

# **UAS - SISTEM CERDAS**

Kelompok 1

# TUGAS 7

PEMBUATAN MACHINE LEARNING MENGGUNAKAN  
NAIVE BAYES

# STUDI KASUS

Klasifikasi buah pisang, termasuk kedalam kategori bagus atau tidak.

# ALGORITMA

Algoritma Naive Bayes adalah algoritma klasifikasi berbasis teori probabilitas Bayes yang digunakan dalam pembelajaran mesin (machine learning).

# DATASET

Pada dataset terdapat 7 karakter pisang yang bisa digunakan sebagai variable dalam input (independen/X), yang akan memprediksi hasil label output (dependen/Y), yaitu Quality.

# DATASET

- DATASET ORIGINAL

```
1 Size,Weight,Sweetness,Softness,HarvestTime,Ripeness,Acidity,Quality
2 -1.9249682,0.46807805,3.0778325,-1.4721768,0.2947986,2.4355695,0.27129033,Good
3 -2.4097514,0.48686993,0.34692144,-2.4950993,-0.8922133,2.0675488,0.30732512,Good
4 -0.3576066,1.4831762,1.5684522,-2.6451454,-0.64726734,3.0906434,1.427322,Good
5 -0.8685235,1.5662014,1.8896049,-1.2737614,-1.0062776,1.8730015,0.47786173,Good
6 0.65182525,1.3191992,-0.022458995,-1.2097088,-1.430692,1.0783454,2.8124418,Good
7 -2.8077223,1.1381357,3.4476268,-1.7133021,-2.2209115,2.07941,2.2812028,Good
8 -0.23020804,2.7834713,1.6811839,-0.52977854,-1.9584678,1.3481431,2.1817663,Good
9 -1.3485153,3.2322812,4.0118165,-0.89060634,-0.031993963,2.3959174,1.0428779,Good
10 -2.0122256,1.9280338,0.69874644,-0.95977193,-1.3497207,1.3118018,1.048762,Good
11 0.05303478,1.3099926,-0.26413944,-2.9692972,0.30398348,3.8893588,1.9313319,Good
12 -1.5878931,1.4466584,0.13883868,-2.8589776,-1.7612225,2.4337273,1.7631273,Good
13 -2.0318587,1.3767682,2.2734401,-2.2057126,0.41215587,3.757726,3.1004484,Good
```

# DATASET

- DATASET TRAINING

```
Size,Weight,Sweetness,Softness,HarvestTime,Ripeness,Acidity,Quality
-2.8077223,1.1381357,3.4476268,-1.7133021,-2.2209115,2.07941,2.2812028,Good
-0.23020804,2.7834713,1.6811839,-0.52977854,-1.9584678,1.3481431,2.1817663,Good
-1.3485153,3.2322812,4.0118165,-0.89060634,-0.031993963,2.3959174,1.0428779,Good
-2.0122256,1.9280338,0.69874644,-0.95977193,-1.3497207,1.3118018,1.048762,Good
0.05303478,1.3099926,-0.26413944,-2.9692972,0.30398348,3.8893588,1.9313319,Good
-1.5878931,1.4466584,0.13883868,-2.8589776,-1.7612225,2.4337273,1.7631273,Good
-2.0318587,1.3767682,2.2734401,-2.2057126,0.41215587,3.757726,3.1004484,Good
2.0621698,1.5615295,2.5703752,-1.8647795,-0.44219488,2.2499652,4.70331,Good
-0.8612225,1.4269999,3.349791,-0.022131119,-1.8131022,-0.07608548,0.37977022,Good
-4.1782064,-1.4760928,-3.2664933,-5.056084,-3.8299162,4.167432,-0.4254814,Good
0.6720202,0.5424222,0.00204250,2.0155004,0.22510462,2.2022051,0.2020200,Good
```

# DATASET

- DATASET TESTING

Size	Weight	Sweetness	Softness	HarvestTime	Ripeness	Acidity
-1.9249682	0.46807805	3.0778325	-1.4721768	0.2947986	2.4355695	0.27129033
-1.3120325	1.0096532	1.1234508	-2.1500666	-2.495496	1.7741758	0.25558656
-2.4097514	0.48686993	0.34692144	-2.4950993	-0.8922133	2.0675488	0.30732512
-0.3576066	1.4831762	1.5684522	-2.6451454	-0.64726734	3.0906434	1.427322
-1.3120325	1.0096532	1.1234508	-2.1500666	-2.495496	1.7741758	0.25558656
-0.8685235	1.5662014	1.8896049	-1.2737614	-1.0062776	1.8730015	0.47786173
-3.4158523	0.66085887	-1.2913183	-3.0183363	-2.2248836	2.830545	-1.4335482
0.65182525	1.3191992	-0.022458995	-1.2097088	-1.430692	1.0783454	2.8124418

# OUTPUT

- OUTPUT

Pada label yang dihasilkan dari beberapa karakteristik buah pisang/variable maka output akan dilabel kan menjadi 2 kategori yaitu:

1. **Good** (Pisang berkualitas baik)
2. **Bad** (Pisang berkualitas buruk)



# OUTPUT

- OUTPUT DATA TESTING

```
Hasil Prediksi:  
Data 1: Good  
Data 2: Bad  
Data 3: Good  
Data 4: Good  
Data 5: Bad  
Data 6: Good  
Data 7: Bad  
Data 8: Good
```

# CODE

Code dari program ini dapat dilihat di screen yang akan ditampilkan.

# TUGAS 9

Implementasi algoritma K-Means Clustering yang dirancang untuk menganalisis dan mengelompokkan data ke dalam sejumlah cluster

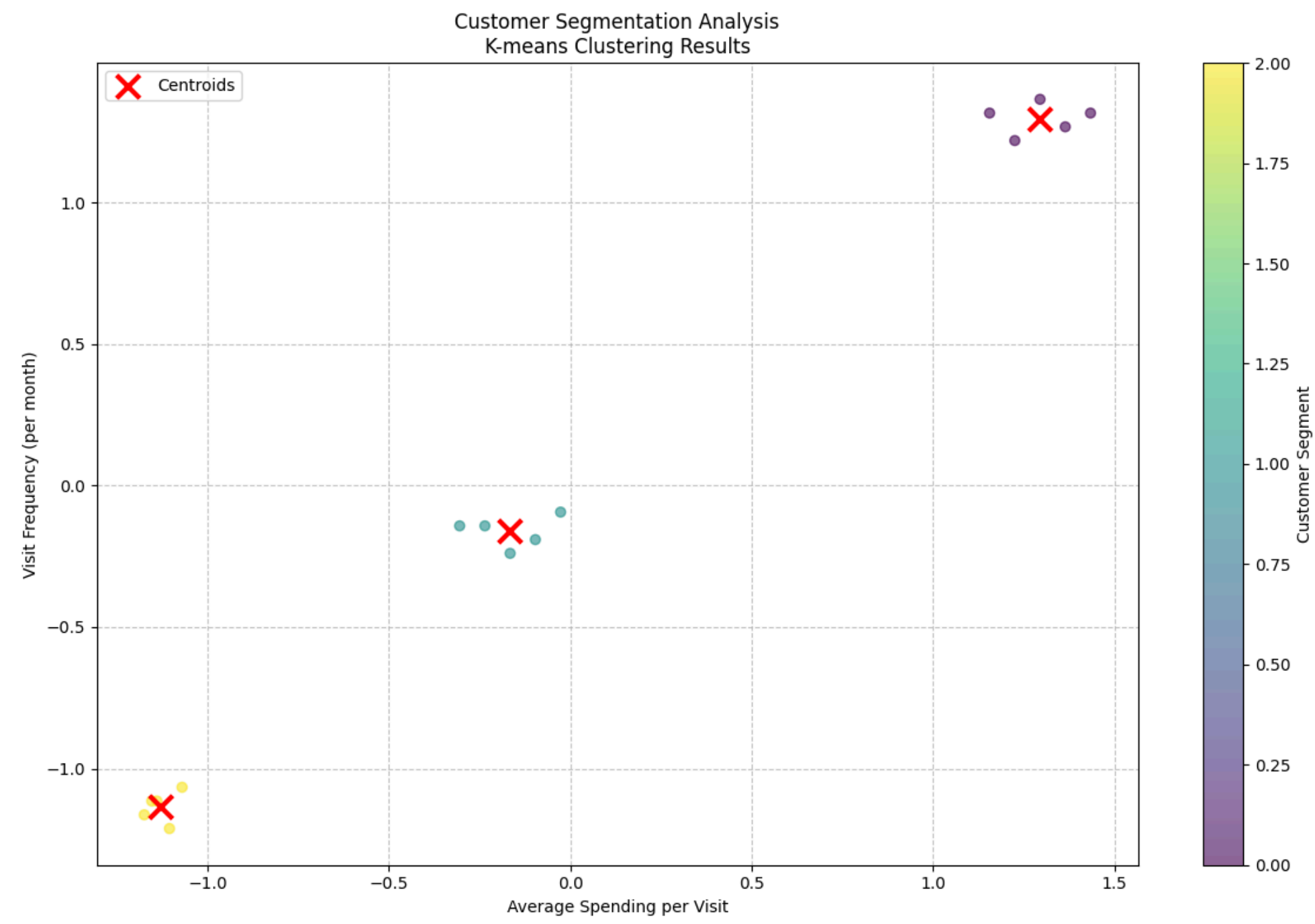
# TUJUAN

Tujuan utama program ini adalah membantu pengguna dalam memisahkan data menjadi beberapa kelompok berdasarkan pola atau karakteristik tertentu.

# SAMPLE DATA

average_spending	visit_frequency
120	3.2
115	3.1
125	3.0
130	3.3
118	3.2
450	8.2
480	8.1
460	8.0
470	8.3
490	8.2
250	5.2
270	5.1
260	5.0
280	5.3
240	5.2

# HASIL



# CODE

Code dari program ini dapat dilihat di screen yang akan ditampilkan.

# TUGAS 10

PENDETEKSI ANOMALI (ANOMALY DETECTION) dengan  
Unsupervised Learning (Isolation Forest)



# TUJUAN

Deteksi anomali (anomaly detection) adalah salah satu jenis tugas dalam machine learning yang bertujuan untuk mengidentifikasi data atau pola yang menyimpang secara signifikan dari mayoritas data lainnya. Data yang menyimpang ini disebut sebagai anomali, outlier, atau abnormal.

# DATASET

- DATASET ORIGINAL

Dataset original dapat diakses secara publik pada kaggle.

<https://www.kaggle.com/datasets/krishd123/log-data-for-anomaly-detection>

# DATASET

- DATASET TRAINING

```
081110 221444 33 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.250.13.240:50010 is added to blk_-6338935086151592174 size 67108864,0.0
081110 211832 14757 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_-3486529834145466157 terminating,0.0
081110 210135 28 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-597126776337351848 is added to invalidSet of 10.251.31.85:50010,1.0
081111 084116 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.42.191:50010 is added to blk_-4986903601822694664 size 67108864,0.0
081111 031036 18191 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-4004403128765485825 terminating,0.0
081111 105407 26900 INFO dfs.DataNode$PacketResponder: PacketResponder 2 for block blk_-233864572521798916 terminating,0.0
081111 080806 24022 INFO dfs.DataNode$DataXceiver: Receiving block blk_-2680543682853942128 src: /10.251.111.209:45920 dest: /10.251.111.209:50010,0.0
081110 222405 16115 INFO dfs.DataNode$DataXceiver: Receiving block blk_-5414362684456468711 src: /10.251.75.143:43739 dest: /10.251.75.143:50010,0.0
081111 092200 25274 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3198110276492484939 src: /10.251.29.239:45413 dest: /10.251.29.239:50010,0.0
081110 103745 19 INFO dfs.FSNamesystem: Deleting block blk_-6596276709178665528 file /mnt/hadoop/dfs/data/current/subdir0/blk_-6596276709178665528,0.0
081111 091537 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.193.175:50010 is added to blk_-7711695658484215379 size 67108864,0.0
081111 070439 22405 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_-3420753066685222136 terminating,0.0
081111 044341 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-9120005470237070778 is added to invalidSet of 10.251.107.50:50010,1.0
081111 080437 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.250.13.240:50010 is added to blk_-7761473888625136493 size 67108864,0.0
081111 085909 27 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-6783506473540833076 is added to invalidSet of 10.251.194.147:50010,1.0
081110 223820 16471 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-7437893091613684055 terminating,0.0
081111 044306 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-8463956587421115412 is added to invalidSet of 10.251.39.179:50010,1.0
081109 205152 34 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.91.32:50010 is added to blk_-984155786703949688 size 67108864,0.0
081111 073427 23489 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_-396585923727413058 terminating,0.0
```

Dataset Training yang kami gunakan ada sebanyak 1.048.576 data.

# DATASET

- DATASET TESTING

```
081110 221444 33 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.250.13.240:50010 is added to blk_-6338935086151592174 size 67108864,0.0
081110 211832 14757 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_-3486529834145466157 terminating,0.0
081110 210135 28 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-597126776337351848 is added to invalidSet of 10.251.31.85:50010,1.0
081111 084116 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.42.191:50010 is added to blk_-4986903601822694664 size 67108864,0.0
081111 031036 18191 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-4004403128765485825 terminating,0.0
081111 105407 26900 INFO dfs.DataNode$PacketResponder: PacketResponder 2 for block blk_-233864572521798916 terminating,0.0
081111 080806 24022 INFO dfs.DataNode$DataXceiver: Receiving block blk_-2680543682853942128 src: /10.251.111.209:45920 dest: /10.251.111.209:50010,0.0
081110 222405 16115 INFO dfs.DataNode$DataXceiver: Receiving block blk_-5414362684456468711 src: /10.251.75.143:43739 dest: /10.251.75.143:50010,0.0
081111 092200 25274 INFO dfs.DataNode$DataXceiver: Receiving block blk_-3198110276492484939 src: /10.251.29.239:45413 dest: /10.251.29.239:50010,0.0
081110 103745 19 INFO dfs.FSNamesystem: Deleting block blk_-6596276709178665528 file /mnt/hadoop/dfs/data/current/subdir0/blk_-6596276709178665528,0.0
081111 091537 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.193.175:50010 is added to blk_-7711695658484215379 size 67108864,0.0
081111 070439 22405 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_-3420753066685222136 terminating,0.0
081111 044341 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-9120005470237070778 is added to invalidSet of 10.251.107.50:50010,1.0
081111 080437 29 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.250.13.240:50010 is added to blk_-7761473888625136493 size 67108864,0.0
081111 085909 27 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-6783506473540833076 is added to invalidSet of 10.251.194.147:50010,1.0
081110 223820 16471 INFO dfs.DataNode$PacketResponder: PacketResponder 1 for block blk_-7437893091613684055 terminating,0.0
081111 044306 35 INFO dfs.FSNamesystem: BLOCK* NameSystem.delete: blk_-8463956587421115412 is added to invalidSet of 10.251.39.179:50010,1.0
081109 205152 34 INFO dfs.FSNamesystem: BLOCK* NameSystem.addStoredBlock: blockMap updated: 10.251.91.32:50010 is added to blk_-984155786703949688 size 67108864,0.0
081111 073427 23489 INFO dfs.DataNode$PacketResponder: PacketResponder 0 for block blk_-396585923727413058 terminating,0.0
```

Dataset Testing yang kami gunakan ada sebanyak 1.048.576 data.

# OUTPUT

- OUTPUT DATA TESTING

```
Total log yang dianalisis: 2239541
Jumlah anomali terdeteksi: 1158967

Anomali Teratas:
-----
Index: 76
Skor: 0.7259
Log: 081111 110513 19 INFO dfs.FSNamesystem: BLOCK* ask 10.250.5.237:50010 to delete blk_6240211623402979493 blk_51818173222797850
-----
Index: 1758
Skor: 0.7259
Log: 081110 103316 19 INFO dfs.FSNamesystem: BLOCK* ask 10.250.14.38:50010 to delete blk_-223438534269164604 blk_85141115273658588
-----
Index: 2970
Skor: 0.7259
Log: 081110 103416 19 INFO dfs.FSNamesystem: BLOCK* ask 10.251.91.229:50010 to delete blk_-6027962917485800010 blk_808826542825825
-----
Index: 3416
Skor: 0.7259
Log: 081110 220726 19 INFO dfs.FSNamesystem: BLOCK* ask 10.251.71.193:50010 to delete blk_-1473034037204051337 blk_-70943122921603
-----
Index: 4523
Skor: 0.7259
Log: 081111 075719 19 INFO dfs.FSNamesystem: BLOCK* ask 10.251.43.115:50010 to delete blk_6231963138498022919 blk_-550458548339484
-----
```

# CODE

Code dari program ini dapat dilihat di screen yang akan ditampilkan.



**TERIMAKASIH**