****KEMENTERIAN PENDIDIKAN DAN KEBUDAYAAN

UNIVERSITAS SULTAN AGENG TIRTAYASA

**FAKULTAS TEKNIK**

Jalan Jenderal Sudirman KM 03 Cilegon, Banten 42435

Telepon 0254-395502, 376712 Faximile 0254-395540, 376712

Website : [www.ft.untirta.ac.id](http://www.ft.untirta.ac.id), e-mail : ft@untirta.ac.id

**UJIAN TENGAH SEMESTER**

|  |  |  |
| --- | --- | --- |
| Nama Mata Kuliah | : | Kecerdasan Buatan |
| Semester | : | Genap 2021/2022 SKS : 3 |
| Waktu Ujian | : | 4 April 2022 |
| Sifat | : |  |

**INSTRUKSI :**

**Kerjakan Soal yang dianggap mudah terlebih dahulu. Listing program akan diberikan pada saat ujian. Program yang digunakan adalah python versi 3. Semua file di-link ke** [**https://drive.google.com/drive/folders/1HNTZvv1skTeY4kqwGxyDuHNnf4FFlTUx?usp=sharing**](https://drive.google.com/drive/folders/1HNTZvv1skTeY4kqwGxyDuHNnf4FFlTUx?usp=sharing)

**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 dikirim ke email saya di *rian.fahrzal@untirta.ac.id*

|  |  |  |  |
| --- | --- | --- | --- |
|  | Diusulkan oleh | Diperiksa oleh | Disetujui oleh |
| Tanggal/Bulan/Tahun | 04/04/2020 |  | Ketua Prodi Teknik Elektro |
| Ttd |  |  |  |
| Nama Lengkap | Rian Fahrizal  NIP.197510262005011001 | Muhammad Iman Santoso, Phd NIP.19770130 200312 1 007 | Dr. Romi Wiryadinata, M.Eng.  NIP. 198307032009121006 |

**Nama: Mutiara Ramadhani**

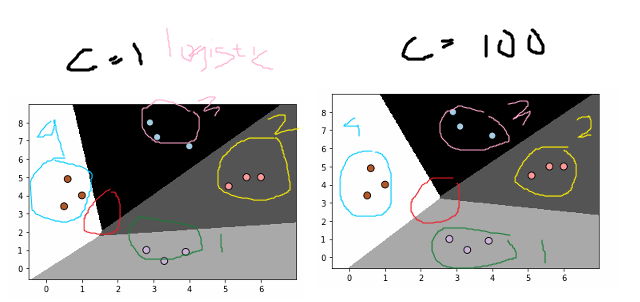
**NIM: 3332190064**

JAWABAN:

1. Logistic Regression

Membuat perbandingan pada nilai c atau nilai kurve-nya untuk model logistic regression. Karena logistic regression akan berhubungan dengan kurva, default kurvanya adalah satu atau nilai c itu nilainya harus lebih dari 0, untuk kurva perbandingannya diberikan nilai c=100, maka akan terjadi perbedaan pada bentuk klasifikasinya pada segmentasi areanya, untuk nilai c=1 dilihat kurang smooth, sedangkan untuk c=100 terlihat lebih smooth.

Apabila memakai nilai 100 bentuk kurvanya akan lebih presisi, memiliki 4 buah data yang terlabel masing-masing, bagaimana caranya kita bisa plot data tersebut diarea yang seharusnya.



Logistic regression fungsinya bukan linier regression, bentuknya seperti sigmoid atau seperti huruf s.

Linier regression bentuknya linier activation, jadi bentuknya lurus keatas.

Kalau pada linier regression antara data 1 dan data 2 dipisahkan oleh garis lurus. Logistic regression, klasterisasi datanya tersebar untuk data 1 dan 2, tidak seperti linier regression

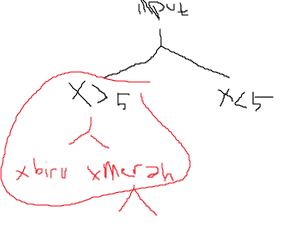


Apabila diibaratkan c=100 adalah nilai ideal dan c=1 adalah tidak ideal, maka akan terjadi perbedaan pada bentuk klasterisasinya. Saat memakai nilai 100, bentuk segmentasi areanya benar" terpisahkan areanya tanpa ada jarak yang jauh. Kalau pada nilai 1 terlihat jelas terdapat jarak pada titik tengahnya.

Saat nilai c=<1 maka data akan underfit atau terdapat data yang keluar dari area yang seharusnya, sehingga dikatakan data jelek atau tidak sesuai.

1. Decission Trees

Secara konsep untuk metode decission, step pertama akan mengambil nilai input, kemudian akan ada data latih dimana yang akan menentukan. Misal terdapat nilai x>5 dan x<5 maka input yang masuk harus x>5 sehingga akan dihasilkan kembali x biru atau x merah, apabila x merah maka akan bercabang kembali.



Pada evaluasi model ada yang namanya,

Precision adalah ketepatan data menempati plot yang sesuai dan seharusnya

Recall adalah banyaknya data yang terpanggil kembali

F1 score adalah nilai harmonic means (menentukan model apakah ideal atau tidak?)

Support adalah banyaknya data

Rumus:

Precision =

Recall =

F1 - Score =

TP= True Positif, data positif yang diprediksi true

TN= True Negatif, ketika data yang harusnya terbaca negatif tetapi ternyata terbaca true

FP= False Positif, data positif namun diprediksi sebagai data negatif

FN= False Negatif, data negatif namun diprediksi sebagai data positif

Precission merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan hasil yang diprediksi positif.

Recall merupakan rasio prediksi benar positif dibandingkan dengan keseluruhan data yang benar positif.

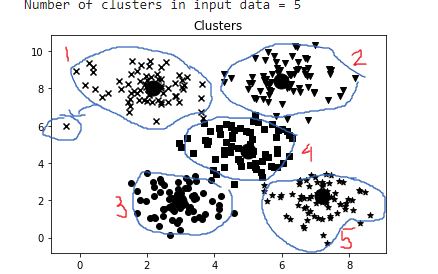
Dilihat dari hasil running algoritma, menunjukkan bahwa data F1-Score nya adalah 1 semua, dapat dikatakan model yang baik karena dasarnya menggunakan data dummy. Apabila untuk penelitian sebenarnya, angka 100% tidak bisa digunakan karena akaan menghasilkan overfitting sehingga akan terjadi kegagalan prediksi, karena nilai prediksi tidak dimungkinkan sampai 100%.

F1-Score adalah rata-rata harmonis, pada data testing dihasilkan nilai 0,97 dapat dikatakan model yang cukup baik, karena tidak dimungkinkan nilai 100% dipengujian nilai harmonis rata-rata.

Support pada data training pada class-0 ada 137 data training. Untuk data testing pada class-0 terdapat 43 data testing.

1. Data Clustering

Clustering bisa dikatakan pengelompokan data.



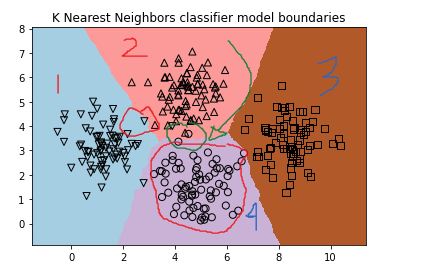
Pengelompokan 5 data, masing-masing data memiliki keidentikannya masing-masing seperti bentuknya. Fungsi clustering adalah menggolongkan atau mengklasifikasikan nilai berdasarkan kemiripannya atau keseragamannya.

Termasuk unsupervised karena tidak memiliki data label, karena tidak ada proses pelabelan. Dalam clustering ini berdasarkan kedekatan informasinya dilihat dari bentuk dari setiap areanya.

Semakin dekat dengan lingkaran tengah maka semakin mirip, apabila semakin jauh bisa dikatakan cut off dari areanya, contoh pada area dengan bentuk cross (x).

1. K-Nearest Neighbour

K-Nearest Neighbour masuk kedalam *supervised learning*. Dilihat modelnya sama hampir mirip seperti data clustering, namun perbedaannya menggunakan kedekatan nilai tetangga (*distance neighbour*).



K-Nearest Neighbour memiliki data latih, dan pelebelan. Dilihat pada gambar bisa dilihat untuk lingkaran pada region 4 berwarna ungu, dilihat antara bentuk segitiga dan lingkaran saling berdekatan jadi secara kedekatan bentuk lingkaran pada region ungu masuk ke region 2 yaitu warna pink dengan bentuk segitiga. Sehingga hasilnya segmentasi dapat terprediksi di region 4 atau region 2.

Dapat dilihat juga pada bentuk segitiga yang berada diperbatasan garis segmentasi antara region 1 dan 2, dapat dikatakan juga bahwa segitiga tersebut dapat masuk ke data region 1 dan data region 2.

1. States

Is Nevada adjacent to Louisiana?:

No

Sehingga dapat diclusterisasikan jawabannya bahwaa tidak, karena nevada tidak bersebelahan dengan louisiana.

List of states adjacent to Oregon:

Washington

California

Nevada

Idaho

Menghasilkan jawaban mengenai list negara yang berdekatan dengan oregon. Dengan pengurutan negara dari yang terdekat data arraynya.

List of coastal states adjacent to Mississippi:

Louisiana

Alabama

Data pesisir dan data negara tetangga, dari dua data tersebut saling berhubungan satu sama lain agar saling mengenali rule base berdasarkan aturan dengan bentuk pertanyaan.

List of 7 states that border a coastal state:

Maine

Tennessee

Georgia

Alabama

Louisiana

North Dakota

Connecticut

Menghasilkan jawaban dengan 7 negara yang mengelilingi negara pesisir.

List of states that are adjacent to Arkansas and Kentucky:

Missouri

Tennessee

Menghasilkan jawaban dengan list negara yang berdekatan dengan arkansas dan kentucky.