

**NAMA : ZAHRA FAUZIYAH EKAPUTRI**

**NIM : 3332200029**

**KELAS : B**

**UTS KECERDASAN BUATAN**

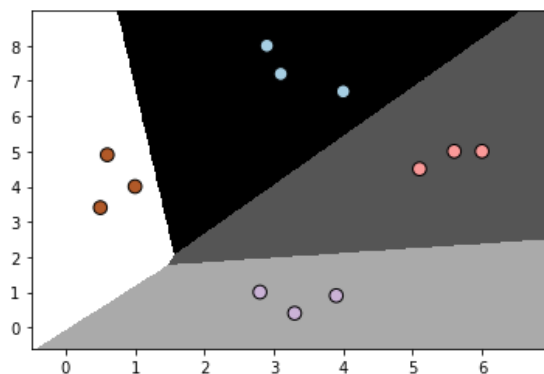
**SOAL :**

1. Analisa algoritma untuk *logistic\_regression.py*. Dan analisa algoritmanya dan jalankan di komputer anda. (Untuk Chapter 2)
2. Analisa algoritma untuk *decision\_trees.py*. Dan analisa algoritmanya dan jalankan di komputer anda. (Untuk Chapter 3)
3. Analisa algoritma untuk *mean\_shift.py*. Dan analisa algoritmanya dan jalankan di komputer anda. (untuk Chapter 4)
4. Analisa algoritma untuk *nearest\_neighbors\_classifier.py*. Dan analisa algoritmanya dan jalankan di komputer anda (untuk Chapter 5)
5. Analisa algoritma untuk *states.py*. Dan analisa algoritmanya dan jalankan di komputer anda (untuk Chapter 6)

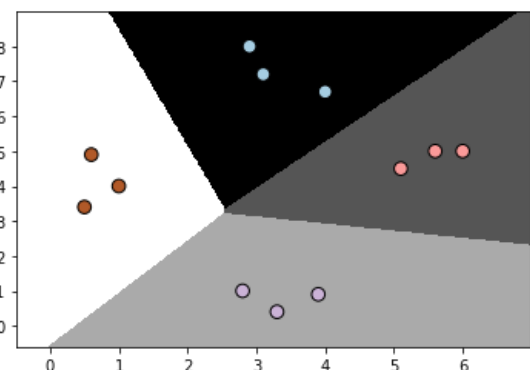
**JAWABAN :**

1. Percobaan ini termasuk pada supervised learning yang merupakan melalui proses learning yang sudah memiliki label dengan ditraining algoritmanya. Pada percobaan ini, yang awalnya berbentuk import file, dibuat visualisasi classifiernya menjadi fungsi . Dengan memasukkan librarynya terlebih dahulu dijadikan satu supaya nanti sistemnya memanggil library. Kemudian pada dasarnya logistic regression ini menggunakan nilai curve yang menjadi patokan. Pada percobaan ini menggunakan nilai curva yang harus bernilai  $> 0$ , maka dari itu menggunakan perbandingan antara kurva bernilai 1 dan 100.

Berikut adalah hasilnya :



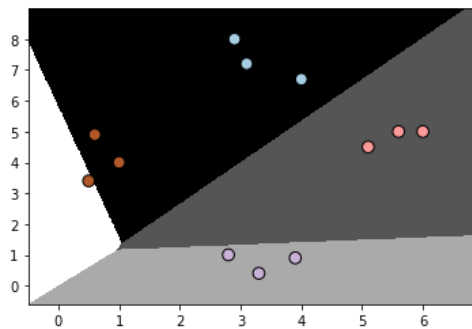
Gambar 1. Logistic Regression  $C = 1$



Gambar 2. Logistic Regression  $C=100$

Jika dibandingkan keduanya itu berbeda, kurva dengan nilai 100 akan lebih smooth dibandingkan dengan curva bernilai 1. Pada kurva ini masing-masing terdapat 4 data, yang masing-masing memiliki titik-titik berjumlah 3 yang posisinya sama tidak

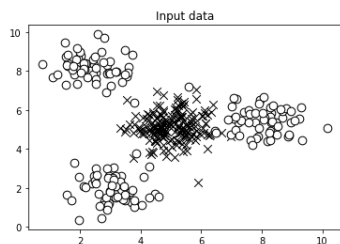
dipengaruhi oleh besarnya curva, dengan ini dapat memplot data tersebut pada area sejarusnya. Perbedaan berada di titik pertemuan kluster. Jadi dengan nilai 100 ini angkanya jauh lebih presisi bentuk kurvanya dengan pertemuan kluster yang lebih simetris dibandingkan dengan kurva 1. Pada kurva 100, ia berbentuk smooth yang terpisahkan tanpa adanya gap, berbeda dengan kurva bernilai 1 yang memiliki gap. Pada regression ini dapat menggunakan logistic dan linier regression. Dimana pada logistic regression memiliki karakteristik berbentuk sigmoid, dengan representasi data yang tersebar antara garis yang berbentuk sigmoid yang membagi 2 kluster misalnya. Sedangkan pada linier regression ia akan berbentuk activation, dengan garis lurus linear ini sebagai pemisah atau klasifikasinya, dengan sebaran datanya cenderung lebih rapi. Jika angkanya dirubah menjadi 0,3 maka akan ada data yang keluar dari kluster, hal ini maka dapat diklasifikasikan sebagai underfitted yang diakibatkan adanya model yang terlalu sederhana sehingga algoritma tak bisa membedakan dan meningkatkan kegagalan prediksi. Berikut adalah gambaranya :



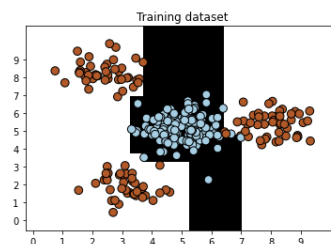
Gambar 3. Logistic Regression  $C = 0.3$

Karena disini memakai data tiruan atau bukan real, sehingga tidak diketahui nilai pasti data tersebut berasal dari mana. Jadi kesimpulannya pada percobaan ini menggunakan data yang sudah dilatih, sehingga dapat memprediksi dengan baik serta nilai kurva yang besar mempengaruhi smoothnya cluster yang dibuat, dimana cluster tersebut memiliki 4 data dengan segmentasi berbeda dan masing-masing memiliki 3 data lagi di dalamnya yang posisinya tetap. Namun ketika menggunakan data baru diluar data training, akan tinggi kegagalannya, karena tidak dapat memprediksi secara baik.

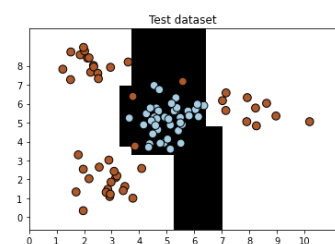
- Decision tree merupakan salah satu bagian dari supervised learning juga. Pertama yaitu tak lupa untuk menghubungkan antara google collab dengan google drive karena akan dilakukan import data pada keduanya. Pada percobaan ini menggunakan data tiruan juga, yang secara konseptual decision trees ini seperti suatu pohon dengan nilai input memiliki cabang atau 2 kondisi, dimana ketika kondisinya benar, ia akan melanjutkan prosesnya seperti layaknya logika if statement. Berikut adalah hasil percobaannya :



Gambar 4. Input Data



Gambar 5. Training Dataset



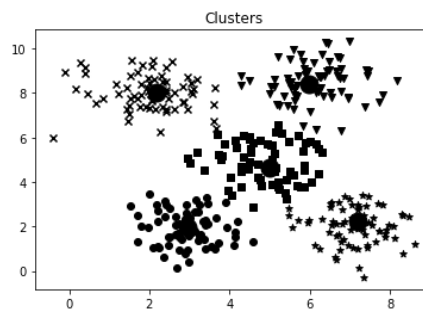
Gambar 6. Test Dataset

Pada percobaan ini terdapat komponen precision yang merupakan ketepatan data yang sesuai menmpati plot yang seharusnya , recall merupakan banyaknya data yang terpanggil kembali, f1 score merupakan nilai harmonis rata-rata, dan support yang merupakan banyaknya data. Untuk menentukan nilai pada beberapa komponen tersebut terdapat kondisi true dan false yang perlu menggunakan sebuah matrix. Dimana ketika nilai true negatif yaitu ketika nilai negatif itu bernilai 1, false positif sistem deteksinya sesuai namun seharusnya salah. Dari nilai situ dapat menentukan komponen-komponen tadi. Kemudian pada nilai harmonis rata-rata untuk menentukan apakah model tersebut bagus atau tidak. Pada nilai F1 Score ini biasanya tidak bernilai 100% karena pada setiap prediksi akan ada kegagalan seperti pada data training, sehingga ketika data testing nilainya 0,97 ia berarti sangat bagus. Sedangkan support sebagai banyaknya total data, ketika terdata ada 137, maka berarti ada 137 data training di dalamnya, sedangkan pada data testing terdapat 43. Seperti berikut :

Classifier performance on training dataset					
precision	recall	f1-score			
support					
Class-0	0.99	1.00	1.00	137	
Class-1	1.00	0.99	1.00	133	
accuracy			1.00	270	
macro avg	1.00	1.00	1.00	270	
weighted avg	1.00	1.00	1.00	270	

Classifier performance on test dataset				
precision	recall	f1-score		
support				
Class-0	0.93	1.00	0.97	43
Class-1	1.00	0.94	0.97	47
accuracy			0.97	90
macro avg	0.97	0.97	0.97	90
weighted avg	0.97	0.97	0.97	90

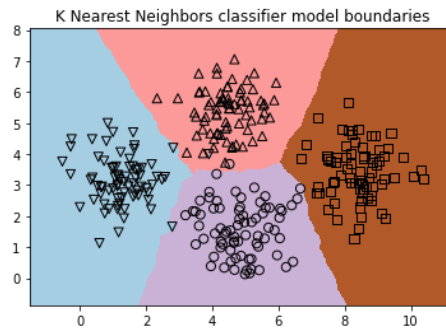
3. Percobaan ini termasuk pada unsupervised learning karena ia tidak memiliki suatu proses pelabelan, misalnya pada suatu ecommerce ketika kecenderungan sering menelusuri suatu hal, maka hal tersebut akan tersugesti terus, karena tak ada label namun dapat mengelompokkan berdasarkan kedekatan informasi yang dimiliki sehingga dapat diklasterisasikan. Pada percobaan ini terdapat 5 data berbeda yaitu persegi, segitiga, lingkaran, bintang, dan eks. Pengelompokan ini berfungsi untuk mengklasifikasikan nilai berdasarkan klasifikasinya, dengan masing-masing data memiliki keidentikan. Berikut adalah hasilnya :



Gambar 7. Pengklasteran

Dengan pada tiap kategori merupakan data region yang setiap datanya yang ada didekatnya sebagai nilai tetangganya. Klasterisasi ini melakukan pengelompokan kedekatan sesuai dengan kedekatan informasi yang dimilikinya karena tak ada pelabelan. Sebagai contoh, ketika suatu data jauh dari nilai tetangga, berarti ia sudah tidak berarti masuk ke dalam cutoff yaitu memiliki hubungan namun sudah tidak terlalu melekat, karena semakin dekat dengan pusat cluster bisa dibilang bahwa semakin mirip karena proses ini dinilai dari similaritasnya. Selanjutnya ketika data terletak antara 2 cluster, seharusnya ia lebih dekat atau lebih sesuai dengan bentuknya yang lebih mirip. Tapi, jarak pun dapat mempengaruhi kemiripannya, sehingga ketika suatu data berada di antara 2 cluster, ia sewaktu-waktu bisa masuk ke kemiripan lainnya meski bentuknya tidak mirip.

4. Percobaan ini masuk ke Supervised learning, namun bisa masuk ke unsupervised learning juga. Modelnya hampir mirip seperti clustering, namun ia memiliki data training dan terdapat labelisasi, dengan memiliki kedekatan nilai tetangga. Ketika ia memiliki data informasi aslinya maka akan diklasterisasikan, sehingga setiap informasinya akan terkelompokkan masuk ke data-data yang ada ini, data yang ada berjumlah 4 yaitu ungu, coklat, biru, dan pink. Berikut adalah gambar grafiknya :



Gambar 8. Pengklasteran dengan KNN

Misal ketika suatu nilai seharusnya berada pada region ungu, namun dia juga masuk ke ranger pink, maka ia bisa teridentifikasi menjadi pink secara kedekatan jaraknya, intinya mereka sewaktu-waktu bisa teridentifikasi sebagai kedua nilai tersebut, tetap berdasarkan jarak serta similaritas. Pada percobaan ini dapat diaplikasikan sebagai smart parking berdasarkan golongan serta pendeteksi pengenalan wajah. Area segmentasi dibuat dengan data dami maka seolah olah diketahui segmentasinya pasti, sedangkan jika pakai nilai realita, bisa tidak dapat diketahui segmentasinya seperti apa. Kemudian bentuk grafik pada KNN ini tidak dapat dibentuk simetris, berbeda dengan logistic regression bentuk nilai aktivasinya simoid yang membuat data terklasterisasi dengan jelas, sedangkan pada KNN ini grafik berdasarkan kedekatan nilai tetangganya, sehingga ia memiliki nilai desimal yang tak beraturan, sehingga segmentasi areanya tidak mungkin dapat smooth.

5. Expert Sistem, merupakan sistem pakar yang mana biasanya terdapat pada pengelompokan atau mengkualifikasian yang berdasarkan aturan. Pada percobaan ini menghubungkan antar relasi, pada dasarnya bergantung pada suatu kondisi seperti if steatment. Algoritmanya akan mengukur berdasarkan peraturan. Misal yang pertama hasilnya yaitu untuk steatment apakah nevada bersebelahan dengan lousiana? Maka klasterisasi jawabannya adalah no. Hal ini karena memang nevada tidak bersebalahan dengan lousiana pada data yang dimiliki. Selanjutnya adalah list kota yang bertetangga /bersebelahan dengan oregon, antara lain Washington, Nevada, Idaho, dan California. Kemudian pada list wilayah pesisir yang dekat dengan Mississippi yaitu Alabama dan Lousiana. Jadi pada percobaan ini terdapat data yang terbagi dua yang mana saling dihubungkan satu sama lain supaya dapat saling mengenal rule basenya. Jadi pada ini ia bermain dengan rule base dimana peraturannya sudah dibuat sebelumnya dari data yang kita miliki namun berdasarkan aturan, ketika ia bernilai nol, langsung ditolak, dan lanjut ke statement lainnya. Dengan dibentuk suatu pertanyaan dan

dibandingkan dengan rule basenya, sehingga akan mengeluarkan hasil yang sesuai dengan aturan yang dibuatnya. Kemudian pernyataan lainnya yaitu daftar 7 wilayah yang mengitari pesisir wilayah, ada North Carolina, Tennessee, Maine, Connecticut, Mississippi, Arizona, Massachusetts. Kemudian Wilayah yang sebelah dengan Arkansas dan Kentucky adalah Tennessee dan Missouri, Untuk lebih jelasnya hasil dari percobaan ini seperti berikut :

```
Is Nevada adjacent to Louisiana?:  
No
```

```
List of states adjacent to Oregon:  
Washington  
Nevada  
Idaho  
California
```

```
List of coastal states adjacent to Mississippi:  
Alabama  
Louisiana
```

```
List of 7 states that border a coastal state:  
North Carolina  
Tennessee  
Maine  
Connecticut  
Mississippi  
Arizona  
Massachusetts
```

```
List of states that are adjacent to Arkansas and  
Kentucky:  
Tennessee  
Missouri
```