

Nama : Ivan Daniar A. P

NIM : 1103204146

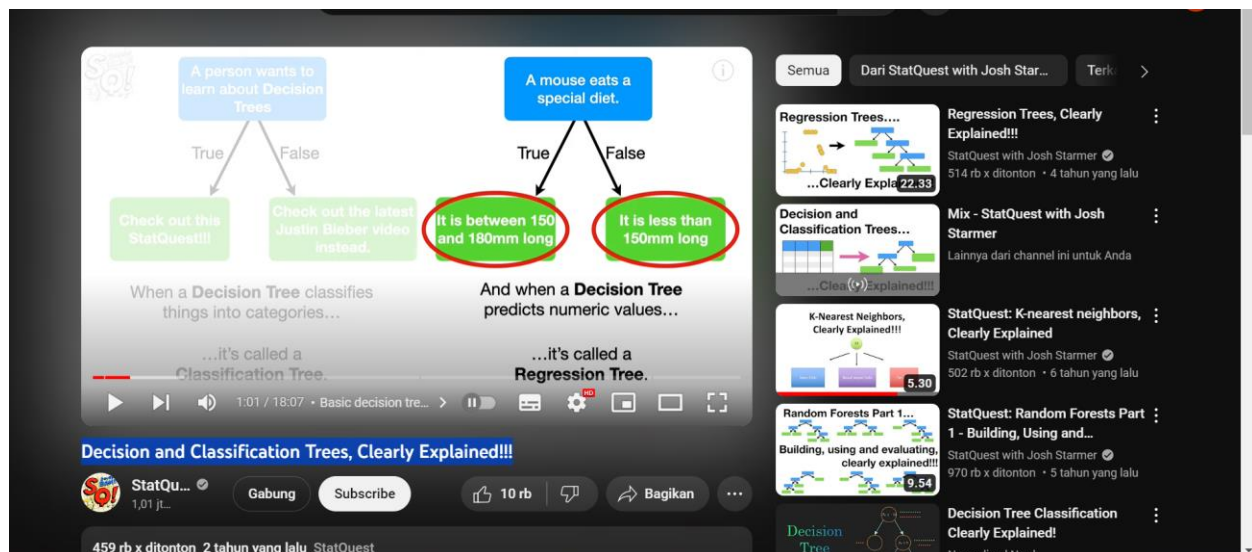
Machine Learning

Vid 1 : https://youtu.be/_L39rN6gz7Y?si=UVpUKh20PSuLM0mK

Decision and Classification Trees, Clearly Explained!!!



Example for a simple Decision Tree



When a decision tree classifies things into categories, its called a classification tree and when a decision tree predict numeric values its called a regression tree.

One of the most popular methods is called **Gini Impurity**, but there are also fancy sounding methods like **Entropy** and **Information Gain**.

Decision and Classification Trees, Clearly Explained!!!

StatQu... 1,01 jL...

Gabung Subscribe

10 rb

Bagikan

459 rb x ditonton 2 tahun yang lalu StatQuest

Semua Dari StatQuest with Josh Star... Terk...

Regression Trees... Clearly Explained!!! StatQuest with Josh Starmer 514 rb x ditonton · 4 tahun yang lalu ...Clearly Expla 22.33

Decision and Classification Trees... Mix - StatQuest with Josh Starmer Lainnya dari channel ini untuk Anda ...Clea(✓)explained!!!

K-Nearest Neighbors, Clearly Explained!!! StatQuest: K-nearest neighbors, Clearly Explained StatQuest with Josh Starmer 502 rb x ditonton · 6 tahun yang lalu 5.30

Random Forests Part 1... StatQuest: Random Forests Part 1 - Building, Using and... StatQuest with Josh Starmer 970 rb x ditonton · 5 tahun yang lalu 9.54

Decision Tree Classification Clearly Explained! Normalized Nerd

Most popular methods is gini impurity

Gini Impurity = 0.375

Gini Impurity for a Leaf = $1 - ((\text{the probability of "Yes"})^2 - (\text{the probability of "No"})^2)$

$$= 1 - \left(\frac{1}{1+3}\right)^2 - \left(\frac{3}{1+3}\right)^2$$

= 0.375

So let's put 0.375 under the Leaf on the left so we don't forget it.

Decision and Classification Trees, Clearly Explained!!!

StatQu... 1,01 jL...

Gabung Subscribe

10 rb

Bagikan

459 rb x ditonton 2 tahun yang lalu StatQuest

Semua Dari StatQuest with Josh Star... Terk...

Regression Trees... Clearly Explained!!! StatQuest with Josh Starmer 514 rb x ditonton · 4 tahun yang lalu ...Clearly Expla 22.33

Decision and Classification Trees... Mix - StatQuest with Josh Starmer Lainnya dari channel ini untuk Anda ...Clea(✓)explained!!!

K-Nearest Neighbors, Clearly Explained!!! StatQuest: K-nearest neighbors, Clearly Explained StatQuest with Josh Starmer 502 rb x ditonton · 6 tahun yang lalu 5.30

Random Forests Part 1... StatQuest: Random Forests Part 1 - Building, Using and... StatQuest with Josh Starmer 970 rb x ditonton · 5 tahun yang lalu 9.54

Decision Tree Classification Clearly Explained! Normalized Nerd

How to calculate to gini impurity by some example leaf

How To Prune... Regression and Decision Trees!!!

One method is called **Pruning**, and there's a whole **StatQuest** dedicated to it, so check it out.

Decision and Classification Trees, Clearly Explained!!!

StatQuest with Josh Starmer

459 rb x ditonton 2 tahun yang lalu StatQuest

Another method is pruning

Vid 2 : <https://youtu.be/HVXimeOnQeI?si=xv4s2Kk8VAJLmtEH>

StatQuest: K-nearest neighbors, Clearly Explained

The K-Nearest Neighbors Algorithm

- A super simple way classify data.

If you already had a lot of data that defined these cell types...

Stem Cells Blood Vessel Cells Fat Cells

We could use it to decide which type of cell this guy is...

StatQuest: K-nearest neighbors, Clearly Explained

StatQuest with Josh Starmer

502 rb x ditonton 6 tahun yang lalu StatQuest

K-nearest neighbors Algorithm is a super simple way classify data

Step 1: Start with a dataset with known categories. In this case, we have different cell types from an intestinal tumor. Then cluster that data. In this case, we used PCA.



StatQuest: K-nearest neighbors, Clearly Explained

Semua Dari StatQuest with Josh Star... Terkait

Naive Bayes... Naive Bayes, Clearly Explained!!!
StatQuest with Josh Starmer
825 rb x ditonton • 3 tahun yang lalu

Decision and Classification Trees... Decision and Classification Trees, Clearly Explained!!!
StatQuest with Josh Starmer
459 rb x ditonton • 2 tahun yang lalu

K-Nearest Neighbors, Clearly Explained!!! Mix - StatQuest with Josh Starmer
Lainnya dari channel ini untuk Anda

K Nearest Neighbor classification with Intuition an...
Krish Naik
137 rb x ditonton • 4 tahun yang lalu

PCA... StatQuest: Principal Component Analysis (PCA), Step-by-Step
StatQuest with Josh Starmer

502 rb x ditonton 6 tahun yang lalu StatQuest

There is 3 step to classify the data. Step 1 start with a dataset with known categories.

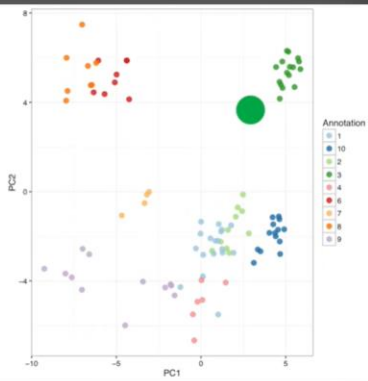
Step 3: We classify the new cell by looking at the nearest annotated cells. (i.e. the "nearest neighbors").

If the "K" in "K-nearest neighbors" is equal to 1, then we only use the nearest neighbor to define the category.

In this case, the category is **GREEN**.

If K=11, we would use the 11 nearest neighbors.

In this case, the category is still **GREEN**.



StatQuest: K-nearest neighbors, Clearly Explained

Semua Dari StatQuest with Josh Star... Terkait

Naive Bayes... Naive Bayes, Clearly Explained!!!
StatQuest with Josh Starmer
825 rb x ditonton • 3 tahun yang lalu

Decision and Classification Trees... Decision and Classification Trees, Clearly Explained!!!
StatQuest with Josh Starmer
459 rb x ditonton • 2 tahun yang lalu

K-Nearest Neighbors, Clearly Explained!!! Mix - StatQuest with Josh Starmer
Lainnya dari channel ini untuk Anda

K Nearest Neighbor classification with Intuition an...
Krish Naik
137 rb x ditonton • 4 tahun yang lalu

PCA... StatQuest: Principal Component Analysis (PCA), Step-by-Step
StatQuest with Josh Starmer

502 rb x ditonton 6 tahun yang lalu StatQuest

And step 3 classify the new cell by looking at the nearest annotated cells.

If $K=11$ and the new cell is between two (or more) categories, we simply pick the category that "gets the most votes".

In this case....

7 nearest neighbors are **RED**.
 3 nearest neighbors are **ORANGE**.
 1 nearest neighbor is **GREEN**.

Since **RED** got the most votes, the final assignment is **RED**.



StatQuest: K-nearest neighbors, Clearly Explained

StatQu... 1,01 jL... Gabung Subscribe 9,5 rb Bagikan ...

502 rb x ditonton 6 tahun yang lalu StatQuest

Semua Dari StatQuest with Josh Star... Terkait

Naive Bayes... Naive Bayes, Clearly Explained!!!
StatQuest with Josh Starmer 825 rb x ditonton · 3 tahun yang lalu

Decision and Classification Trees... Decision and Classification Trees, Clearly Explained!!!
StatQuest with Josh Starmer 459 rb x ditonton · 2 tahun yang lalu

K-Nearest Neighbors, Clearly Explained!!! Mix - StatQuest with Josh Starmer
Lainnya dari channel ini untuk Anda

K Nearest Neighbor classification with Intuition an... Krish Naik 137 rb x ditonton · 4 tahun yang lalu

PCA... StatQuest: Principal Component Analysis (PCA), Step-by-Step
StatQuest with Josh Starmer

And the final assignment is red.

A few thoughts on picking a value for "K"

- There is no physical or biological way to determine the best value for "K", so you may have to try out a few values before settling on one. Do this by pretending part of the training data is "unknown".
- Low values for K (like $K=1$ or $K=2$) can be noisy and subject to the effects of outliers.
- Large values for K smooth over things, but you don't want K to be so large that a category with only a few samples in it will always be out voted by other categories.

StatQuest: K-nearest neighbors, Clearly Explained

StatQu... 1,01 jL... Gabung Subscribe 9,5 rb Bagikan ...

502 rb x ditonton 6 tahun yang lalu StatQuest

Semua Dari StatQuest with Josh Star... Terkait

Naive Bayes... Naive Bayes, Clearly Explained!!!
StatQuest with Josh Starmer 825 rb x ditonton · 3 tahun yang lalu

Decision and Classification Trees... Decision and Classification Trees, Clearly Explained!!!
StatQuest with Josh Starmer 459 rb x ditonton · 2 tahun yang lalu

K-Nearest Neighbors, Clearly Explained!!! Mix - StatQuest with Josh Starmer
Lainnya dari channel ini untuk Anda

K Nearest Neighbor classification with Intuition an... Krish Naik 137 rb x ditonton · 4 tahun yang lalu

PCA... StatQuest: Principal Component Analysis (PCA), Step-by-Step
StatQuest with Josh Starmer

A few thoughts on picking a value for "K"

Vid 3 : https://youtu.be/_UVHneBUBW0?si=5pP44ggzTWGMQ35h

Principal Component Analysis (PCA) clearly explained (2015)

Background: An Introduction to Dimensions

- This is going to seem very, very simple.
- Just hang in there, you'll be glad we did this.
 - It will keep your head from exploding.

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jt...

Gabung Subscribe 13 rb Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

Semua Dari StatQuest with Josh Starmer Terkait

Linear Regression Clearly Explained!!! 27:27

StatQuest: Linear Regression, Clearly Explained!!! StatQuest with Josh Starmer 1,2 jt x ditonton • 6 tahun yang lalu

PCA... Step-by-Step 21:58

StatQuest: Principal Component Analysis (PCA), Step-by-Step StatQuest with Josh Starmer 2,4 jt x ditonton • 5 tahun yang lalu

Mix - StatQuest with Josh Starmer Lainnya dari channel ini untuk Anda

Decision and Classification Trees... Clearly Explained!!! 18:08

StatQuest: Decision and Classification Trees, Clearly Explained!!! StatQuest with Josh Starmer 459 rb x ditonton • 2 tahun yang lalu

PCA Main Ideas...

StatQuest: PCA main ideas in only 5 minutes!!! StatQuest with Josh Starmer

Introduction to dimension

Dimensions So Far...

- 1 cell = 1-D graph (number line)
- 2 cells = 2-D graph (normal x/y graph)
- 3 cells = 3-D graph (fancy graph with depth)
- 4 cells = 4-D graph (you can't draw it)
- 200 cells = 200-D graph (etc..)

Are all those dimensions super important? Or are some more important than others?

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jt...

Gabung Subscribe 13 rb Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

Semua Dari StatQuest with Josh Starmer Terkait

Linear Regression Clearly Explained!!! 27:27

StatQuest: Linear Regression, Clearly Explained!!! StatQuest with Josh Starmer 1,2 jt x ditonton • 6 tahun yang lalu

PCA... Step-by-Step 21:58

StatQuest: Principal Component Analysis (PCA), Step-by-Step StatQuest with Josh Starmer 2,4 jt x ditonton • 5 tahun yang lalu

Mix - StatQuest with Josh Starmer Lainnya dari channel ini untuk Anda

Decision and Classification Trees... Clearly Explained!!! 18:08

StatQuest: Decision and Classification Trees, Clearly Explained!!! StatQuest with Josh Starmer 459 rb x ditonton • 2 tahun yang lalu

PCA Main Ideas...

StatQuest: PCA main ideas in only 5 minutes!!! StatQuest with Josh Starmer

Dimension so far

Summary of Dimensions

- Each cell we sequence adds another “dimension”
- Some dimensions are more important than others...

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jt...

Gabung Subscribe

13 rb

Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

Semua Dari StatQuest with Josh Star... Terkait

Linear Regression Clearly Explained!!! 27.27

StatQuest: Principal Component Analysis (PCA), Step-by-Step 21.58

Mix - StatQuest with Josh Starmer

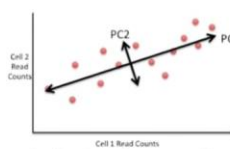
Decision and Classification Trees... Clearly Explained!!! 18.08

PCA Main Ideas...

Summary for example TV and Movies

General ideas so far...

- For each gene, we plotted a point based on how many reads were from each cell.



- PC1 captures the direction where most of the variation is.
- PC2 captures the direction with the 2nd most variation.

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jt...

Gabung Subscribe

13 rb

Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

Semua Dari StatQuest with Josh Star... Terkait

Linear Regression Clearly Explained!!! 27.27

StatQuest: Principal Component Analysis (PCA), Step-by-Step 21.58

Mix - StatQuest with Josh Starmer

Decision and Classification Trees... Clearly Explained!!! 18.08

PCA Main Ideas...

General ideas for PCA

Using the two Principle Components to plot cells
Combining the read counts for all genes in a cell to get a single value.

The original read counts

Gene	Cell1	Cell2
a	10	8
b	0	2
c	14	10
d	33	45
e	50	42
f	80	72
g	95	90
h	44	50
i	60	50
etc	etc	etc

Gene	Influence on PC1	In numbers
a	high	10
b	low	0.5
c	low	0.2
d	low	-0.2
e	high	13
f	high	-14
g	---	---
h	---	---
i	---	---
etc	etc	etc

Gene	Influence on PC2	In numbers
a	medium	3
b	high	10
c	high	8
d	high	-12
e	low	0.2
f	low	-0.1
g	---	---
h	---	---
i	---	---
etc	etc	etc

Cell1 PC1 score = $(10 * 10) + (0 * 0.5) + \dots \text{etc} = 12$

Cell1 PC2 score = $(10 * 3) + (0 * 10) + \dots \text{etc} = 6$

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jt... Gabung Subscribe 13 rb Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

Using two principle components to plot cells

Hooray! We know how they plotted all of the cells!!!

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jt... Gabung Subscribe 13 rb Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

They plotted all of the cells

How to identify key genes.

See how the cells are spread out left/right, above/below?

If we wanted to find out which genes had a big influence in putting dermal cells on the left and neural cells on the right, we could look at the influence scores in PC1.

Principal Component Analysis (PCA) clearly explained (2015)

StatQu... 1,01 jL...

Gabung Subscribe

13 rb

Bagikan

969 rb x ditonton 8 tahun yang lalu High Throughput Sequencing

Semua Dari StatQuest with Josh Star... Terkait

Linear Regression Clearly Explained!!! 27.27

PCA... Step-by-Step 21.58

Decision and Classification Trees... Clearly Explained!!! 18.08

PCA Main Ideas...

Linear Regression, Clearly Explained!!! StatQuest with Josh Starmer 1,2 jt x ditonton • 6 tahun yang lalu

StatQuest: Principal Component Analysis (PCA), Step-by-Step StatQuest with Josh Starmer 2,4 jt x ditonton • 5 tahun yang lalu

Mix - StatQuest with Josh Starmer Lainnya dari channel ini untuk Anda

Decision and Classification Trees, Clearly Explained!!! StatQuest with Josh Starmer 459 rb x ditonton • 2 tahun yang lalu

StatQuest: PCA main ideas in only 5 minutes!!! StatQuest with Josh Starmer

How to identify key genes