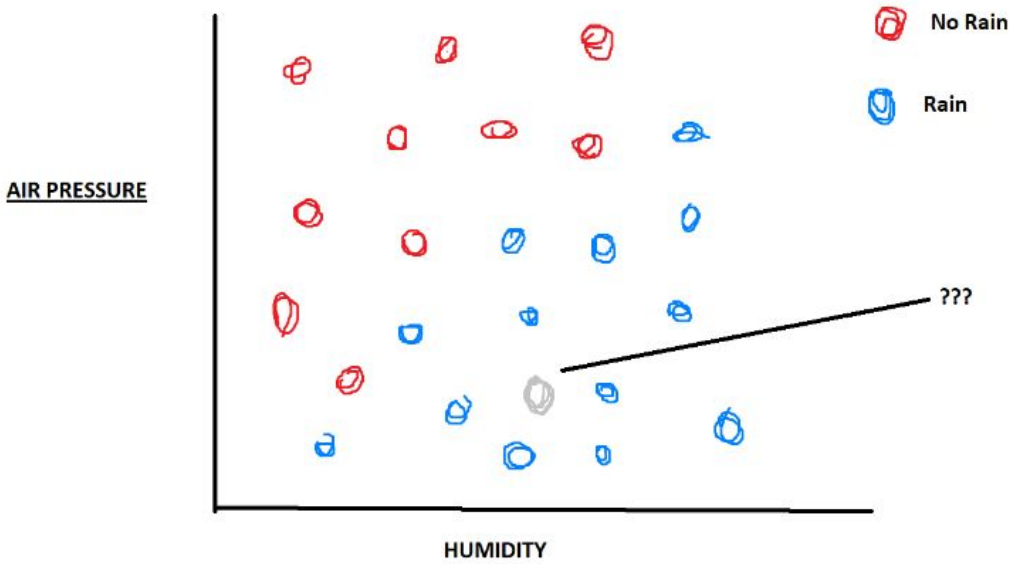

365
SPAM
HOSTPAPA


501
Junk
Not Junk
Block Sender

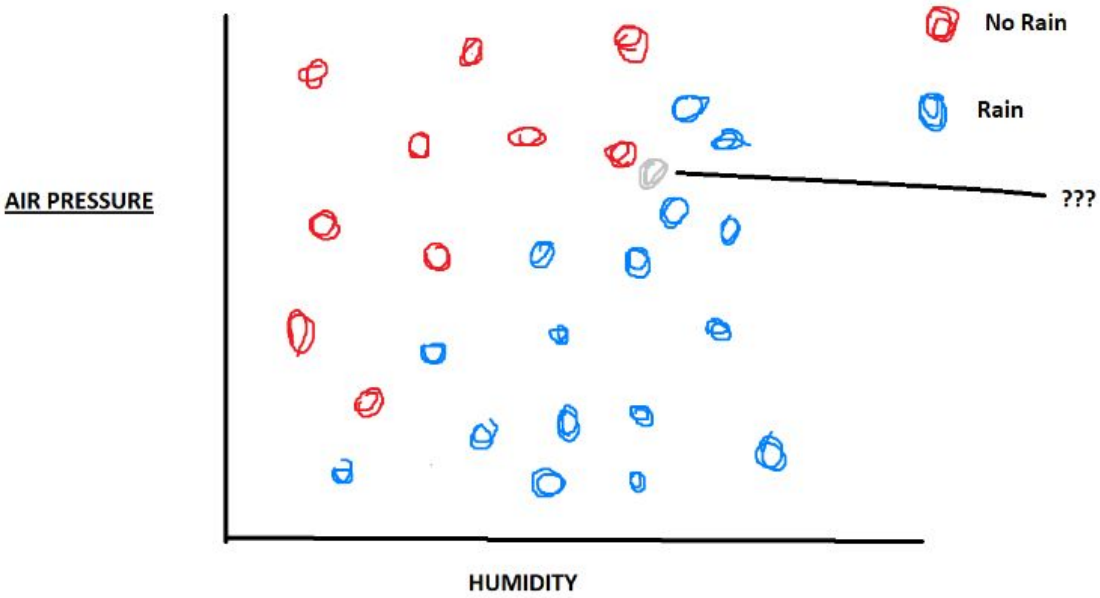
	Gender	Married	ApplicantIncome	LoanAmount	Loan_Status
0	Male	Yes	4583	128000.0	N
1	Male	Yes	3000	66000.0	Y
2	Male	Yes	2583	120000.0	Y
3	Male	No	6000	141000.0	Y
4	Male	Yes	5417	267000.0	Y

Humidity	Air pressure	Rain
18	25	Rain
19	3	No Rain
25	5	No Rain
20	7	No Rain
8	25	Rain
10	40	Rain
6	25	-

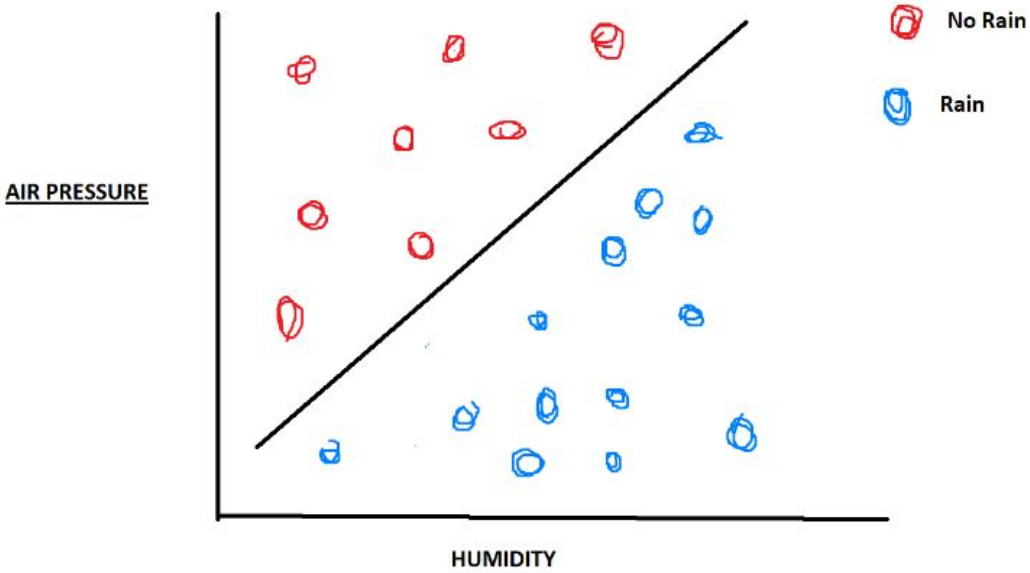

No Rain
Rain
AIR PRESSURE
HUMIDITY

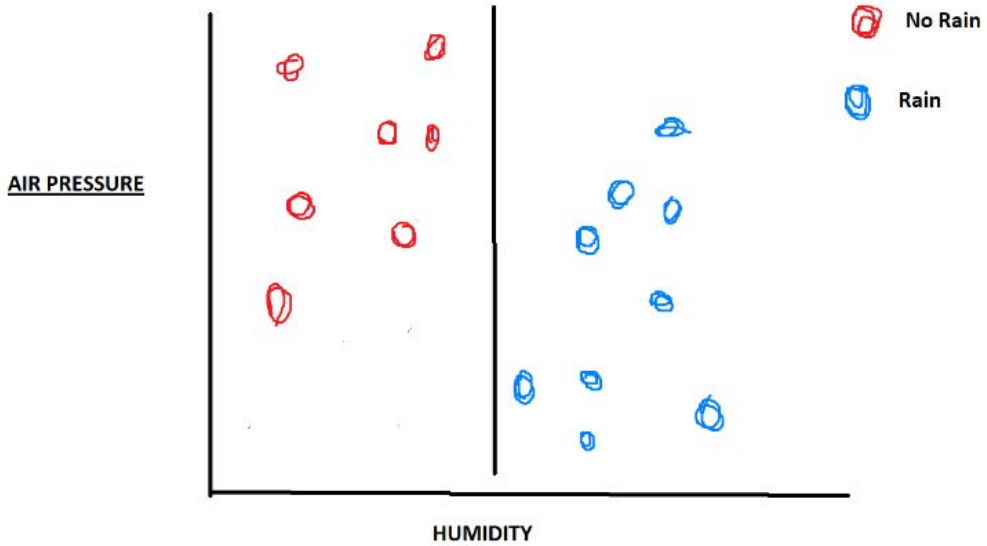

No Rain
Rain
AIR PRESSURE
???
HUMIDITY

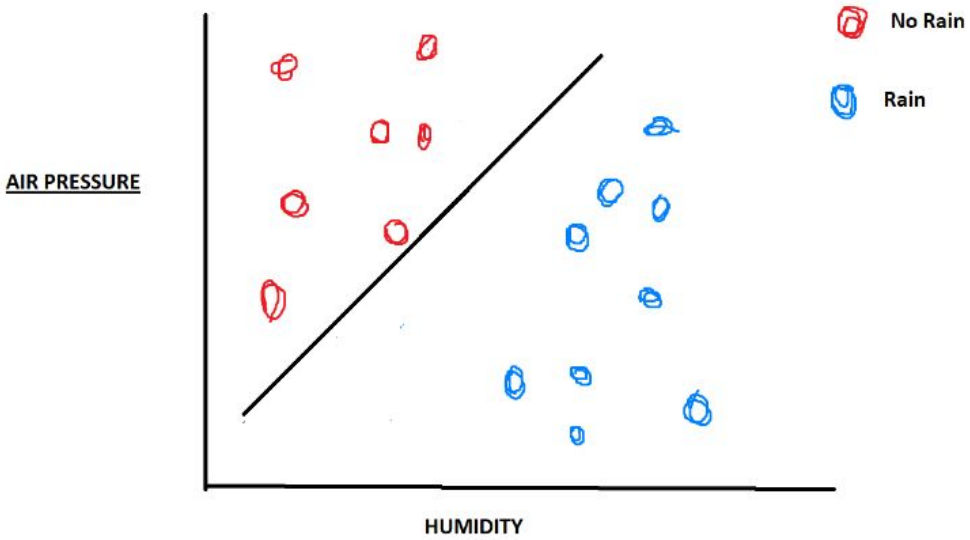
Nearest Neighbour Classification; An algorithm that, given an input chooses the nearest data point to that input.

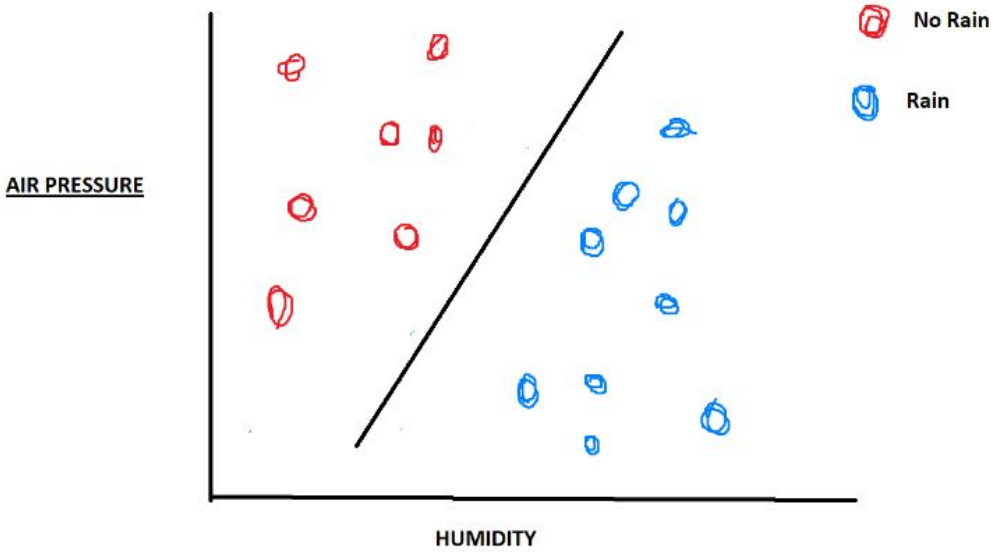

No Rain
Rain
???
AIR PRESSURE
HUMIDITY

K-nearest Neighbour Classification; An algorithm that, given an input chooses the most common class out of the k - nearest data points to that input.

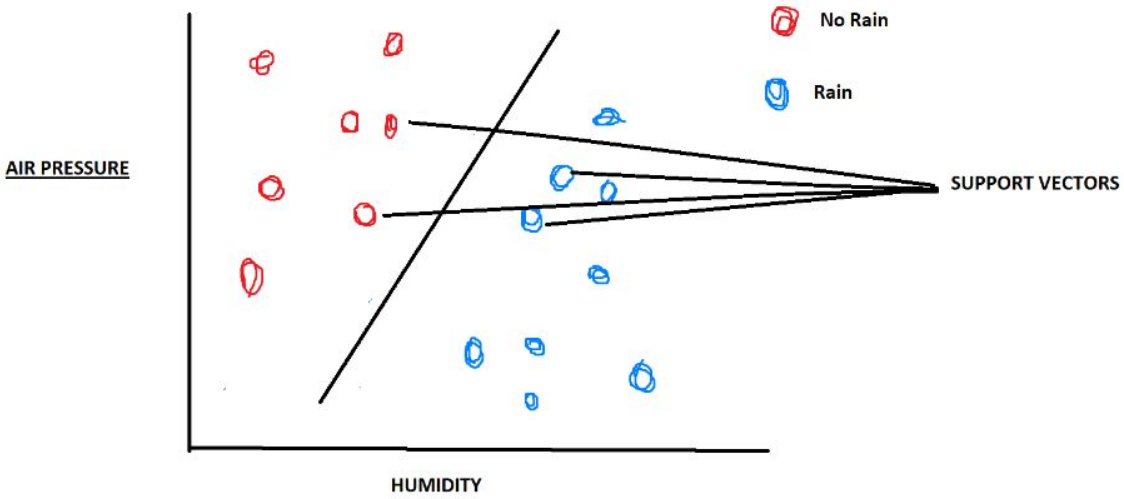

No Rain
Rain
AIR PRESSURE
HUMIDITY


No Rain
Rain
AIR PRESSURE
HUMIDITY

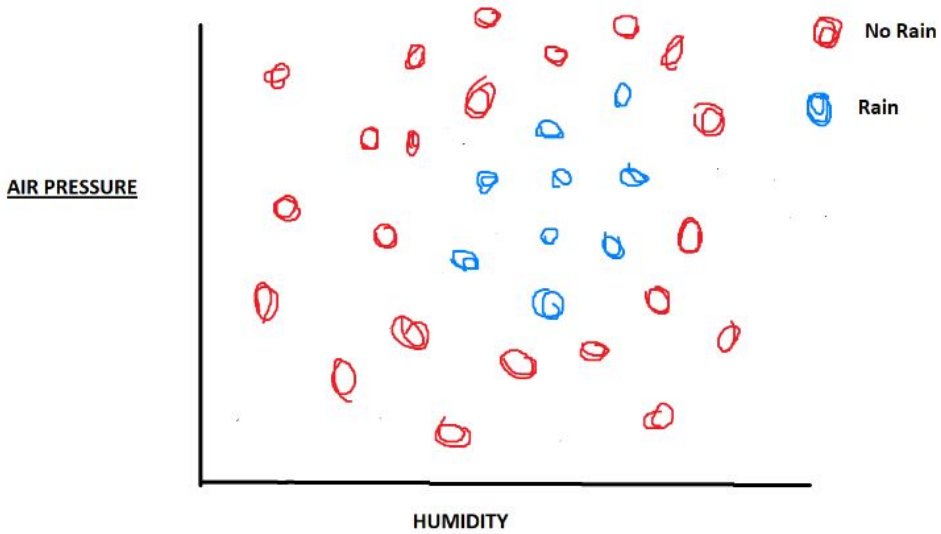

No Rain
Rain
AIR PRESSURE
HUMIDITY

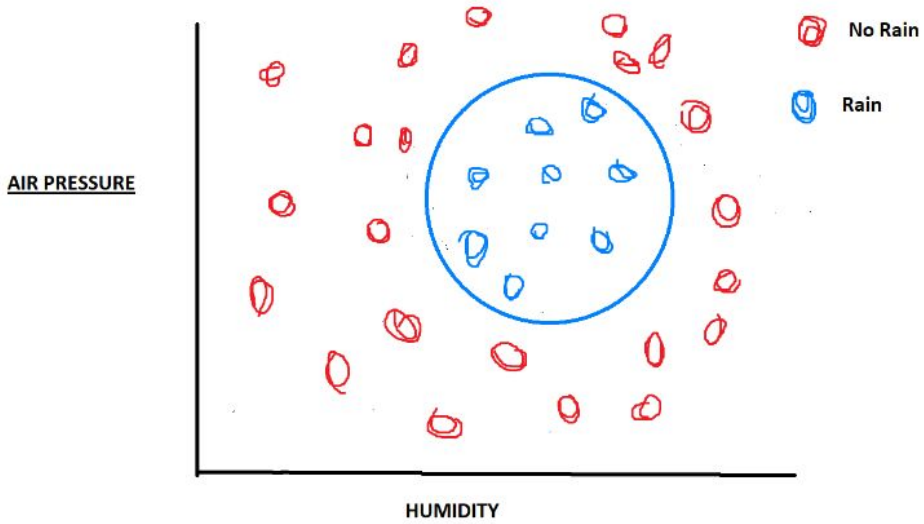

No Rain
Rain
AIR PRESSURE
HUMIDITY

Maximum margin separator; Boundary that maximises the distance between any of the data points


No Rain
Rain
AIR PRESSURE
SUPPORT VECTORS
HUMIDITY

Support Vector Machines;
SVM maps training examples to points in space so as to maximise the width of the gap between the two categories.


No Rain
Rain
AIR PRESSURE
HUMIDITY


No Rain
Rain
AIR PRESSURE
HUMIDITY

Humidity	Air pressure	Rain
18	25	Rain
19	3	No Rain
25	5	No Rain
20	7	No Rain
8	25	Rain
10	40	Rain
6	25	-

Humidity	Air pressure	Rain
18	25	Rain
19	3	No Rain
25	5	No Rain
20	7	No Rain
8	25	Rain
10	40	Rain
6	25	-

Decision Tree;
We categorize or make predictions based on how a previous set of questions were answered

Regression;
A task in supervised learning that deals with mapping an input to a continuous value.

Common algorithms for regression include linear regression, polynomial regression, etc. Applications of regression include stock price prediction and weather forecasting.

Humidity	Air pressure	Temperature
30	25	5.3
5	3	20.2
25	5	18.5
20	7	24.8
8	25	2.3
10	40	4.6
6	25	-

Evaluating models

0 - 1 Loss;
Each time a wrong category is predicted, the model scores a loss of 1

L1 Loss;
Measures how close the predicted value is to the actual value i.e. (Actual - Predicted)

L2 Loss;
A variant of the L1 loss in which the difference between the Actual value and the predicted value is squared.

Overfitting;
A model that fits too closely to a particular dataset and may therefore fail to generalise on future datasets

holdout cross-validation;
splitting data into a training set and a testing set, such that learning happens on the training set and is evaluated on the test set

K-fold cross-validation;
splitting data into k sets and experimenting k times, using each set as a test set once and using the remaining data as a train set.

Reinforcement learning

How does it differ;
Given a set of rewards or punishments, an agent learns what actions to take in the future.
Examples; game playing and robotics.

Unsupervised learning

How does it work;
Given data without any labels, learn patterns

Clustering;
An unsupervised learning task that looks at organising a set of objects into groups in such a way that similar objects tend to be in the same group

Applications

Of Clustering

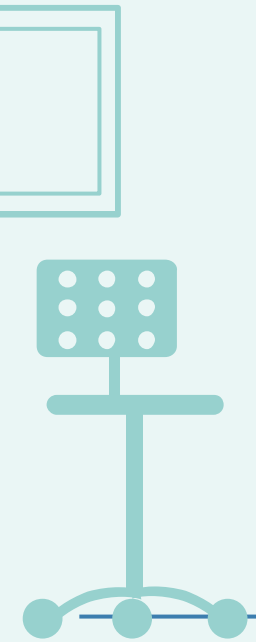
- Market Research
- Image segmentation
- Social Network Analysis
- Medical imaging

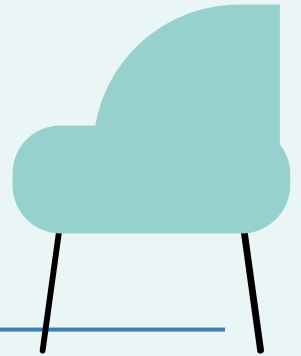
K-means Clustering; Repeatedly assigning points to clusters and updating clusters' centers

So what is Data Mining?

Data mining is the process of discovering patterns and insights from large datasets. It involves applying techniques from statistics, machine learning, and database systems to extract meaningful information from raw data.

Any Questions...





Demo Time!!

Survey time…

Sources;

- Harvard’s CS50
- Introduction to Machine Learning Class Notes (By Carnegie Mellon University)
- University of Toronto (http://www.cs.toronto.edu/~mbrubake/teaching/C11/notesReadings.html)