

PENERAPAN ALGORITMA REGULARIZED DISCRIMINANT ANALYSIS UNTUK KLASIFIKASI KANKER PARU

Evy Priyanti

Program Studi Sistem Informasi Akuntansi
Fakultas Teknik dan Informatika
Universitas Bina Sarana Informatika (UBSI)
Jl. Kramat Raya No 98 Jakarta Pusat
evy.evp@bsi.ac.id

Abstract

Lung cancer is a type of cancer that starts from the lungs with abnormal cell growth. Lung cancer is a type of cancer with the largest number of sufferers. In this research, Regularized Discriminant Analysis (RDA) analysis is used to classify the types of lung cancer that exist. Regularized Discriminant Analysis (RDA) is a generalization of Linear Discriminant Analysis (LDA) and Quadratic Discriminant Analysis (QDA). If the alpha parameter is set to 1, this operator performs an LDA. Likewise if the alpha parameter is set to 0, this operator performs QDA. Discriminant analysis is used to determine which variables distinguish between two or more groups that occur naturally. With an accuracy value of 60%, it can be ascertained the type of lung cancer suffered by patients for ease of care and treatment.

Keyword : Regularized Discriminant Analysis (RDA), Lung Cancer, Accuracy

Abstrak

Kanker paru merupakan jenis kanker yang onsetnya dimulai dari paru-paru dengan pertumbuhan sel yang tidak normal. Kanker paru merupakan jenis kanker dengan jumlah penderita terbesar didunia di bandingkan dengan kanker lainnya. Pada penelitian kali ini digunakan algoritma Regularized Discriminant Analysis (RDA) dalam mengelompokkan jenis kanker paru yang ada. Regularized Discriminant Analysis (RDA) adalah generalisasi dari the Linear Discriminant Analysis (LDA) dan the Quadratic Discreminant Analysis (QDA). Jika parameter alpha diatur ke 1, operator ini melakukan LDA. Demikian pula jika parameter alpha diatur ke 0, operator ini melakukan QDA. Analisis diskriminan digunakan untuk menentukan variabel mana yang membedakan antara dua atau lebih kelompok yang terjadi secara alami. Dengan nilai akurasi sebesar 60% dapat dipastikan jenis kanker paru yang diderita pasien untuk mempermudah dalam penanganan dan pengobatan.

Kata Kunci : Regularized Discriminant Analysis (RDA), kanker paru, akurasi

1. Pendahuluan

Jumlah penderita kanker paru saat ini sangat memprihatinkan ditengah pandemi berupa covid-19, penderita kanker menjadi sangat riskan dalam tertularnya virus. Di Indonesia kanker paru merupakan salah satu kanker dengan jumlah terbesar setelah kanker serviks dan kanker payudara sedangkan didunia berdasarkan (Cancer, 2018) data dunia menunjukkan tingkat penderita kanker paru terbesar

nomor satu dibandingkan jenis kanker yang lainnya yaitu sebanyak 2.093.876 kasus.

Beberapa faktor yang mempengaruhi dan menjadikan resiko tersendiri yang tidak dapat dikendalikan diantaranya : usia, jenis kelamin, dan genetik serta beberapa faktor yang dapat dikendalikan seperti merokok, radon atau gas radioaktif yang dilepaskan secara normal dari peluruhan uranium, thorium dan radium dalam batu dan tanah, selanjutnya polutan dari didalam, dan polutan diluar,

penyakit paru kronik dan karsinogen di lingkungan kerja (RI, 2018).

Data di Indonesia membuktikan bahwa kanker paru merupakan kanker terbesar pertama bagi pria dan terbesar keempat bagi wanita setelah kanker payudara, kanker serviks dan kanker ovarium (Fariha Ramadhaniah, Desy Khairina, Dian Triana Sinulingga & Mulawarman, 2019).

Sejumlah 13% keganasan tumor dunia berasal dari tumor paru dan menjadikan sepertiga faktor kematian utama yang ada di Indonesia, ditemukan pada pria perokok dan berusia diatas 50 tahun (Iqbalawaty et al., 2019).

Pada penderita kanker paru juga akan mengalami beberapa batuk, nafas pendek, nyeri dada, batuk darah, mual, nyeri, kelelahan dan beberapa keluhan lainnya (Ananda, Ermayanti, & Abdiana, 2018).

Pada penelitian kali ini dilakukan dengan mengambil sample penderita kanker paru-paru dari Stefan Aeberhard dengan jumlah contoh sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3. Dikarenakan data set berupa nominal maka dilakukankah sebuah pengujian dengan menggunakan Regularized Discriminant Analysis (RDA) dimana RDA merupakan salah satu metode yang dapat menganalisis diskriminan yang digunakan untuk menentukan variable yang mana membedakan dua atau lebih kelompok yang terjadi secara alami.

The Regularized Discriminant Analysis (RDA) adalah generalisasi dari the Linear Discriminant Analysis (LDA) dan the Quadratic Discriminant Analysis (QDA). Jika parameter alpha diatur ke 1, operator ini melakukan LDA. Demikian pula jika parameter alpha diatur ke 0, operator ini melakukan QDA. Analisis diskriminan digunakan untuk menentukan variabel mana yang membedakan antara dua atau lebih kelompok yang terjadi secara alami. Secara komputasi, analisis fungsi diskriminan sangat mirip dengan analysis of variance (ANOVA). Gagasan dasar yang mendasari analisis diskriminan adalah untuk menentukan apakah kelompok berbeda sehubungan dengan rata-rata variabel, dan kemudian menggunakan variabel itu untuk memprediksi keanggotaan grup (RapidMiner, 2020).

Analisis Diskriminan dapat digunakan untuk dua tujuan: dalam menilai kecukupan klasifikasi, mengingat keanggotaan kelompok dari objek yang diteliti; atau dalam menetapkan objek ke salah satu dari sejumlah grup objek. Analisis Diskriminan dengan demikian dapat memiliki tujuan deskriptif atau prediksi. Dalam kedua kasus tersebut, beberapa tugas kelompok harus diketahui sebelum melakukan Analisis Diskriminan. Tugas kelompok seperti itu, atau pelabelan, dapat diterima dengan cara apa pun. Oleh karena itu Analisis Diskriminan dapat digunakan sebagai pelengkap yang berguna untuk Analisis Cluster (untuk menilai hasil yang terakhir) atau *Principal Components Analysis* (RapidMiner, 2020).

2. Metode Penelitian

Penelitian kali ini menggunakan data sekunder dengan metode penelitian survey oleh Stefan Aeberhard dengan jumlah sample sebanyak 32 orang dan jumlah atribut sebanyak 57 buah, satu buah atribut class label dan 56 atribut berupa nominal integer dengan batasan antara 0-3.

Data set yang telah dikumpulkan akan diuji dengan menggunakan Regularized Discriminant Analysis dimana The Regularized Discriminant Analysis (RDA) adalah generalisasi dari the Linear Discriminant Analysis (LDA) dan the Quadratic Discriminant Analysis (QDA). Jika parameter alpha diatur ke 1, operator ini melakukan Linear Discriminant Analysis (LDA). Demikian pula jika parameter alpha diatur ke 0, operator ini melakukan Quadratic Discriminant Analysis (QDA). Analisis diskriminan digunakan untuk menentukan variabel mana yang membedakan antara dua atau lebih kelompok yang terjadi secara alami.

Perbedaan mendasar diantaranya

- a. Linear Discriminant Analysis yaitu Operator RDA melakukan analisis diskriminan yang diatur The Regularized Discriminant Analysis (RDA) yang merupakan generalisasi dari Linear Discriminant Analysis (LDA) yang merupakan kasus khusus dari algoritma ini. Jika parameter alpha diatur ke 1, operator RDA melakukan LDA.
- b. Quadratic Discriminant Analysis yaitu Operator RDA melakukan analisis diskriminan yang diatur (RDA) yang merupakan generalisasi dari QDA yang merupakan kasus khusus dari algoritma

ini. Jika parameter alpha diatur ke 0, operator RDA melakukan QDA.

Input yang dibutuhkan pada RDA berupa training set yaitu output dari operator Retrieve dalam Proses. Output dari operator lain juga dapat digunakan sebagai input.

Sedangkan *output* atau keluarannya adalah berupa model atau analisis diskriminan yang dilakukan.

Parameter yang digunakan pada Regularized Discriminant Analysis (RDA) sebagai berikut:

a. Alfa

Parameter ini menentukan kekuatan regularisasi. Jika diatur ke 1, hanya kovarians global yang digunakan. Jika diatur ke 0, hanya kovarians per kelas yang digunakan.

Range : Real

b. `approximate_covariance_inverse`

Parameter ini menunjukkan apakah kebalikan dari matriks kovarians harus diperkirakan jika invers yang sebenarnya tidak ada. Ini diaktifkan secara default.

Range : Boolean

3. Hasil dan Pembahasan

Hasil dari penelitian tentang kanker paru menggunakan Regularized Discriminant Analysis (RDA) memperoleh hasil akurasi sebesar 60% dengan 9 data pertama sebagai berikut:

Tabel 1 data set Kanker Paru

| class | atribut 1 | atribut 2 | atribut 3 |
|-----------|-----------|-----------|-----------|
| 1 | 0 | 3 | 0 |
| 1 | 0 | 3 | 3 |
| 1 | 0 | 3 | 3 |
| 1 | 0 | 2 | 3 |
| 1 | 0 | 3 | 2 |
| 1 | 0 | 3 | 3 |
| 1 | 0 | 3 | 2 |
| 1 | 0 | 2 | 2 |
| 1 | 0 | 3 | 1 |
| atribut 4 | atribut 5 | atribut 6 | atribut 7 |
| 0 | 0 | 2 | 2 |
| 1 | 0 | 3 | 1 |
| 2 | 0 | 3 | 3 |
| 2 | 1 | 3 | 3 |

| 1 | 1 | 3 | 3 |
|------------|------------|------------|------------|
| 2 | 0 | 3 | 3 |
| 1 | 0 | 3 | 3 |
| 1 | 0 | 3 | 1 |
| 1 | 0 | 3 | 1 |
| atribut 8 | atribut 9 | atribut 10 | atribut 11 |
| 2 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 |
| 3 | 1 | 1 | 1 |
| 3 | 1 | 2 | 1 |
| 3 | 2 | 2 | 2 |
| 3 | 1 | 2 | 2 |
| 3 | 1 | 2 | 1 |
| 3 | 3 | 3 | 3 |
| 3 | 1 | 1 | 1 |
| atribut 12 | atribut 13 | atribut 14 | atribut 15 |
| 1 | 3 | 2 | 2 |
| 1 | 1 | 3 | 3 |
| 0 | 3 | 3 | 3 |
| 0 | 3 | 3 | 1 |
| 1 | 1 | 2 | 2 |
| 0 | 3 | 3 | 3 |
| 2 | 3 | 3 | 3 |
| 2 | 1 | 3 | 3 |
| 3 | 2 | 3 | 3 |
| atribut 16 | atribut 17 | atribut 18 | atribut 19 |
| 1 | 2 | 2 | 0 |
| 1 | 2 | 2 | 0 |
| 1 | 2 | 1 | 0 |
| 1 | 2 | 2 | 0 |
| 2 | 2 | 2 | 0 |
| 2 | 2 | 1 | 0 |
| 3 | 2 | 2 | 0 |
| 1 | 2 | 2 | 0 |
| 1 | 2 | 2 | 0 |
| atribut 20 | atribut 21 | atribut 22 | atribut 23 |
| 2 | 2 | 2 | 2 |
| 0 | 2 | 2 | 2 |
| 0 | 2 | 2 | 2 |
| 0 | 2 | 2 | 2 |

| | | | | | | | |
|---------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| 0 | 2 | 2 | 2 | 2 | 2 | 1 | 2 |
| 0 | 1 | 2 | 2 | 2 | 2 | 1 | 2 |
| 0 | 2 | 2 | 2 | 2 | 1 | 2 | 1 |
| 0 | 1 | 1 | 2 | 2 | 1 | 2 | 2 |
| 0 | 2 | 2 | 2 | 2 | 2 | 1 | 2 |
| atribut 24 | atribut 25 | atribut 26 | atribut 27 | atribut 40 | atribut 41 | atribut 42 | atribut 43 |
| 1 | 2 | 2 | 2 | 2 | 1 | 2 | 2 |
| 1 | 2 | 1 | 3 | 2 | 2 | 1 | 2 |
| 1 | 2 | 2 | 3 | 2 | 1 | 2 | 2 |
| 2 | 1 | 3 | 2 | 2 | 2 | 1 | 2 |
| 1 | 1 | 2 | 3 | 2 | 2 | 1 | 2 |
| 2 | 1 | 3 | 3 | 2 | 2 | 1 | 2 |
| 2 | 1 | 3 | 2 | 2 | 1 | 2 | 2 |
| 1 | 2 | 1 | 3 | 2 | 1 | 2 | 2 |
| 1 | 2 | 1 | 2 | 2 | 1 | 2 | 2 |
| atribut 28 | atribut 29 | atribut 30 | atribut 31 | atribut 44 | atribut 45 | atribut 46 | atribut 47 |
| 3 | 2 | 1 | 1 | 2 | 2 | 2 | 2 |
| 2 | 3 | 1 | 1 | 2 | 1 | 2 | 2 |
| 2 | 3 | 1 | 3 | 1 | 2 | 2 | 2 |
| 3 | 3 | 1 | 3 | 2 | 2 | 1 | 2 |
| 2 | 2 | 1 | 1 | 2 | 2 | 2 | 2 |
| 1 | 2 | 2 | 3 | 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 3 | 2 | 2 | 1 | 2 |
| 2 | 1 | 1 | 3 | 2 | 1 | 2 | 2 |
| 1 | 1 | 1 | 3 | 1 | 2 | 2 | 2 |
| atribut 32 | atribut 33 | atribut 34 | atribut 35 | atribut 48 | atribut 49 | atribut 50 | atribut 51 |
| 1 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| 1 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 1 | 2 | 2 | 2 | 2 | 2 |
| 3 | 3 | 1 | 1 | 2 | 2 | 2 | 2 |
| 1 | 3 | 2 | 1 | 2 | 2 | 2 | 2 |
| 3 | 3 | 2 | 1 | 2 | 2 | 2 | 2 |
| 3 | 3 | 2 | 1 | 2 | 2 | 2 | 2 |
| 3 | 3 | 2 | 1 | 2 | 2 | 2 | 1 |
| 3 | 3 | 2 | 2 | 2 | 1 | 2 | 2 |
| 3 | 3 | 3 | 2 | 2 | 2 | 2 | 2 |
| atribut 36 | atribut 37 | atribut 38 | atribut 39 | atribut 52 | atribut 53 | atribut 54 | atribut 55 |
| 2 | 1 | 2 | 2 | 1 | 1 | 1 | 2 |
| 2 | 2 | 1 | 2 | 2 | 2 | 1 | 2 |
| 2 | 1 | 2 | 2 | 2 | 2 | 2 | 1 |
| 1 | 1 | 2 | 2 | 2 | 2 | 2 | 2 |

| | | | |
|---|---|---|---|
| 2 | 2 | 1 | 2 |
| 2 | 2 | 2 | 1 |
| 2 | 2 | 2 | 1 |
| 2 | 2 | 1 | 2 |
| 2 | 2 | 1 | 2 |

| | | | | |
|---------|---|---|---|---|
| atribut | | | | |
| 56 | 2 | 2 | 2 | 2 |
| 2 | 2 | 2 | 2 | 2 |

Berikut ini hasil akurasi menggunakan Regularized Discriminant Analysis (RDA).

| Multiclass Classification Performance Annotations | | | | |
|---|---------|--------|--------|-----------------|
| Table View Plot View | | | | |
| accuracy: 60.00% +/- 11.06% (mikro: 59.38%) | | | | |
| | true 1 | true 2 | true 3 | class precision |
| pred. 1 | 19 | 9 | 4 | 59.38% |
| pred. 2 | 0 | 0 | 0 | 0.00% |
| pred. 3 | 0 | 0 | 0 | 0.00% |
| class recall | 100.00% | 0.00% | 0.00% | |

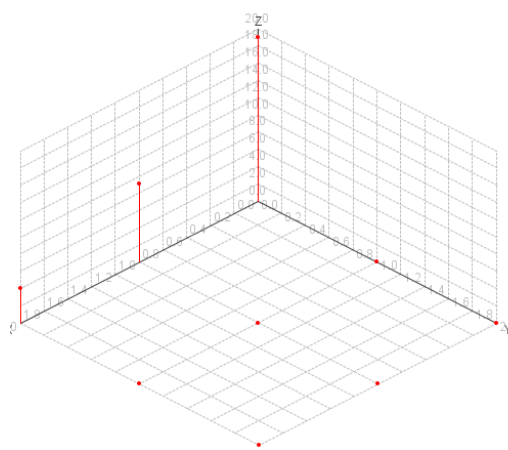
Gambar 1 Hasil Akurasi RDA pada Kanker Paru

| PerformanceVector | | | |
|---|----|---|---|
| PerformanceVector: | | | |
| accuracy: 60.00% +/- 11.06% (mikro: 59.38%) | | | |
| ConfusionMatrix: | | | |
| True: | 1 | 2 | 3 |
| 1: | 19 | 9 | 4 |
| 2: | 0 | 0 | 0 |
| 3: | 0 | 0 | 0 |
| kappa: 0.000 +/- 0.000 (mikro: 0.000) | | | |
| ConfusionMatrix: | | | |
| True: | 1 | 2 | 3 |
| 1: | 19 | 9 | 4 |
| 2: | 0 | 0 | 0 |
| 3: | 0 | 0 | 0 |

Gambar 2. PerformanceVector RDA pada Kanker Paru

| kappa: 0.000 +/- 0.000 (mikro: 0.000) | | | | |
|---------------------------------------|---------|--------|--------|-----------------|
| | true 1 | true 2 | true 3 | class precision |
| pred. 1 | 19 | 9 | 4 | 59.38% |
| pred. 2 | 0 | 0 | 0 | 0.00% |
| pred. 3 | 0 | 0 | 0 | 0.00% |
| class recall | 100.00% | 0.00% | 0.00% | |

Gambar 3. Kappa RDA pada Kanker Paru



Gambar 4. Plot View RDA pada Kanker Paru

| Regularized Discriminant Model | |
|--------------------------------|--------|
| Apriori probabilities: | |
| 1 | 0.5938 |
| 2 | 0.2812 |
| 3 | 0.1250 |

Gambar 5. Regularized Discriminat Model pada Kanker Paru

Penelitian kali ini membuktikan bahwa class 1 mempunyai peluang yang lebih besar terkena kanker paru dan class 3 mempunyai peluang paling kecil terkena kanker paru.

4. Kesimpulan

Penelitian ini dilakukan dengan menggunakan data set berupa kanker paru dengan jumlah contoh sebanyak 32 pasien dan jumlah atribut sebanyak 57 buah dengan satu buah berupa class label dan 56 buah berupa atribut. Penelitian kali ini

menggunakan *Regularized Discriminant Analysis (RDA)* dengan nilai akurasi sebesar 60%.

Berikut saran yang didapat dari hasil penelitian ini:

1. perlu dilakukan pengujian dengan algoritma yang lain untuk meningkatkan akurasi.
2. Perlu dilakukan pengembangan dengan variasi model yang lain.

Referensi

- Ananda, R. R., Ermayanti, S., & Abdiana. (2018). Hubungan Staging Kanker Paru dengan Skala Nyeri pada Pasien Kanker Paru yang Dirawat di Bagian Paru RSUP DR M Djamil Padang. *Jurnal Kesehatan Andalas*, 7(3). Retrieved from file:///D:/jurnal/lung cancer/kanker paru 2.pdf
- Cancer, G. B. O. (2018). Estimated number of new cases in 2018. Retrieved from https://gco.iarc.fr/today/online-analysis-table?v=2018&mode=cancer&mode_population=continents&population=900&populations=900&key=asr&sex=0&cancer=39&type=0&statistic=5&prevalence=0&population_group=0&ages_group%5B%5D=0&ages_group%5B%5D=17&group_cancer=1&include_nmsc=1&include_nmsc_other=1
- Fariha Ramadhaniah, Desy Khairina, Dian Triana Sinulingga, E. S., & Mulawarman, A. (2019). Gambaran Pasien Kanker Paru di Rumah Sakit Kanker Dharmais (RSKD) Tahun 2008-2012. *Jurnal Respirologi Indonesia*, 39, 31. Retrieved from <https://jurnalrespirologi.org/index.php/jri/article/viewFile/1/23>
- Iqbalawaty, I., Machillah, N., Farjriah, Abdullah, A., Yani, M., Ilzana, T. M., ... Khaled, T. M. (2019). Profil hasil pemeriksaan CT-Scan pada pasien tumor paru di Bagian Radiologi RSUD Dr. Zainoel Abidin periode Juli 2018-Oktober 2018. *Intisari Sains Medis*, 10, 625–630. Retrieved from <https://isainsmedis.id/index.php/ism/article/viewFile/661/411>
- RapidMiner. (2020). Regularized Discriminant Analysis. Retrieved from [https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/discriminant_analysis/regularized_discriminant_analysis.html#:~:text=The regularized discriminant analysis \(RDA,quadratic discredinant analysis \(QDA\).&text=Discriminant analysis is used to determine which variables discriminate,or more naturally occurring groups.](https://docs.rapidminer.com/latest/studio/operators/modeling/predictive/discriminant_analysis/regularized_discriminant_analysis.html#:~:text=The regularized discriminant analysis (RDA,quadratic discredinant analysis (QDA).&text=Discriminant analysis is used to determine which variables discriminate,or more naturally occurring groups.)
- RI, K. K. (2018). Pedoman Pengendalian Faktor Risiko Kanker Paru. Jakarta. Retrieved from <http://yayasankankerindonesia.org/storage/article/59a98dea4515c2f3845fa8d9be4b5c7c.pdf>