



KLASIFIKASI KANKER PAYUDARA MENGGUNAKAN ALGORITMA K-NN DENGAN OPTIMASI BOBOT (FORWARD)

Ahvan Muharam¹, Enda Suhadi², Tiki Ramdhani³, Imam Azmi⁴, Fitriyani⁵

ARS UNIVERSITY¹²³⁴⁵

e-mail: muharamahvan17@gmail.com, endasuhadi@gmail.com, tikiramadhan123@gmail.com,
imamazmi33@gmail.com, fitriyani@ars.ac.id

Abstrak

Tujuan dari studi eksplorasi ini adalah untuk mengembangkan dan menilai model prediksi yang berpotensi dapat digunakan sebagai biomarker kanker payudara, berdasarkan data antropometrik dan parameter yang dapat dikumpulkan di analisis darah rutin, Untuk masing-masing dari 166 peserta beberapa fitur klinis diamati atau diukur, termasuk usia, BMI, Glukosa, Insulin, HOMA, Leptin, Adiponektin, Resistin dan MCP-1. Algoritma pembelajaran mesin (regresi logistik, hutan acak, mesin vektor pendukung) diimplementasikan dengan mengambil sebagai prediktor jumlah variabel yang berbeda. Itu model yang dihasilkan dinilai dengan pendekatan Monte Carlo Cross-Validation untuk menentukan interval kepercayaan 95% untuk sensitivitas, spesifisitas dan AUC model. Mendukung model mesin vektor menggunakan Glukosa, Resistin, Usia, dan BMI sebagai prediktor yang memungkinkan prediksi kehadiran kanker payudara pada wanita dengan sensitivitas berkisar antara 82 dan 88% dan spesifisitas berkisar antara 85 dan 90%. Interval kepercayaan 95% untuk AUC adalah [0.87, 0.91] Temuan ini memberikan bukti yang menjanjikan bahwa model yang menggabungkan parameter usia, BMI dan metabolisme mungkin menjadi alat yang ampuh untuk biomarker kanker payudara yang murah dan efektif

Keywords: Kanker payudara, Glukosa, Resistin, BMI, Usia, Biomarker

Abstract

The aim of this exploratory study was to develop and assess predictive models that could potentially be used as biomarkers of breast cancer, based on anthropometric data and parameters that could be collected in routine blood analyses. For each of the 166 participants several clinical features were observed or measured, including age, BMI, Glucose, Insulin, HOMA, Leptin, Adiponectin, Resistin and MCP-1. Machine learning algorithms (logistic regression, random forest, supporting vector machines) are implemented by taking as predictors the number of different variables. The resulting model was assessed with a Monte Carlo Cross-Validation approach to determine 95% confidence intervals for the sensitivity, specificity and AUC of the model. Supporting a vector machine model using Glucose, Resistin, Age, and BMI as predictors allows prediction of the presence of breast cancer in women with sensitivity ranging between 82 and 88% and specificity ranging between 85 and 90%. The 95% confidence interval for the AUC is [0.87, 0.91] These findings provide promising evidence that a model combining age, BMI and metabolism parameters may be a powerful tool for inexpensive and effective breast cancer biomarkers.

Keywords: Breast cancer, Glucose, Resistin, BMI, Age, Biomarkers



1. Pendahuluan

Skrining kanker payudara adalah strategi penting untuk memungkinkan deteksi dini dan memastikan kemungkinan yang lebih besar untuk memiliki hasil yang baik dalam pengobatan. Prediksi yang kuat model berdasarkan data yang mungkin dikumpulkan secara rutin konsultasi dan analisis darah diupayakan untuk memberikan kontribusi penting dengan menawarkan lebih banyak alat penyaringan. Dalam pekerjaan ini kami bertujuan untuk menilai bagaimana model berdasarkan data yang dapat dikumpulkan dalam analisis darah rutin - tidak mampu, Glukosa, Insulin, HOMA, Leptin, Adiponektin, Resistin, MCP-1, Usia dan Indeks Massa Tubuh (BMI) – mungkin digunakan untuk memprediksi adanya kanker payudara. Kita percaya bahwa parameter ini adalah kumpulan tanggal yang bagus, karena kami baru-baru ini memverifikasi deregulasi di profil mereka tentang kanker payudara terkait obesitas, Beberapa kandidat untuk biomarker kanker payudara telah dilaporkan dalam literatur, [2]. Sensitivitas 93% dan spesifisitas 72% ketika membedakan kontrol dari pasien dengan kanker payudara, [3]. BMI, Leptin, CA15-3 dan rasio antara Leptin dan Adiponektin yang digunakan bersama dinilai sebagai biomarker untuk kanker payudara pada [4] (2013). Meskipun nilai yang sangat tinggi disajikan untuk accuracy: 85.30% +/- 10.35% (micro average: 85.34%)

2. Metode Penelitian

Metode penelitian yang kami gunakan yaitu komparasi antara dua metode klasifikasi data mining antara algoritma K-Nearest Neighbor (KNN) dengan Optimasi Weights (Forward) salah satu optimasi yang sering digunakan pada data mining

Data ini diambil pada halaman web yaitu Machine Learning Repository (UCI) yaitu dataset Kanker Payudara dimana data tersebut telah dirujuk beberapakali dan diterbitkan pada tahun 2018. Aplikasi yang digunakan oleh kami adalah aplikasi Rapidminer 9.9.002 dan menjelaskan kronologis penelitian, termasuk desain penelitian, prosedur penelitian , bagaimana untuk menguji dan akuisisi data. Deskripsi dari program penelitian harus didukung referensi, sehingga penjelasan tersebut dapat diterima secara ilmiah.

2.1 Data Mining

Data mining adalah suatu proses pengerukan atau pengumpulan informasi penting dari suatu data yang besar. Proses data mining seringkali menggunakan metode statistika, matematika, hingga memanfaatkan teknologi artificial intelligence.

Nama alternatifnya yaitu Knowledge discovery (mining) in databases (KDD), knowledge extraction, data/pattern analysis, data archeology, data dredging, information harvesting, business intelligence, dan lain-lain.

Jika dilihat dilihat pada gambar dalam proses KDD tersebut, Banyak konsep dan teknik yang digunakan dalam proses data mining. Proses tersebut membutuhkan beberapa langkah untuk mendapatkan sebuah data yang diinginkan.

Dalam proses KDD tersebut termasuk melakukan pembersihan data, integrasi data, seleksi data, transformasi, penambangan data, evaluasi pola, dan presentasi pengetahuan.

2.2 Klasifikasi

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan untuk membedakan konsep dan kelas data yang bertujuan untuk mendapatkan akurasi kelas dari suatu objek yang labelnya tidak diketahui.

2.3 K-Nearest Neighbor (KNN)

KNN adalah algoritma yang berfungsi untuk melakukan klasifikasi dengan proyeksi data pembelajaran pada ruang berdimensi banyak. Ruang ini dibagi menjadi bagian-bagian yang merepresentasikan kriteria data pembelajaran. Setiap data pembelajaran direpresentasikan menjadi titik-titik c pada ruang dimensi banyak.

2.4 Optimize Weights (Forward)

Adalah untuk mengoptimasi bobot dari suatu dataset yang telah kami teliti

2.5 Rapidminer

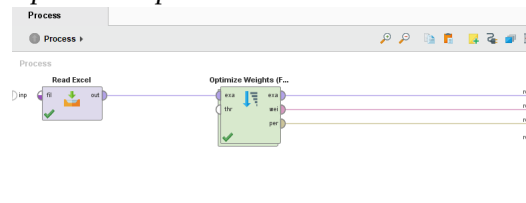
Adalah sebuah aplikasi perangkat lunak untuk mencari akurasi dari data yang ingin kita teliti dengan menggunakan beberapa algoritma yang telah tersedia didalamnya dan juga berfungsi untuk persiapan data, pembelajaran mesin, pembelajaran mendalam, penambangan teks dan analitik prediktif

3. Hasil dan Pembahasan

Pada Tahapan yang akan dibahas ini menggunakan klasifikasi K-Nn dengan oprimize weights (forward) dan tanpa menggunakan optimize weights (forward) atau menggunakan metode K-NN saja.

Pada bagian ini, dijelaskan hasil

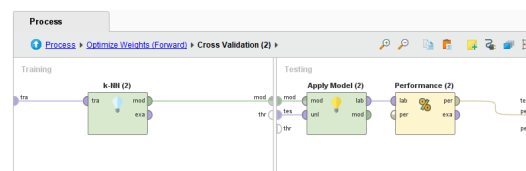
Gambar 3.1 Model proses RapidMiner



3.1. Hasil K-NN Dengan Optimize Weights (Forward)

Pada tahapan ini penyusun menggunakan algoritma K-NN dengan Optimize Weights (forward) . berikut hasilnya :

gambar 3.2 Model K-NN



Dari Model gambar di atas data akan di olah menggunakan algoritma K-NN dengan menggunakan Optimasi Bobot (forward) kemudian hasil dari data tersebut akan muncul hasil sebagai berikut.

Gambar 3.3 Hasil Tabel K-NN dan Forward

Table View
 Plot View

accuracy: 85.30% +/- 10.35% (micro average: 85.34%)

	true 1	true 2	class precision
pred. 1	39	4	90.70%
pred. 2	13	60	82.19%
class recall	75.00%	93.75%	

Gambar 3.4 Deskripsi hasil

```

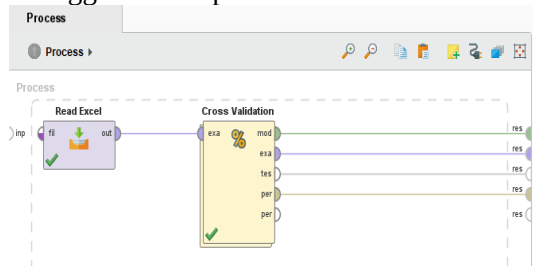
PerformanceVector

PerformanceVector:
accuracy: 85.30% +/- 10.35% (micro average: 85.34%)
ConfusionMatrix:
True: 1 2
1: 39 4
2: 13 60
  
```

Dari hasil data di atas yang berhasil di olah oleh algoritma K-NN dengan menggunakan optimasi bobot (forward) maka akurasi yang muncul adalah 85,30%, berikut gambar deskripsinya

3.2. K-NN tanpa Optimize Weights (Forward)

Pada tahapan kedua ini penyusun menggunakan algoritma K-NN saja tanpa menggunakan optimasi tambahan. Berikut model pada aplikasi RapidMiner :



Gambar 3.5 Model RapidMiner

Dari model gambar di atas data akan di teruskan menggunakan algoritma K-NN saja. Berikut adalah hasilnya

	true 1	true 2	class precision
pred. 1	17	22	43.59%
pred. 2	35	42	54.55%
class recall	32.69%	65.62%	

Gambar 3.6 hasil Tabel K-NN

Dari hasil data di atas yang berhasil di olah oleh algoritma K-NN saja maka akurasi yang muncul adalah 50,86 %, berikut gambar deskripsinya

Gambar 3.7 Deskripsi K-NN

```
KNNClassification
Weighted 5-Nearest Neighbour model for classification.
The model contains 116 examples with 9 dimensions of the following classes:
1
2
```

4. Kesimpulan

Berdasarkan uraian yang telah dibahas pada bab-bab sebelumnya mengenai Klasifikasi dataset Kardiotokografi menggunakan algoritma Neural network, baik secara teoritis maupun penerapannya. Maka penulis dapat mengambil kesimpulan sebagai berikut :

1. Model klasifikasi algoritma K-NN dapat melakukan klasifikasi dataset Kanker Payudara dengan baik, dibuktikan oleh nilai akurasi yang dihasilkan sebesar 85.30% hal itu memperlihatkan bahwa klasifikasi Kanker Payudara menggunakan algoritma

K-NN dapat diimplementasikan menggunakan suatu aplikasi dalam membantu pada saat melakukan klasifikasi.

2. Dengan adanya klasifikasi ini diharapkan dapat membantu ahli Kanker Payudara dalam melakukan proses klasifikasi payudara dengan menggunakan dataset. Dengan adanya aplikasi proses klasifikasi dapat dilihat dengan cepat.

Referensi

- V.G. Vogel, Epidemiology, genetics, and risk evaluation of postmenopausal women at risk of breast cancer. Menopause 15, 782–789 (2008)
- A.G. Renehan, M. Tyson, M. Egger, R.F. Heller, M. Zwahlen, Body-mass index and incidence of cancer: a systematic review and meta-analysis of prospective observational studies. Lancet 371, 569–578 (2008)



- L. Vona-Davis, D.P. Rose, Type 2 diabetes and obesity metabolic interactions: common factors for breast cancer risk and novel approaches to prevention and therapy. *Curr. Diab. Rev.* 8, 116–130 (2012)
- D.P. Rose, L. Vona-Davis, The cellular and molecular mechanisms by which insulin influences breast cancer risk and progression. *Endocr. Relat. Cancer* 19, R225–R241 (2012)
- M. Dalamaga, Obesity, insulin resistance, adipocytokines and breast cancer: New biomarkers and attractive therapeutic targets. *World J. Exp. Med.* 20, 34–42 (2013)
- D.H. Cohen, D. LeRoith, Obesity, type 2 diabetes, and cancer: the insulin and IGF connection. *Endocr. Relat. Cancer* 19, F27–F45 (2012)
- P. Matafome, D. Santos-Silva, C.M. Sena, R. Seica, Common mechanisms of dysfunctional adipose tissue and obesity-related cancers. *Diabetes Metab. Res. Rev.* 29, 285–295 (2013)
- A.M. Lorincz, S. Sukumar, Molecular links between obesity and breast cancer. *Endocr. Relat. Cancer* 13, 279–292 (2006)
- G.L. Anderson, M.L. Neuhouse, Obesity and the risk for premenopausal and postmenopausal breast cancer. *Cancer Prev. Res.* 5, 515–521 (2012)