

Developing a Flight Delay Prediction Model using Machine Learning

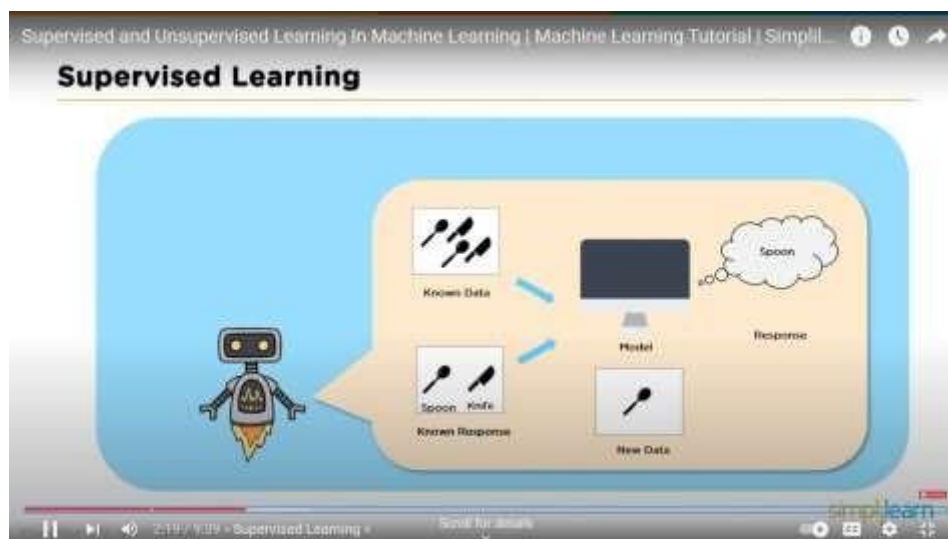
TEAM LEADER : Saranya M

TEAM MEMBER :

1. Raja Rathna M
2. Ramya M
3. Uma M

Prior Knowledge:

Supervised and unsupervised learning:



Regression Classification and Clustering:

The image displays two screenshots of a YouTube video player, showing a presentation on machine learning problems. The video is titled "Common ML Problems" and is part of a series called "Introduction to Machine Learning" by DataCamp. The presenter, a man in a dark shirt, is visible in the bottom right corner of the video frame.

Top Screenshot: The video is at the 0:22 mark. The slide titled "Common ML Problems" lists three bullet points: "Classification", "Regression", and "Clustering".

Bottom Screenshot: The video is at the 6:21 mark. The slide titled "k-Means" has the subtitle "Cluster data in k clusters!". It features two scatter plots. The left plot shows three distinct clusters of black data points. The right plot shows the same data points, but each cluster is assigned a different color (red, green, and blue) to represent the three clusters.

Below the video player, the video title "Machine Learning in R - Classification, Regression and Clustering Problems" is visible, along with the upload date "05,674 views • Dec 5, 2015". The video has 582 likes and 1 dislike. The video player interface includes a search bar, a sign-in button, and a progress bar.

Python Flask Tutorial For Beginners | Flask Web Development Tutorial | Python Training | Edureka

```
from flask import Flask
app = Flask(__name__)


@app.route('/hello/')
def hello_name(name):
    return 'Hello %s' % name

if __name__ == '__main__':
    app.run(debug = True)
```

WARNING: Do not use the development server in a production environment.
Use a production WSGI server instead.
* Debug mode: on
* Running on http://127.0.0.1:5000/ (Press CTRL+C to quit)
127.0.0.1 - - [14/Dec/2018 11:44:47] "GET / HTTP/1.1" 200 -

Python Flask Tutorial For Beginners | Flask Web Development Tutorial | Python Training | Edureka

Flask – Redirect & Errors

 Standardized status codes

Prototype \Rightarrow Flask.abort(code)

Sl.no	Status Code
1	HTTP_300_MULTIPLE_CHOICES
2	HTTP_301_MOVED_PERMANENTLY
3	HTTP_302_FOUND
4	HTTP_303_SEE_OTHER
5	HTTP_304_NOT_MODIFIED
6	HTTP_305_USE_PROXY
7	HTTP_306_RESERVED

Sl.no	Code	Description
1	400	Bad Request
2	401	Unauthenticated
3	403	Forbidden
4	404	Not Found
5	406	Not Acceptable
6	415	Unsupported Media Type
7	429	Too Many Requests

Tutorial 37: Entropy In Decision Tree Intuition

DECISION TREE ENTROPY

Entropy ↓

the purity of split

$H(s) = -P_{+} \log_2(P_{+}) - P_{-} \log_2(P_{-})$

P_{+} / P_{-} : % of +ve class / % of -ve class

Training Example

9 Yes / 0 No

f_1

f_2 f_3 o/p

Yes No

split/no

split/yes

split/no

split/yes

2:50 / 8:52

Scroll for details

Tutorial 37: Entropy In Decision Tree Intuition

DECISION TREE ENTROPY

3 Yes / 2 No = 1 bits

Entropy: Means the purity of split

$H(s) = -P_{+} \log_2(P_{+}) - P_{-} \log_2(P_{-})$

P_{+} / P_{-} : % of +ve class / % of -ve class

S = Subset of Training Example

$= -\frac{3}{5} \log_2(\frac{3}{5}) - (\frac{2}{5} \log_2(\frac{2}{5}))$

0.72 bits

9 Yes / 0 No

f_1

f_2 f_3 o/p

Yes No

split/no

split/yes

split/no

split/yes

2 Yes 0 No

0 Yes 2 No

2 Yes 0 No

0 Yes 2 No

8:05 / 8:52

Scroll for details

Decision Tree information gain:

Tutorial 38- Decision Tree Information Gain

DECISION TREE INFORMATION GAIN

ENTROPY

f_1 9Y/5N
 f_2 3Y/3N

$H(f_1) = 0.91$
 $H(f_2) = 0.91$

② Information Gain

$Gain(S, A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$
 $H(S) = 0.91$

f_1 6Y/2N f_2 3Y/3N
 $H(f_1) = 0.91$ $H(f_2) = 1$

$Gain(S, f_1)$
 $= H(S) - \frac{8}{14} H(f_1) - \frac{6}{14} H(f_2)$
 $= 0.91 - \frac{8}{14} \times 0.91 - \frac{6}{14} \times 1$
 $= 0.049$

$H(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$
 $= -\frac{8}{14} \log_2 \frac{8}{14} - \frac{6}{14} \log_2 \frac{6}{14}$

0.06 / 12:39

Tutorial 38- Decision Tree Information Gain

DECISION TREE INFORMATION GAIN

ENTROPY

f_1 9Y/5N
 f_2 3Y/3N
 f_3 1Y/1N

$H(f_1) = 0.91$
 $H(f_2) = 0.91$
 $H(f_3) = 1$

② Information Gain

$Gain(S, A) = H(S) - \sum \frac{|S_v|}{|S|} H(S_v)$
 $H(S) = 0.91$

f_1 6Y/2N f_2 3Y/3N f_3 1Y/1N
 $H(f_1) = 0.91$ $H(f_2) = 0.91$ $H(f_3) = 1$

$Gain(S, f_1)$
 $= H(S) - \frac{8}{14} H(f_1) - \frac{6}{14} H(f_2) - \frac{1}{14} H(f_3)$
 $= 0.91 - \frac{8}{14} \times 0.91 - \frac{6}{14} \times 0.91 - \frac{1}{14} \times 1$
 $= 0.049$

$H(S) = -P_+ \log_2 P_+ - P_- \log_2 P_-$
 $= -\frac{8}{14} \log_2 \frac{8}{14} - \frac{6}{14} \log_2 \frac{6}{14}$

11:11 / 12:39

Gini Impurity intuition in depth in Decision Tree:

Tutorial 39- Gini Impurity Intuition In Depth In Decision Tree

GINI IMPURITY DT

f_1	f_2	f_3	q/p
G_1	D_1		Yes
G_2	D_2		Yes
			No
			No
			Yes
			...

① Entropy

$$H(S) = -P_1 \log_2 P_1 - P_2 \log_2 P_2$$



② GINI IMPURITY

$$GI = 1 - \sum_{i=1}^n (P_i)^2$$

$$= 1 - [(P_1)^2 + (P_2)^2]$$

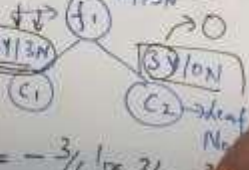
Tutorial 39- Gini Impurity Intuition In Depth In Decision Tree

GINI IMPURITY DT

f_1	f_2	f_3	q/p
G_1	D_1		Yes
G_2	D_2		Yes
			No
			No
			Yes
			...



$$H(S) = -P_1 \log_2 P_1 - P_2 \log_2 P_2$$



$$GI = 1 - \sum_{i=1}^n (P_i)^2$$

$$= 1 - [(P_1)^2 + (P_2)^2]$$

$$= 1 - [(3/6)^2 + (3/6)^2]$$

$$= 1 - [0.25 + 0.25]$$

$$= 0.5$$

$$= -3/6 \log_2 3/6 - 3/6 \log_2 3/6$$

$$= 1$$