

*Skripsi Fisika*

**KLASIFIKASI BUAH KURMA MENGGUNAKAN  
PEMBELAJARAN MESIN DENGAN PROGRAM RAPIDMINER**

**BETUEL NABYAL**

**H021 17 1701**



**DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR  
2023**

**HALAMAN JUDUL**

**KLASIFIKASI BUAH KURMA MENGGUNAKAN  
PEMBELAJARAN MESIN DENGAN PROGRAM RAPIDMINER**

**Diajukan Sebagai Salah Satu Syarat Untuk Memperoleh Gelar Sarjana Sains  
Pada Departemen Fisika  
Fakultas Matematika Dan Ilmu Pengetahuan Alam  
Universitas Hasanuddin**

**Oleh:**

**BETUEL NABYAL**

**H021 17 1701**

**DEPARTEMEN FISIKA  
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM  
UNIVERSITAS HASANUDDIN  
MAKASSAR**

**2023**

**HALAMAN PENGESAHAN**

**KLASIFIKASI BUAH KURMA MENGGUNAKAN  
PEMBELAJARAN MESIN DENGAN PROGRAM RAPIDMINER**

Oleh:

**BETUEL NABYAL**

**H021 17 1701**

**Makassar Februari 2023**

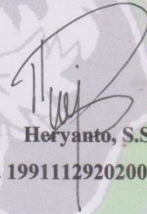
Disetujui Oleh :

**Pembimbing Utama**



**Eko Juarlin, S.Si, M.Si  
NIP. 198111062008121002**

**Pembimbing Pertama**



**Heryanto, S.Si, M.Si  
NIP. 199111292020053001**

**Ketua Program Studi,**



**Prof. Dr. Arifin, M.T  
NIP. 196705201994031002**

## PERNYATAAN

Dengan ini saya menyatakan bahwa skripsi ini merupakan karya tulis saya dalam sepanjang pengetahuan saya tidak memuat bahan yang telah dipublikasikan atau telah ditulis oleh orang lain dalam rangka tugas akhir untuk suatu gelar akademik di universitas hasanuddin atau di lembaga pendidikan lainnya dimanapun, kecuali bagian yang telah dikutip sesuai kaidah yang berlaku. Saya juga menyatakan bahwa skripsi ini merupakan hasil kerja saya sendiri dan dalam batas tertentu dibantu oleh pihak pembimbing.

Penulis



**BETUEL NABYAL**

## **ABSTRAK**

Klasifikasi Buah Kurma berhasil dilakukan. Metode yang digunakan adalah naïve bayes, K-Nearest Neighbour (KNN), dan Decision Tree. Program yang digunakan adalah RapidMiner. Ada 15 parameter masukan dan 2 kelas target. Ada 2 tipe data yaitu numerik dan polinomial. Ada 4 variasi perbandingan. Akurasi hasil penelitian sekitar 83%. Nilai akurasi berubah maksimum 1% akibat perubahan variasi. Karena ada banyak parameter dan anggota himpunan dalam parameter, nilai akurasi relatif tidak berubah terhadap variasi.

## KATA PENGANTAR

Segala Puji Syukur kepada Tuhan Yang Maha Esa atas berkat, bimbingan dan kasih karunia-NYA yang dilimpahkan kepada penulis, sehingga akhirnya penulis dapat menyelesaikan penulisan skripsi ini yang berjudul **“Klasifikasi Buah Kurma Menggunakan Pempelajaran Mesin Dengan Program RapidMiner** yang merupakan tugas akhir untuk melengkapi persyaratan mencapai gelar Sarja Fisika Jurusan Fisika Fakultas Matematika dan Ilmu Pengetahuan Alam Universitas Hasanuddin. Puji syukur kepadah Tuhan karena berkat dan rahmat-Nya saya bisa menyelesaikan skripsi ini.

Dalam menyusun skripsi ini, penulis tidak luput dari berbagai kesulitan dan hambatan, namun atas bantuan dan dorongan dari berbagai pihak akhirnya penulisan skripsi ini dapat terselesaikan. Untuk itu, pada kesempatan ini penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya dan penghargaan yang setinggi-tingginya kepada semua pihak yang telah membantu serta mendukung penulis dalam menyusun dan menyelesaikan skripsi ini, yaitu kepada:

1. Kedua orang tua ayah **Benyemin Nabyal** dan Ibunda **Maryana kipka** serta kaka dan adik-adik yang tak henti-hentinya memberikan doa dan dukungan serta kasih sayang yang tulus sehingga penulis dapat menyelesaikan skripsi ini.
2. Bapak Prof. Dr. Arifin, M.T Selaku Ketua Departemen Fisika Kakultas MIPA Unhas
3. Bapak **Eko Juarlin, S,Si.M,Si** selaku pembimbing utama yang telah banyak meluangkan waktu dan pikiran serta memberikan motivasi, masukan dan arahnya dalam penyusun skripsi ini.

4. Bapak **Heryanto, S.Si.M,Si**, selaku dosen pembimbing dua yang telah banyak menyediakan waktu, tenaga, pemikiran dalam membimbing dan mengarahkan penulis dalam menyelesaikan skripsi ini.
5. Bapak **Prof. Dr. rer-nat Wira Baharudin Nurdin** dan ibu **Dr. Sri Dewi Astuty, S.Si.M,Si** tim penguji yang telah banyak memberikan masukan, saran serta kritiknya demi penyempurnaan skripsi ini.
6. **Dosen-dosen Pengajar** yang telah membagikan ilmunya serta memberikan bimbingan selama perkuliahan.
7. Staf Pegawai baik itu dari departemen Fisika maupun Fakultas MIPA yang telah banyak membantu penulis dalam menyelesaikan segala jenis administrasi selama menempuh study
8. Untuk teman-teman seperjuangan, rekan-rekan mahasiswa/i Jurusan (**Fisika Angkatan 2017**), selama 5 tahun yang telah kita lewati bersama, merupakan kenangan yang tak terlupakan. Ini bukanlah akhir dari perpisahan kita.
9. Terimakasih kepada **Helti Lisaria Malaipada** yang selalu kasih sayang dan selalu menemani dalam penyusunan skripsi memberi semangat dalam penulisan skripsi ini.
10. Ucapan Terimakasih kepada teman-teman **Afirmasi Unhas 2017 (Ridwan, Usman, Hery, Fredi, Marlon Alfin, Yustinus, Sem, Habel, Jimmy, Mariyes, Imba, Grace, Heni, Elsy, Sarlota, , Marlince Dan Ateng)** Yang selalu menemani saya dari awal sampai selesai skripsi ini, walaupun mereka punya kesibukan masing-masing namun selalu ada untuk saya membantu menyelesaikan skripsi ini.

12. Keluarga **Besar Ikatan Mahasiswa Papua Universitas Hasanuddin (IMP-UNHAS) Makassar** yang telah memberikan dukungan moral maupun moril selalu penulis melaksanakan kuliah.

12. Untuk semua pihak yang tidak dapat disebutkan satu persatu, yang secara langsung maupun tidak langsung telah membantu penulisan skripsi ini. Penulis telah berusaha dengan sebaik mungkin dengan kemampuan yang ada dalam menyelesaikan skripsi ini untuk mendapatkan hasil yang sebaik-baiknya. Namun penulis menyadari bahwa skripsi ini masih jauh dari sempurna. Oleh karena itu, dengan segala kerendahan hati penulis sangat menghargai segala kritik dan saran yang membangun. Akhir kata, penulis berharap semoga skripsi ini dapat memberikan manfaat bagi semua pihak yang memerlukannya.

*Tuhan Memberkati Kita Semua Amin*

Makassar, ..Februari 2023

Betuel Nabyal

Penulis



## DAFTAR ISI

HALAMAN SAMPUL.....	i
HALAMAN JUDUL.....	ii
HALAMAN PENGESAHAN .....	iii
PERNYATAAN .....	iv
ABSTRAK .....	v
KATA PENGANTAR .....	vi
DAFTAR ISI .....	ix
BAB 1. PENDAHULUAN.....	1
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah .....	2
1.3 Tujuan Penelitian .....	2
1.4 Manfaat penelitian .....	2
BAB 2. TINJAUAN PUSTAKA.....	3
2.1 Kurma .....	3
2.2 Naive Bayes .....	6
2.3. K-Nearest Neighbour (KNN).....	8
2.4 Parameter Citra .....	12
2.5 Decision Tree.....	27
BAB 3. METODE PENELITIAN .....	18
3.1 Metode Pembelajaran Mesin .....	18
3.2 Analisis Hasil Penelitian.....	20
BAB 4. HASIL PENELITIAN.....	21
4.1 Metode Naïve Bayes .....	21
4.2 K-Nearest Neighbour (KNN).....	24
4.3 Decision Tree .....	26

BAB 5. KESIMPULAN DAN SARAN.....	30
5.1 Kesimpulan .....	30
4.2 Saran .....	30
DAFTAR PUSTAKA.....	32
LAMPIRAN.....	33

# **BAB I**

## **PENDAHULUAN**

### **1.1 Latar Belakang**

Kurma (*Phoenix Dactylifera L.*) adalah komoditi besar dan tanaman yang penting di daerah tandus dan panas seperti Saudi Arabia, Mesir. Di negara-negara ini, buah kurma biasa digunakan sebagai obat, kosmetik, konsumsi bagi manusia maupun hewan. Sedangkan pohon dan bagian-bagiannya, seperti pelepah kurma, biasa digunakan untuk kayu bakar maupun atap rumah. Selain di negara-negara tersebut, kurma juga terkenal di Indonesia karena cita rasa yang manis, banyak manfaatnya, dan tidak perlu repot bila ingin mengonsumsinya.

Kurma memegang peranan penting dalam perannya sebagai obat dan makanan. Kurma adalah bahan pangan yang kaya akan zat gula, vitamin, mineral, dan serat. Dalam beberapa varietasnya, kandungan zat gulanya dapat mencapai 88%. Samy Selim et. al. Melakukan studi pada dua belas varietas buah kurma guna mengetahui kandungan kimia apa saja yang terdapat di dalamnya. Salah satu kandungan buah kurma adalah senyawa flavonoid

Dalam studinya, beberapa kelas flavonoid teridentifikasi. Diantaranya adalah kuersetin, luteolin, apigenin, isokuersetin, dan rutin. Total kandungan flavonoid dari buah kurma berkisar antara 1,22 hingga 2,82 mg/100 g DW varietas Saffawy merupakan yang tertinggi dan varietas Al Qassim yang memiliki kandungan flavonoid terendah. Sedangkan varietas Ajwa Al Madinah yang 2 digunakan dalam penelitian ini merupakan varietas kurma dengan kandungan flavonoid tertinggi kedua (2,78 mg/100 g DW) tepat setelah varietas Saffawy [1].

Al-Daihan dan Bhat meneliti manfaat buah kurma (*Phoenix dactylifera L.*) secara in vitro dan menemukan bahwa salah satu manfaatnya adalah buah kurma memiliki daya antibakterial. Dari penelitian ini ditemukan bahwa daya antibakterial tersebut diduga berasal dari kandungan kurma, salah satunya adalah flavonoid. Penelitian sebelumnya telah dilakukan oleh Aryoko Widodo (2012) secara in vivo

bahwa flavonoid dapat digunakan sebagai imunomodulator untuk meningkatkan kerja makrofag dalam memfagosit mikroba patogen yang menyerang tubuh. Tubuh memiliki respon imun yang baik bila diserang oleh bakteri patogen. Salah satunya adalah dengan mekanisme fagositosis oleh makrofag. *Nitric oxide* (NO) diproduksi oleh fagolisosom pada makrofag. Ia berfungsi sebagai agen pembunuh mikroba (microbicidal agent) yang membunuh mikroba yang telah dimakan (*ingested organisms*). Produksi NO dipengaruhi oleh enzim NO-sintase yang mengkonversi L-arginin menjadi NO. Makrofag mengeluarkan enzim ini karena induksi dari berbagai macam mikroba dan stimuli sitokin. Percobaan ini menggunakan kadar NO sebagai indikator dikarenakan NO merupakan suatu zat aktif yang digunakan langsung oleh makrofag untuk melawan bakteri *pathogen*

Secara umum warna fisik kurma berbeda dan bentuk berbeda. Karena perbedaan itu, ada kesulitan membedakan jenis kurma dari warna dan bentuk. Salah satu solusi membedakan spesies kurma adaah menggunakan pembeleajaran mesin.[2]

## **1.2 Rumusan Masalah**

Karakter warna buah kurma susah dikenali dengan mata.

## **1.3 Tujuan Penelitian**

Tujuan penelitian ini adalah:

1. Mengklasifikasikan buah kurma menggunakan pembelajaran mesin untuk menentukan karakter warna buah kurma.
2. Membandingkan hasil klasifikasi dengan metode yang berbeda.

## **1.4 Manfaat Penelitian**

Blok diagram klasifikasi dapat bermanfaat bagi penelitian sejenis.

## **BAB II**

### **TINJAUAN PUSTAKA**

#### **2.1 Kurma**

Pohon kurma atau biasa disebut (*Phoenix dactylifera L.*) merupakan tanaman bersifat menahun dan diploid serta diklasifikasikan dalam genus *Phoenix*, yang merupakan anggota tunggal suku Phoeniceae, famili Palmae monokotil. *Phoenix*, yang berarti ungu atau merah dalam bahasa Yunani, mengacu pada warna buah dan “*dactylifera*” berarti jari, mengacu pada bentuk buah. *Spesies Phoenix* memiliki dua jenis cara pertumbuhan yaitu berbatang atau menggumpal. Kedua bentuk ini sangat umum ditemukan di Sudan. Pertumbuhan pohon kurma bergantung di mana pohon tersebut tumbuh. Tinggi batang pohon dapat mencapai 15 hingga 25 meter [1].

Arab Saudi dianggap sebagai negara induk dari pohon kurma dari 10.000 tahun yang lalu. Saat ini, Arab Saudi adalah produsen kurma terbesar kedua di dunia, dengan lebih dari 300 jenis kurma, masing-masing dengan rasa dan teksturnya sendiri. Al-Madinah Al-Munawarah dianggap sebagai produsen kurma pertama di Arab Saudi. Buah kurma atau (*Phoenix dactylifera L.*) kaya akan mineral dan vitamin serta merupakan bahan yang sangat baik untuk memproduksi gula rafinasi, jus pekat, pasta kembang gula dan produk fermentasi. Penelitian telah menunjukkan bahwa kurma memiliki aktivitas antioksidan yang kuat, antikanker, dan antivirus. Kurma juga mengandung *riboflavin*, *niasin*, *piridoksal*, dan *folat* dimana dalam 100 gram kurma memenuhi lebih dari 9% kebutuhan vitamin dalam sehari. Buah kurma yang telah matang juga kaya akan kandungan kalsium dan zat besi [2].

Nilai gizi manfaat dari buah kurma atau (*Phoenix dactylifera.L*) telah lama diklaim untuk konsumsi dan kesehatan manusia. Penduduk di daerah Gurun Sahara menganggap buah kurma sebagai buah yang sangat kaya akan manfaat sehingga para penduduk sekitar menganggap buah kurma sebagai makanan yang sangat penting untuk mereka. Kelembutan daging buah kurma yang manis, menjadikan buah kurma

sebagai buah yang kaya akan sumber energi. 80% kandungan dari buah kurma menjelaskan nilai energi yang tinggi. Selain itu, buah kurma juga mengandung berbagai senyawa antioksidan.

Kurma juga mengandung sukrosa, fruktosa, dan glukosa alami. Tanin dalam kurma bersifat anti infeksi dan anti inflamasi, kalsium dalam kurma sangat penting untuk pertumbuhan gigi, sedangkan kandungan zat besi dan kalium diperlukan untuk menetralkan cairan tubuh bila berada dalam kondisi asam[2].

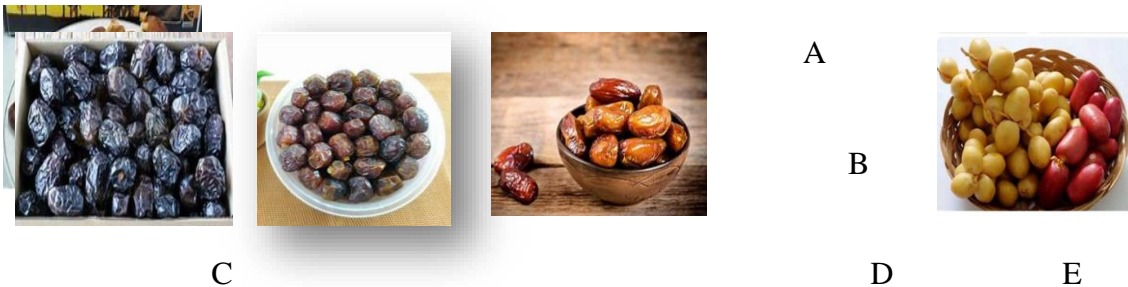
### **2.1.1 Karakteristik Buah Kurma**

Kurma (*Phoenix Dactylifera*) adalah sejenis tumbuhan palem yang buahnya dapat dimakan karena rasanya yang manis. Pohon kurma memiliki tinggi sekitar 15 – 25 meter dan daun yang menyirip dengan panjang 3 – 5 meter. Pohon kurma mulai berbuah rata-rata pada umur 5 tahun, dan memberikan hasil maksimal hanya setelah 30-40 tahun. Buah kurma adalah buah berbentuk bulat telur dan panjangnya mencapai 7 cm. Kurma berbuah dari Februari hingga Juni, dan mencapai kematangan penuh pada akhir musim gugur, ketika warna buahnya berubah dari hijau menjadi merah-kuning. Sampai mengering dan muncul warna coklat tua, dibiarkan di pohon. Karena perakaran yang dalam dan kuat, menggunakan air tanah yang dalam dan dapat bertahan hidup di oasis di daerah gurun dengan kelembaban dan curah hujan rendah, bahkan selama lebih dari 100 tahun. Buah kurma merupakan buah yang terdiri dari kulit, daging buah, lapisan dalam dan bijinya. Ada lebih dari 600 jenis kurma di seluruh dunia yang berbeda dalam bentuk, ukuran, dan sifat daging buahnya. Selain itu kurma memiliki warna, bentuk, dan tekstur yang bervariasi, tergantung pada fase pemasakan di mana buah kurma tersebut dipanen. Warna kurma juga tergantung pada jenis tanaman kurma dan kondisi iklim.

Dalam keadaan belum matang, kurma berwarna hijau. Namun selama proses pemasakan yang tergantung varietasnya, warnanya berubah dari kuning menjadi merah atau coklat. Kandungan air dan gula dalam kurma juga berbeda tergantung pada fase pematangan. Kandungan gula pada tahap awal pemasakan sekitar 20% dan jika buah telah matang kandungan karbohidrat mencapai 72-88%. Dari segi

kelembapan dan tekstur, kurma dibagi menjadi segar dan lunak, semi kering dan kering. Sekitar 90% pohon kurma ditanam di Timur Tengah dan Afrika Utara. Produsen kurma utama adalah Mesir, Arab Saudi, Iran, Aljazair, Irak, Pakistan, Sudan, Oman Uni Emirat Arab dan Tunisia.

Kurma (*Phoenix dactylifera L.*) adalah buah dari tanaman yang berasal dari keluarga Arecaceae dimana buah ini memiliki biji dengan satu lembaga (monokotil). Buah kurma mempunyai karakteristik yang sangat bervariasi, yaitu memiliki berat 2-60 gram, panjang 3-7 cm, konsistensi buah lunak sampai kering, memiliki warna kuning kecokelatan bervariasi dan berbiji [1]



Gambar 2.1

(a) Spesies kurma ajwa (b) Spesies kurma lulu (c) Spesies kurma amer haji (d) Spesies kurma palemfluit (e) Spesies kurma muda

Buah kurma mempunyai karakteristik yang sangat bervariasi, yaitu memiliki berat 2-60 gram, panjang 3-7 cm, konsistensi buah lunak sampai kering, memiliki warna kuning kecokelatan bervariasi dan berbiji. Gambar buah kurma beberapa spesies dijelaskan pada gambar 2.1.

Taksonomi Buah Kurma (*Phoenix dactylifera*) memiliki taksonomi kurma sebagai berikut:

- Kingdom : Plantae (Tumbuhan)
- Subkingdom : Tracheobionta (Tumbuhan berpembuluh)
- Super Divisi : Spermatophyta (Menghasilkan biji)
- Divisi : Magnoliophyta (Tumbuhan berbunga)
- Kelas : Liliopsida (Berkeping satu / monokotil)

Sub Kelas	: Arecidae
Ordo	: Arecales
Famili	: Arecaceae / Palmae (Suku pinang-pinangan)
Genus	: Phoenix
Spesies	: Phoenix dactylifera L

Kurma hanya bisa tumbuh di daerah Afrika Utara dan Timur Tengah. Kurma mempunyai karakteristik yang hampir sama dengan kelapa, yaitu roset batang, monokotil, diaceous (berumah dua), panjang dan bertulang, daun menyirip, bunga bentuk tandan, buah warna hijau dan ketika tua berubah menjadi merah kecoklatan [2].

## 2.2 Naive Bayes

*Naive Bayes Classifier* merupakan sebuah metoda klasifikasi yang berakar pada teorema *Bayes*. Metode pengklasifikasian dengan menggunakan metode probabilitas dan statistik yg dikemukakan oleh ilmuwan Inggris Thomas Bayes, yaitu memprediksi peluang di masa depan berdasarkan pengalaman di masa sebelumnya sehingga dikenal sebagai Teorema *Bayes*. Ciri utama dari *Naive Bayes Classifier* ini adalah asumsi yg sangat kuat (*naif*) akan independensi dari masing-masing kondisi / kejadian.

Menurut Olson Delen (2008) menjelaskan *Naive Bayes* untuk setiap kelas keputusan, menghitung probabilitas dengan syarat bahwa kelas keputusan adalah benar, mengingat vektor informasi obyek. Algoritma ini mengasumsikan bahwa atribut obyek adalah independen. Probabilitas yang terlibat dalam memproduksi perkiraan akhir dihitung sebagai jumlah frekuensi dari "master" tabel keputusan.

*Naive Bayes Classifier* bekerja sangat baik dibanding dengan model classifier lainnya. Hal ini dibuktikan oleh Xhemali, Hinde Stone dalam jurnalnya "*Naive Bayes vs. Decision Trees vs. Neural Networks in the Classification of Training Web Pages*" mengatakan bahwa "*Naive Bayes Classifier* memiliki tingkat akurasi yg lebih baik dibanding model classifier lainnya".

Atribut yang digunakan dalam perhitungan naive bayes [6].



Tabel 2.2 Contoh Dataset Play Tennis (UCI Machine Learning Repository)

<b>Id</b>	<b>outlook</b>	<b>Temperature</b>	<b>Humidity</b>	<b>Windy</b>	<b>play (class)</b>
1	sunny	Hot	High	False	no
2	sunny	Hot	High	True	no
3	overcast	Hot	High	False	yes
4	rainy	Mild	High	False	yes
5	rainy	Cool	Normal	False	yes
6	rainy	Cool	Normal	True	no
7	overcast	Cool	Normal	True	yes
8	sunny	Mild	High	False	no
9	sunny	Cool	Normal	false	yes
10	rainy	Mild	Normal	false	yes
11	sunny	Mild	Normal	true	yes
12	overcast	Mild	High	true	yes
13	overcast	Hot	Normal	False	yes
14	rainy	Mild	High	True	no

Dari data, dihitung kejadian setiap parameter terhadap target yang dijelaskan dalam table..... lalu dihitung probabilitasnya seperti dalam table...

Table 2.3 frekuensi setiap atribut

	<b>Outlook</b>		<b>temperature</b>		<b>humidity</b>		<b>windy</b>		<b>play</b>			
	Yes	no	yes	no	yes	no	yes	ye	no	no		
Sunny	2	3	hot	2 3	higt	3	4	false	6	2	9	5
Overcast	4	0	mild	4 2	nprmal	6 1	true		3	3		
Rainy	3	2	hool	3 1								

Table 2.4 probabilitas setiap atribut

	Outlook		temperature		humidity		windy		play		
	Yes	no	yes	no	yes	no	yes	ye	no	no	
Sunny	2/9	3/5	hot	2/9	3/5	high	3/9	4/5	false	6/9	2/5
	9/14	5/14									
Overcast	4/9	0/5	mild	4/9	2/5	nprmal	6/9	1/5	true	3/9	3/5
Rainy	3/9	2/5	hool	3/9	1/5						

Ada data uji sebagai berikut:

Outlook sunny, temperatur cool, humidity high, windy true, play ?

Tahapan menentukan Play adalah menghitung kedekatannya dengan yes dan no

$$p_{asingmen}(C_i) = \frac{likelihood(C_i)}{\sum_{c_j \in C} Likelihood C_i}$$

$$\hat{C}_i \arg \max_{c_i \in C} p_{asingmen}(C_i)$$

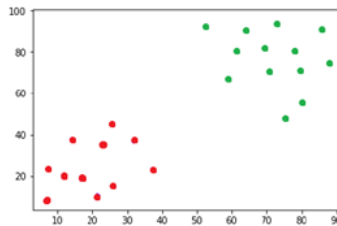
### 2.3 KNN

K-Nearest Neighbor (KNN) adalah suatu metode yang menggunakan algoritma supervised dimana hasil dari query instance yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN. Tujuan dari algoritma KNN adalah untuk mengklasifikasi objek baru berdasarkan atribut dan training samples. Dimana hasil dari sampel uji yang baru diklasifikasikan berdasarkan mayoritas dari kategori pada KNN.

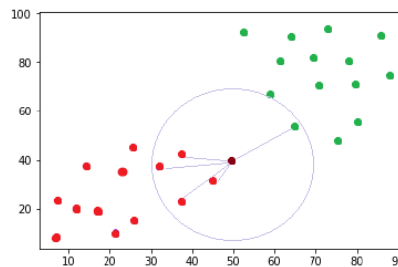
Algoritma KNN merupakan salah satu jenis supervised ML yang sering digunakan dalam aplikasi klasifikasi dan regresi. Pengklasifikasi KNN adalah pengklasifikasi berbasis instance (memori), yang berarti pengklasifikasi tersebut mengandalkan fitur serupa antar kelas (masukan) untuk memprediksi kelas titik masukan atau data baru (independen). Secara umum, Memerlukan kumpulan data

pelatihan, di mana model dilatih untuk memprediksi, dan kami mengevaluasi kinerja model pada kumpulan data independen, untuk memeriksa keakuratannya.

Bagaimana tepatnya cara kerja KKN? Nah, katakanlah kita memiliki kumpulan data (array) titik, yang perlu diklasifikasikan. Periksa gambar di bawah.



Kemudian, titik data baru ditambahkan (coklat) di lokasi (50, 40). Memeriksa nomor K dari titik terdekat ("tetangga") ke titik baru tersebut. Kita asumsikan,  $K = 5$ . Jadi kita perlu mencari 5 titik terdekat ke entri baru (titik coklat).



Jika dipilih 5, 5 titik terdekat dalam hal jarak Euclidian ditemukan (garis ungu muda), kita dapat melihat bahwa dari 5 titik, 4 di antaranya berwarna merah, oleh karena itu titik coklat kita diklasifikasikan ke sudut “merah”.

$K = 5$  menyajikan keseimbangan antara model yang berisik dengan pencilan dan kemungkinan data yang salah label (angka rendah  $K = 1$ , atau  $K = 2$ ), atau  $K$  besar, yang menyebabkan underfitting, jadi pada dasarnya, hasilnya kurang detail, atau dalam skenario terburuk, semuanya menjadi satu kelas.  $K$  yang tepat bergantung pada kumpulan data. Paling sering  $K$  adalah angka ganjil, untuk menghindari keputusan acak jika sebuah titik berjarak sama dari 2 kelas yang diketahui [7].

### **Langkah-Langkah Algoritma KNN**

Langkah-langkah pada algoritma KNN:

1. Tentukan jumlah tetangga (K) yang akan digunakan untuk pertimbangan penentuan kelas.
2. Hitung jarak dari data baru ke masing-masing data point di dataset.
3. Ambil sejumlah K data dengan jarak terdekat, kemudian tentukan kelas dari data baru tersebut.

Untuk mencari dekat atau jauhnya jarak antar titik pada kelas k biasanya dihitung menggunakan jarak Euclidean. Jarak Euclidean adalah formula untuk mencari jarak antara 2 titik dalam ruang dua dimensi. Berikut rumus untuk menghitung jarak Euclidean:

$$d = \sqrt{(x_2 - x_1)^2 + (y_2 - y_1)^2}$$

### Penjelasan Cara Kerja Algoritma KNN dengan Contoh Soal

Age	Income	Class
29	350	A
51	430	B
33	290	A
24	255	A
40	410	B
45	380	B
34	390	?

Data yang digunakan yaitu data Customer, diman harus menentukan Class dari data baru berdasarkan data-data di atasnya. Detail dari data tersebut sebagai berikut :

- Ada 2 Class yaitu A dan B
- Age dan Income adalah independent variables atau variabel yang nilainya tidak dipengaruhi oleh variabel lain dan akan digunakan untuk menghitung jarak
- Class merupakan dependent variable, variabel yang nilainya dipengaruhi oleh variabel lain (Age dan Income)

Selanjutnya gunakan langkah-langkah algoritma KNN untuk menentukan Class dari data terakhir.

### 1) Menentukan nilai K

Penentuan nilai K ini tidak ada rumus pastinya. Namun satu tips yang dapat dipertimbangkan, yakni jika kelas berjumlah genap maka sebaiknya nilai K-nya ganjil, sebaliknya jika kelas berjumlah ganjil maka sebaiknya nilai K-nya genap. Karena di contoh jumlah kelasnya ada 2, nilai K nya ditentukan jadi 3.

### 2) Hitung jarak antara data baru dan masing-masing data lainnya

Gunakan rumus Euclidean untuk menghitung jarak antara data baru dengan masing-masing data lainnya. Diperoleh hasil sebagai berikut :

#### Data 1

$$dis = \sqrt{(34 - 29)^2 + (390 - 350)^2} = 40.31$$

#### Data 2

$$dis = \sqrt{(34 - 51)^2 + (390 - 430)^2} = 43.46$$

#### Data 3

$$dis = \sqrt{(34 - 33)^2 + (390 - 290)^2} = 100.01$$

#### Data 4

$$dis = \sqrt{(34 - 24)^2 + (390 - 255)^2} = 135.37$$

#### Data 5

$$dis = \sqrt{(34 - 40)^2 + (390 - 410)^2} = 20.88$$

#### Data 6

$$dis = \sqrt{(34 - 45)^2 + (390 - 380)^2} = 14.87$$

3) Ambil sejumlah K data dengan jarak terdekat, kemudian tentukan kelas dari data baru tersebut.

Nilai dari K yaitu 3, sehingga perlu mengambil 3 data dengan jarak terdekat dengan data baru. Jarak yang terdekat yaitu data ke-6, data ke-5, dan data ke-1.

Data ke-	Age	Income	Class	Jarak dengan data baru
1	29	350	A	40.31
5	40	410	B	20.88
6	45	380	B	14.87

Dari ketiga data terdekat tersebut terdapat dua kelas B dan satu kelas A. Sehingga dapat disimpulkan bahwa kelas dari data uji adalah B [5]

#### 2.4 Parameter Citra

Sebuah piksel dalam citra digital ditransformasi menjadi komponen nilai warna merah (R), nilai warna biru (B), nilai warna hijau (G). Rentang nilai setiap komponen warna dari 0 – 255. Sebuah citra memiliki sangat banyak piksel yang bergantung pada ukuran piksel dalam panjang dan dalam lebar citra. Setelah ditransformasi, atribut yang menjadi karakter citra dihitung.

Atribut kelompok pertama adalah luas dan keliling objek. Dari komponen warna dengan menggunakan segmentasi, dihitung luas objek. Batas objek ditentukan dengan menggunakan operator deteksi tepi. Jumlah piksel batas dikenakan operasi erosi sehingga didapatkan garis batas. Jumlah piksel garis batas adalah keliling objek. Data luas dan keliling objek digunakan untuk menghitung kebulatan dan kekompakan. Kebulatan dihitung dengan rumus  $r = 4\pi \frac{luas}{keliling}$  dan kekompakan dihitung dengan rumus  $C = \frac{keliling^2}{luas}$ .

Atribut kelompok kedua adalah sumbu utama dan sumbu pendek. Koordinat posisi ujung objek ditentukan. Kedua ujung objek dihubungkan dengan garis lalu membentuk sumbu utama untuk garis yang panjang (M) dan sumbu pendek (m) untuk garis yang pendek. Kelonjongan objek dihitung dengan rumus  $c = \sqrt{M^2 - m^2}$ . Aspek rasio dihitung dengan membagi garis panjang dengan garis pendek.

Atribut kelompok ketiga adalah membangun parameter dari komponen warna. Nilai rata-rata komponen warna merah dihitung secara statistik dan standar deviasi warna merah dihitung secara statistik. Demikian pula untuk warna hijau dan biru. Gambar 2.1 menunjukkan bahwa luas dan keliling setiap jenis kurma berbeda-beda. Data luas dan keliling langsung dijadikan atribut. Bentuk setiap kurma bisa dikenali melalui kebulatan dan kekompakan. Penelusuran ciri dengan menggunakan atribut tersebut dilakukan dalam atribut kelompok satu yang mengidentifikasi bentuk berdasarkan luas dan keliling.

Dalam atribut kelompok dua, ciri khas objek sisi keelipsan dikaji. Pengajian diawali dari menghitung panjang sumbu utama dan panjang sumbu pendek. Sumbu utama dan sumbu pendek adalah karakter sebuah elips. Dari sumbu utama dan sumbu pendek, kelonjongan dan aspek rasio dihitung. Gambar 2.1 menunjukkan bentuk kurma berbeda-beda sehingga identifikasi keelipsan yang berawal dari panjang sumbu utama dan sumbu pendek, dapat membedakan setiap target.

Dalam atribut kelompok ketiga ciri khas objek sisi warna ditelusuri. Ini mengidentifikasi pewarnaan objek. Dibedakan dari warna-warnanya. Gambar 2.1 menunjukkan warna setiap jenis kurma berbeda sehingga pengidentifikasian warna bisa membedakan setiap target.

Semua parameter dalam penelitian ini adalah angka yang meliputi bilangan bulat dan bilangan riil. Atribut yang berupa angka langsung dihitung dalam pembelajaran mesin. Angka tidak dikelompokkan menjadi variabel polinomial.

## **2.5 Decision Tree**

Dikutip dari ([Mind Tools](#), *decision tree*) adalah diagram yang bisa membantumu memilih salah satu dari beberapa pilihan tindakan. Umumnya, (*Decision Tree*) dimulai dengan satu *node* atau simpul. Kemudian, *node* tersebut bercabang untuk menyatakan pilihan-pilihan yang ada. Selanjutnya, setiap cabang tersebut akan memiliki cabang-cabang baru. Oleh karenanya, metode ini disebut '*tree*' karena bentuknya menyerupai pohon yang

memiliki banyak cabang. Dalam *decision tree*, bisa menyusun berbagai pilihan dan menyelidiki kemungkinan hasil dari pilihan tersebut. Selain itu, kita juga bisa melihat kemungkinan risiko dan kelebihan atas setiap pilihan yang ada. Mengutip dari [Venngage](#), ada tiga elemen dalam satu *decision tree*, yaitu:

1. **root node (akar)**: tujuan akhir atau keputusan besar yang ingin diambil
2. **branches (ranting)**: berbagai pilihan tindakan
3. **leaf node (daun)**: kemungkinan hasil atas setiap tindakan

Biasanya, ada dua jenis *leaf node*, yakni yang berbentuk persegi dan lingkaran. *Leaf node* persegi menyatakan keputusan yang diambil. Sementara itu, *leaf node* lingkaran menyatakan hasil yang tidak pasti. *Decision tree* adalah metode yang biasa dilakukan untuk mengambil keputusan-keputusan informal atau sederhana.

Namun, menurut [Lucidchart](#), tidak sedikit pula yang menggunakannya untuk memprediksi hasil secara sistematis. Salah satu contohnya adalah dalam [analisis data](#).

Tabel 2.1 Contoh Dataset Play Tennis (UCI) Machine Learning Repository

Id	Outlook	Temperature	Humidity	windy	Play (Class)
1	Sunny	Hot	High	False	No
2	Sunny	Hot	High	True	No
3	Overcast	Hot	High	False	Yes
4	Rainy	Mild	High	False	Yes
5	Rainy	Cool	Normal	False	Yes
6	Rainy	Cool	Normal	True	No
7	Overcast	Cool	Normal	True	Yes
8	Sunny	Mild	High	False	No
9	Sunny	Cool	Normal	False	Yes
10	Rainy	Mild	Normal	False	Yes
11	Sunny	Mild	Normal	True	Yes
12	Overcast	Mild	High	True	Yes
13	Overcast	Hot	Normal	False	Yes
14	Rainy	Mild	High	False	No

Data dalam tabel 2.1 diolah dengan metode *Decision Tree* yang memiliki



beberapa tahapan. Menghitung entropi: Rumus Entropi

$$entropy(x) = - \sum_{i=1}^n x_i \log x_i \quad (2.1)$$

Dalam tabel 2.1, ada 9 kejadian yes dan 5 kejadian no, maka entropi dihitung sebagai berikut:

$$\begin{aligned} \text{Info}([9,5]) &= \text{entropy}\left(\left[\frac{9}{14}, \frac{5}{14}\right]\right) \\ &= -\frac{9}{14} \log\left(\frac{9}{14}\right) - \frac{5}{14} \log\left(\frac{5}{14}\right) \\ &= 0,283 \end{aligned}$$

Nilai entropi dimasukkan sebagai variabel *Information Gain*.

Atribut *temperature* memiliki dua anggota yaitu *False* sebanyak 8 kejadian dan *True* sebanyak 6 kejadian. Kejadian *False* menghasilkan *No* ada dua kejadian dan menghasilkan *Yes* ada enam kejadian. Entropi *False* dapat dihitung dapat dihitung sebagai berikut:

$$\begin{aligned} \text{Info}([2,6]) &= \text{entropy}\left(\left[\frac{2}{8}, \frac{6}{8}\right]\right) \\ &= -\frac{2}{8} \log\left(\frac{2}{8}\right) - \frac{6}{8} \log\left(\frac{6}{8}\right) \\ &= 0,244 \end{aligned}$$

Kejadian *True* menghasilkan *No* ada tiga kejadian dan menghasilkan *Yes* ada tiga kejadian. Entropi *True* dapat dihitung dapat dihitung sebagai berikut:

$$\begin{aligned} \text{Info}([3,3]) &= \text{entropy}\left(\left[\frac{3}{6}, \frac{3}{6}\right]\right) \\ &= -\frac{3}{6} \log\left(\frac{3}{6}\right) - \frac{3}{6} \log\left(\frac{3}{6}\right) \\ &= 0,301 \end{aligned}$$

*Information gain* dihitung sebagai persamaan [6.3](#), dimana  $c_i$  adalah jumlah *instance* untuk kelas ke- $i$ ,  $v$  adalah nilai atribut,  $c^v$  adalah banyaknya *Instance* ketika dicabangkan dengan nilai atribut  $v$ ,  $C^v$  adalah banyaknya *instance* kelas saat percabangan. *Information gain* dapat dimaknai sebagai pengurangan entropi karena melakukan percabangan.

$$IG. (C_1, C_2, \dots, C_N) = \text{info} (C_1, C_2, \dots, C_N) = - \sum_{vev} \frac{c^2}{\sum_{i=1}^n} \text{info} (C_1, C_2, \dots, C_N)$$

Info didefinisikan sebagai persamaan dimana  $c_i$  adalah banyaknya kejadian diklasifikasikan sebagai kelas ke- $i$  (atau secara lebih umum ke  $node-i$ )

$$\text{Info} (C_1, C_2, \dots, C_N) = \text{entropy} C_1, \frac{c_1}{\sum_j^N c_j}, C_1, \frac{c_1}{\sum_j^N c_j}, \dots C_1, \frac{c_1}{\sum_j^N c_j} \quad (1.2)$$

*Information gain* dihitung sebagai persamaan 6.3, dimana  $c_i$  adalah jumlah *instance* untuk kelas ke- $i$ ,  $v$  adalah nilai atribut,  $c^v$  adalah banyaknya *Instance* ketika dicabangkan dengan nilai atribut  $v$ ,  $C^v \times$  adalah banyaknya *instance* kelas saat percabangan. *Information gain* dapat dimaknai sebagai pengurangan entropy karena melakukan percabangan.

$$IG. (C_1, C_2, \dots, C_N) = \text{info} (C_1, C_2, \dots, C_N) = - \sum_{vev} \frac{c^2}{\sum_{i=1}^n} \text{info} (C_1, C_2, \dots, C_N) \quad (1.3)$$

*Information gain* atribut *temperature* dihitung sebagai berikut:

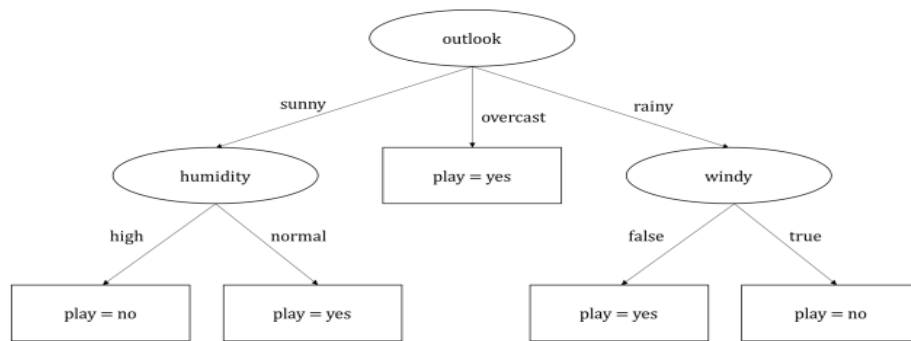
$$\begin{aligned} IG(\text{temperature}) &= \text{Info}[9,5] - \left( \frac{6}{14} \text{info}[3,3] + \frac{8}{14} \text{info}[2,6] \right) \\ &= 0,283 - \left( \frac{6}{14} 0,244 + \frac{8}{14} 0,301 \right) \\ &= 0,078 \end{aligned}$$

Dengan metode yang sama, kita hitung *information gain* untuk atribut lainnya.

- $IG(\text{outlook}) = 0,247$
- $IG(\text{humidity}) = 0.152$
- $IG(\text{windy}) = 0.048$

Karena  $IG$  outlook memiliki nilai paling besar, *Outlook* menjadi *root*.

Dengan teknik sejenis, *branches* dan *leaf* dapat dihitung. Hasil perhitungan dalam contoh di sub bab ini ada dalam gambar ...



Gambar : 2.2 *Decision Tr*