

Deteksi Penyakit Kanker Payudara dengan Seleksi Fitur berbasis Principal Component Analysis dan Random Forest

Ahmad Fauzi¹, Riki Supriyadi², Nurlaelatul Maulidah³

^{1,2,3}Universitas Bina Sarana Informatika

e-mail: ahmad.fzx@bsi.ac.id, riki.rsd@bsi.ac.id, nurlaelatul.nlt@bsi.ac.id

Diterima	Direvisi	Disetujui
30-04-2020	06-05-2020	20-05-2020

Abstrak - Kanker payudara (*Carcinoma mammae*) didefinisikan sebagai suatu penyakit *neoplasma* ganas yang berasal dari *parenchyma*. *World Health Organization (WHO)* memasukan penyakit ini ke dalam *International Classification of Diseases (ICD)* dengan kode nomor 17. *Skrining* merupakan salah satu upaya untuk mengidentifikasi penyakit atau kelainan yang belum jelas dengan menggunakan tes, pemeriksaan atau prosedur tertentu. Upaya ini dapat digunakan secara cepat untuk membedakan orang - orang yang kelihatannya sehat tetapi sesungguhnya menderita suatu kelainan. Tujuan utama penelitian ini adalah untuk meningkatkan performa klasifikasi pada diagnosis kanker payudara dengan menerapkan seleksi fitur pada beberapa algoritme klasifikasi. Database yang digunakan merupakan database kanker payudara *Breast Cancer Coimbra Data Set*. Metode seleksi fitur berbasis *pricipal component analysis* akan dipasangkan dengan beberapa algoritme klasifikasi dan metode, seperti *Logitboost*, *Bagging*, dan *Random Forest*. Penelitian ini menggunakan *10 fold cross validation* sebagai metode evaluasi. Hasil penelitian menunjukkan metode seleksi fitur berbasis *pricipal component analysis* mengalami peningkatan performa klasifikasi secara signifikan setelah dipasangkan dengan seleksi fitur *Random Forest* dan *logitboost*, *Random forest* menunjukkan performa terbaik dengan akurasi 79.3103% dengan nilai AUC sebesar 0,843.

Kata Kunci: Seleksi Fitur, PCA, Kanker Payudara, *Skrining*, *Random Forest*

Abstract - *Breast cancer (Carcinoma mammae)* is defined as a malignant neoplastic disease derived from *parenchyma*. The *World Health Organization (WHO)* enters the disease into *International Classification of Diseases (ICD)* with code number 17. *Screening* is one of the attempts to identify unclear diseases or abnormalities by using certain tests, examinations or procedures. This effort can be used quickly to distinguish people who seem to be healthy but actually suffer from an abnormality. The main purpose of this research is to improve classification performance on the diagnosis of breast cancer by applying a selection of features to several classification algorithms. The database is used as a database of breast Cancer, the *Coimbra Data Set breast cancer*. The method of selection of *Pricipal-based component analysis feature* will be paired with several classification algorithms and methods, such as *Logitboost*, *Bagging*, and *Random Forest*. The study used *10 fold cross validation* as the evaluation method. The results showed a method of selection of *Pricipal-based feature component analysis* experienced a significantly improved classification performance after pairing with the selection of *Random Forest* and *logitboost* features, *Random Forest* showed the best performance with 79.3103% accuracy with AUC value of 0.843..

Keywords: selection of features, PCA, Breast Cancer, Screening, Random Forest

PENDAHULUAN

Kanker payudara (*Carcinoma mammae*) merupakan suatu penyakit *neoplasma* ganas yang berasal dari *parenchyma*. *World Health Organization (WHO)* memasukan penyakit ini ke dalam *International Classification of Diseases (ICD)* dengan kode nomor 17 (Assiri, A. M. A., & Kamel, H. F. M. (2015)). Frekuensi kasus penyakit ini relatif tinggi di negara maju dan merupakan jenis kanker yang banyak diderita dari jenis kanker lainnya. Di Indonesia, kanker payudara menempati peringkat kedua setelah kanker servik. Menurut data terakhir

WHO, angka kematian karena kanker payudara di Indonesia mencapai 20.052 atau sebanyak 1,41% dari seluruh kematian atau angka kematian disesuaikan dengan usia adalah 2.025 per 100.000 penduduk (Cole, K. D., He, H., & Wang, L. (2013)). Kunci untuk bertahan hidup penderita kanker payudara adalah mendeteksi kanker payudara sedini mungkin, sebelum kanker tersebut memiliki kesempatan untuk menyebar (Assiri, A. M. A., & Kamel, H. F. M. (2015)).

Semakin berkembangnya teknologi dan informasi terutama bidang kecerdasan buatan, teknik *machine learning* diperkenalkan untuk membantu

meningkatkan kemampuan pendeteksian otomatis. Dengan bantuan sistem ini, kemungkinan kesalahan diagnosis yang dilakukan oleh para ahli dapat dihindari, dan data medis dapat diperiksa dalam kurun waktu yang singkat serta lebih rinci (Ghaderyan, P., Abbasi, A., & Hossein, M. (2014)). Teknik statistik dan teknik kecerdasan buatan telah digunakan untuk memprediksi kanker payudara oleh beberapa peneliti. Tujuan dari teknik ini adalah untuk menetapkan identifikasi pasien ke dalam grup jinak (yang tidak memiliki kanker payudara) atau kelompok ganas (yang terbukti kuat memiliki kanker payudara) (Hasan, H., Tahir, N., & Mara, U. T. (2010)).

Data medis yang berdimensi tinggi merupakan salah satu kendala dalam penerapan teknik *machine learning* karena akan memberikan efek negatif terhadap proses analisis. Untuk menanganinya data medis berdimensi tinggi tersebut, mereduksi fitur menjadi hal yang sangat penting. Dengan pengurangan fitur tidak mengakibatkan kemampuan diskriminatif menjadi memburuk, bahkan sebaliknya

terdapat banyak keuntungan diantaranya dapat menghindari *over-fitting*, mengurangi kompleksitas analisis data dan meningkatkan kinerja analisis data. Salah satu usaha untuk mengurangi fitur data yang berdimensi tinggi adalah dengan menggunakan seleksi fitur, seleksi fitur merupakan bagian dari *preprocessing* pada proses klasifikasi. Pemilihan fitur sangat mempengaruhi keakuratan klasifikasi dalam kasus kanker payudara.

Pada penelitian ini dua metode fitur seleksi akan diterapkan dan kemudian akan diuji cobakan pada beberapa algoritme klasifikasi yang berbeda, *Principal component analysis* merupakan seleksi fitur yang dapat mengidentifikasi fitur-fitur yang signifikan dan menghilangkan fitur-fitur yang tidak relevan untuk menghasilkan model pembelajaran yang baik, sehingga dapat mengurangi dimensi data tanpa kekurangan informasi yang terkandung dalam kumpulan data tersebut. *F-score* sendiri merupakan seleksi fitur dengan teknik sederhana yang mengukur diskriminasi dua set bilangan real, sehingga fitur yang memiliki nilai *Fscore* rendah dianggap memiliki kemampuan diskriminatif yang rendah pula begitu pula sebaliknya fitur yang memiliki nilai *F-score* tinggi juga akan memiliki kemampuan diskriminatif yang tinggi pula. Dalam penelitian terdahulu diketahui seleksi fitur *Principal component analysis* dan seleksi fitur *F-score* memiliki kemampuan yang sangat baik dalam memilih fitur-fitur yang signifikan terhadap klasifikasi.

1. Kanker Payudara

Kanker adalah penyakit yang berbahaya dan salah satu penyebab kematian tertinggi di Indonesia bahkan dunia (Arifin, Riana, & Hapsari,

2014). Kanker payudara merupakan jenis kanker yang terbanyak di Indonesia. Berdasarkan *Pathological Based Registration* di Indonesia, Data Kanker di Indonesia Tahun 2010 menunjukkan kanker payudara menempati urutan pertama dengan frekuensi relatif sebesar 18,6% berdasarkan data Histopatologik ; Badan Registrasi Kanker Perhimpunan Dokter Spesialis Patologi Indonesia dan Yayasan Kanker Indonesia . Diperkirakan kejadian yang ada di Indonesia mencapai 12/100.000 wanita, sedangkan di Amerika adalah sekitar 92/100.000 wanita dengan mortalitas yang lumayan tinggi yaitu 27/100.000 atau 18 % dari kematian pada wanita. Penyakit ini juga dapat juga diderita laki - laki dengan frekuensi sekitar 1 %. Di Indonesia mencapai lebih dari 80% kasus yang ditemukan pada stadium yang lanjut, dimana upaya pengobatan sulit dilakukan. (KEMENTERIAN KESEHATAN REPUBLIK INDONESIA PANDUAN PENATALAKSANAAN KANKER PAYUDARA, n.d.).

2. Skrining

Pencegahan primer merupakan usaha supaya tidak terkena kanker payudara . Pencegahan berupa mengurangi atau meniadakan faktor-faktor risiko yang diduga sangat erat kaitannya dengan peningkatan insiden kanker payudara. Pencegahan supaya tidak terjadinya kanker secara sederhana adalah mengetahui faktor -faktor risiko kanker payudara, seperti yang telah disebutkan di atas, dan berusaha menghindarinya. Prevensi primer agar tidak terjadi kanker payudara saat ini memang masih sulit, yang bisa dapat dilakukan adalah dengan meniadakan atau memperhatikan beberapa faktor risiko yang erat kaitannya dengan peningkatan insiden kanker payudara (KEMENTERIAN KESEHATAN REPUBLIK INDONESIA PANDUAN PENATALAKSANAAN KANKER PAYUDARA, n.d.).

Pencegahan sekunder yaitu melakukan skrining kanker payudara. Skrining kanker payudara merupakan strategi penting untuk memungkinkan deteksi dini dan memastikan probabilitas yang lebih besar untuk mendapatkan hasil yang baik dalam pengobatan. Skrining kanker payudara dengan pemeriksaan atau usaha untuk menemukan abnormalitas yang menuju pada kanker payudara pada seseorang atau sekelompok orang yang tidak mempunyai keluhan. Model prediktif yang kuat berdasarkan data yang dapat dikumpulkan dalam konsultasi rutin dan analisis darah dicari untuk memberikan kontribusi penting dengan menawarkan lebih banyak alat skrining.

3. Principal Component Analysis

PCA merupakan kombinasi linear dari variabel awal yang secara geometris kombinasi linear ini merupakan sistem koordinat baru yang diperoleh dari rotasi sistem semula. Metode *Principal Component Analysis* sangat berguna jika data yang ada memiliki jumlah variabel dalam jumlah besar dan memiliki korelasi antar variabelnya. Perhitungan dari atau *Principal Component Analysis* didasari pada perhitungan nilai eigen dan vektor eigen yang menyatakan penyebaran data dari suatu dataset.

4. Random Forest

Random Forest merupakan pengembangan dari *Decision Tree* dengan menggunakan beberapa *Decision Tree*, dimana setiap *Decision Tree* telah dilakukan *training* menggunakan sampel individu dan setiap atribut dipecah pada *tree* yang dipilih antara atribut subset yang bersifat acak. Pada proses klasifikasi, individunya berdasarkan pada *vote* dari suara terbanyak pada kumpulan populasi *tree*. *Random Forest* yang didapat memiliki banyak *tree*, dan setiap *tree* ditanam dengan cara yang sama. *Tree* pada variabel x akan ditanam sejauh mungkin dengan *tree* pada variabel y . Dan dalam perkembangannya, sejalan dengan bertambahnya data set, maka *tree* pun ikut berkembang. Penempatan *tree* yang berjauhan membuat jika terdapat *tree* disekitar *tree* x berarti *tree* tersebut merupakan perkembangan dari *tree* x . Beberapa fungsi *learning* yang didapatkan oleh *random forest* digunakan strategi *ensemble* "bagging" untuk mengatasi masalah *overfitting* apabila dihadapkan data set yang kecil. Pada makalah ini *Ensemble* digunakan untuk melakukan *resampled* data dengan mengklasifikasi ulang data *outlayer* sehingga merubah struktur data set yang asli. (Saputra, Sujatmika, & Arifin, 2011).

5. Bagging

Bagging (bootstrap aggregating) adalah salah satu teknik dari metode ensemble dengan memanipulasi data pelatihan, data pelatihan digandakan dalam d kali dengan pengambilan sampel dengan penggantian. Proses tersebut kemudian menghasilkan sebanyak data pelatihan baru, kemudian dari data pelatihan akan dibangun pengklasifikasi yang disebut pengklasifikasi kantong (Muslim, Nurzahputra, & Prasetyo, 2018). Teknik bagging banyak digunakan dengan hasil meningkatkan tingkat akurasi dari penggunaan algoritma dalam data mining.

Bagging menggunakan subdataset (bootstrap) untuk menghasilkan set pelatihan L (learning) yang kemudian melatih dasar belajar menggunakan prosedur pembelajaran yang tidak stabil dan mengambil rata-rata selama pengujian (Arrahimi, Ihsan, Kartini, Faisal, & Indriani, 2019).

Bagging (bootstrap aggregating) salah satu algoritma *ensemble learning* yang paling awal yang dikembangkan oleh Breiman. *Bagging* didasarkan pada konsep *voting* mayoritas, di mana subset data pelatihan yang berbeda secara acak digunakan untuk pelatihan sebuah *base learner* yang berbeda dengan cara yang sama (Wang et al., 2011).

Algoritma merupakan salah satu algoritma yang paling intuitif dan paling sederhana untuk diterapkan. Keanekaragaman pada bagging diperoleh dari replika *bootstrapped* dari data pelatihan. subset data pelatihan yang beda secara acak ditarik dengan penggantian dari keseluruhan data pelatihan. setiap subset data pelatihan digunakan untuk melatih *base learner* yang berbeda dengan tipe yang sama.

Sumber : Wang, 2011

Input: Data set $D = \{(x_1, y_1), (x_2, y_2), \dots, (x_m, y_m)\}$;

Base learning algorithm L ;

Number of learning rounds T .

Process:

For $t = 1, 2, \dots, T$:

$D_t = \text{Bootstrap}(D)$; % Generate a bootstrap sample from D

$h_t = L(D_t)$ % Train a base learner h_t from the bootstrap sample

end.

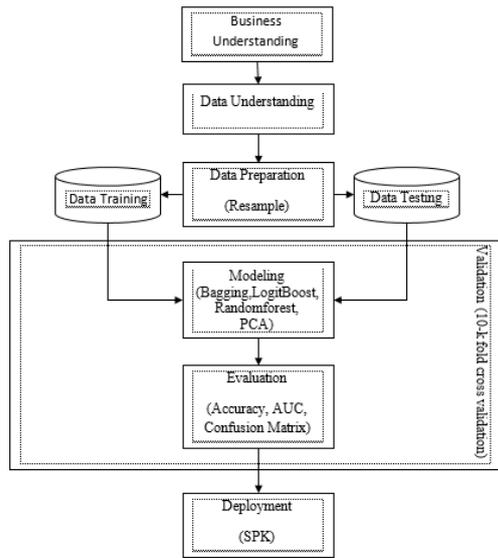
Output: $H(x) = \arg \max_{y \in Y} \sum_{t=1}^T I(y = h_t(x))$ % the value of $I(\alpha)$ is 1 if α is true
% and 0 otherwise

6. Logitboost

Logitboost merupakan algoritma penguat yang berpengaruh untuk klasifikasi. Algoritma Logitboost digunakan sebagai *meta-classifier* bersama dengan Random Forests (RF) sebagai classifier lemah. Kinerja teknik yang diusulkan dievaluasi dalam konteks deteksi intrusi (Kamarudin & Maple, 2017). Pengenalan algoritma Logitboost dirancang sebagai solusi alternatif untuk mengatasi keterbatasan Adaboost dalam menangani noise dan outlier. Algoritma Logitboost menggunakan log-kemungkinan binomial yang mengubah fungsi kerugian secara linear. Sebaliknya, Adaboost menggunakan fungsi kerugian eksponensial yang berubah secara eksponensial dengan kesalahan klasifikasi. Inilah alasan mengapa Logitboost cenderung kurang sensitif terhadap pencilan dan kebisingan (Kamarudin & Maple, 2017).

METODE PENELITIAN

Pada tahap ini subset fitur hasil seleksi akan diklasifikasi dengan menggunakan beberapa algoritme yaitu Logitboost, Bagging, dan RF. secara garis besar skema alur penelitian dapat dilihat pada Gambar 1. Klasifikasi dilakukan dengan bantuan perangkat lunak Weka.



(Sumber: Hasil Penelitian (2019))

Gambar 1. Skema alur penelitian

1. Business Understanding

Tahapan ini menjadi proses dimana penelitian mengarah kepada tujuannya. Perumusan dan definisi masalah diselesaikan pada tahap ini. Penelitian akan menggunakan dataset dari *Breast Cancer Coimbra* untuk melakukan prediksi terhadap fitur yang mempengaruhi terjadinya kanker payudara. Dari dataset darah dapat ditentukan pengaruh dari data darah yang di gunakan untuk uji coba kanker payudara. Klasifikasi diperlukan dengan *data mining* untuk melakukan klasifikasi dengan berdasarkan kepada data yang diperoleh dari tes darah. Hasil dari penelitian diharapkan dapat dijadikan sebagai acuan untuk mendeteksi kanker payudara.

2. Data Understanding

Data yang akan diklasifikasikan perlu dipahami sebelum penelitian dilakukan. Dataset yang digunakan merupakan data dari dataset *Breast Cancer Coimbra* yang berasal dari UCI. Data kanker payudara ini melibatkan 64 pasien dengan kanker payudara dan 52 kontrol sehat. Ada 10 prediktor, semuanya kuantitatif, dan variabel dependen biner, yang menunjukkan ada tidaknya kanker payudara. Prediktor adalah data antropometrik dan parameter yang dapat dikumpulkan dalam analisis darah rutin. Model prediksi berdasarkan prediksi ini, jika akurat, berpotensi dapat digunakan sebagai biomarker kanker payudara.

3. Data Preparation

Setelah data yang akan diujikan dipahami, selanjutnya adalah melakukan persiapan data. Tahapan awal dilakukan dengan *pre-processing* untuk mempersiapkan dataset agar dapat lebih mudah untuk diklasifikasikan. Pengolahan data pada tahap *pre-processing* dilakukan dengan teknik *Resample*. Sebanyak 117 *record* akan diujikan

dengan 9 atribut dan 2 label. Atribut terdiri dari *Glukosa, Insulin, HOMA, Leptin, Adiponektin, Resistin, MCP1, Age* dan *Body Mass Index (BMI)*. Keterangan merupakan label yang yang berisi kontrol sehat dan pasien .

4. Modeling

Pada tahapan ini dataset diujikan dengan menggunakan teknik *data mining*. Klasifikasi dilakukan dengan menerapkan beberapa algoritma *data mining* yang diusulkan. Pertama, pengujian dilakukan seleksi atribut dengan atribut *evaluator principal components analysis*. Selanjutnya dilakukan eksperimen ke dua dengan mengklasifikasikan data seleksi atribut *principal component analysis* dengan menambahkan kalsifikasi *logitBoost, Bagging dan Random Forest* . Setelah pengujian dilakukan, selanjutnya hasil yang diperoleh akan disajikan dalam bentuk tabel dan grafik.

5. Evaluation

Evaluasi dilakukan berdasarkan nilai akurasi, AUC, dan *confussion matrix*. Validasi juga dilakukan dengan *10 k-fold cross validation*. Validasi menggunakan *10-fold cross validation* dilakukan dengan membagi dataset menjadi 10 bagian, satu bagian sebagai data uji, sedangkan bagian lain sebagai data latih. Proses validasi dilakukan berulang, mulai bagian pertama sebagai data uji sampai bagian kesepuluh, sehingga semua data dalam dataset diuji. Tahapan ini akan melihat secara mendetail mengenai hasil kinerja algoritma yang diujikan pada dataset.

6. Deployment

Algoritma dengan nilai akurasi terbaik akan diimplementasikan dalam sistem penunjang keputusan sehingga dapat digunakan sebagai pertimbangan dalam pengambilan keputusan. Selain itu, model yang digunakan juga memiliki nilai AUC yang mencukupi. Kemudian, hasil dari implementasi akan diterapkan untuk mengklasifikasi 50 *record* data testing.

HASIL DAN PEMBAHASAN

Eksperimen dengan menggunakan fitur seleksi
Tabel 1. *Confusion Matrix* dan Akurasi

	True Sehat	True Pasien	Akurasi
Pred.Sehat	38	14	0.731
Pred.Pasien	12	52	0.813

Sumber: Hasil Penelitian (2019)

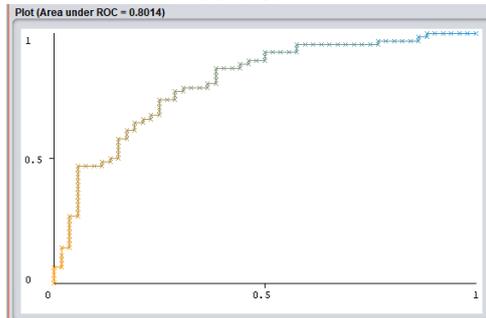
Tabel di atas merupakan *Confusion Matrix* dari klasifikasi *Principal Component Analysis* berbasis *Logitboost* dengan akurasi yang dihasilkan sebesar 77.5862 %.

Tabel 2. *Confusion Matrix* dan Akurasi *Principal Component Analysis* Menggunakan *Bagging*

	True Sehat	True Pasien	Recall
Pred.Sehat	33	19	0.635
Pred.Pasien	14	50	0.781
Precision	0.702	0.725	

Sumber: Hasil Penelitian (2019)

Tabel di atas merupakan *Confusion Matrix* dari klasifikasi *Principal Component Analysis* Menggunakan *Bagging* dengan akurasi yang dihasilkan sebesar 71.5517 %.



Sumber: Hasil Penelitian (2019)
Gambar 2. Hasil AUC Label Sehat

1. *Confusion Matrix* dan *Accuracy*

Tabel 3. *Confusion Matrix* dan Akurasi *Random Forest*

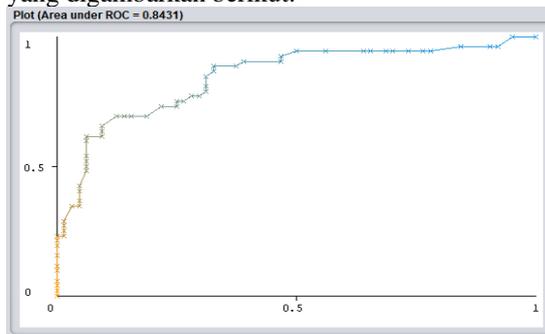
	True Sehat	True Pasien	Precision
Pred.Sehat	36	16	0.818
Pred.Pasien	8	56	0.778
Recall	0.692	0.875	

Sumber: Hasil Penelitian (2019)

Tabel di atas merupakan *Confusion Matrix* dari klasifikasi algoritma *Random Forest* dengan akurasi yang dihasilkan sebesar 79.3103 %.

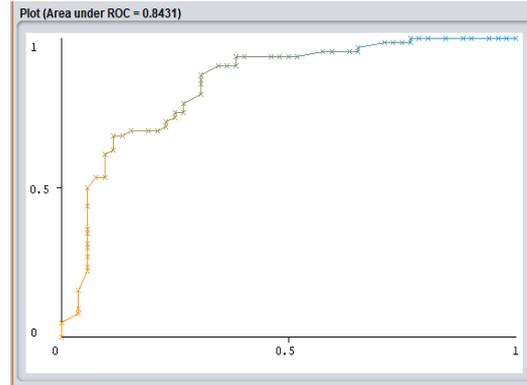
a. AUC

Area Under the Curve (AUC) yang dihasilkan dari klasifikasi algoritma RF sebesar 0.843 seperti yang digambarkan berikut:



Sumber: Hasil Penelitian (2019)

Gambar 3. Hasil AUC Label Sehat Algoritma RF



Sumber: Hasil Penelitian (2019)

Gambar 4. Hasil AUC Label Pasien Algoritma RF *Area Under the Curve* (AUC) yang dihasilkan dari klasifikasi algoritma RF sebesar 0,8431

Tabel 4. Pengujian Algoritma *LogitBoost*, *Bagging* dan *Random Forest*

	<i>LogitBoos</i> <i>t</i>	<i>Bagging</i>	<i>Random</i> <i>Forest</i>
<i>Accuracy</i>	77.5862 %	71.5517 %	79.3103 %
<i>Precision</i>	77.05%	71,05%	79,06%
<i>Recall</i>	77.06%	71,06%	79,03%
AUC	0,822	0,801	0,843

Sumber: Hasil Penelitian (2019)

Berdasarkan hasil pengujian model algoritma *LogitBoost*, *Bagging* dan *Random Forest*, nilai akurasi tertinggi dihasilkan oleh algoritma *Random Forest* sebesar 79.3103 % dengan nilai AUC sebesar 0,843.

KESIMPULAN

Penelitian ini dibuat untuk mendeteksi penyakit kanker payudara dengan tingkat akurasi yang baik. Untuk memprediksi permasalahan tersebut digunakan model klasifikasi yaitu *Principal Component Analysis* berbasis *Random Forest* dimana hasil evaluasi model tersebut akan dilihat nilai akurasinya.

Berdasarkan penelitian tingkat akurasi prediksi dengan *Principal Component Analysis* berbasis *LogitBoost* sebesar 77.5862 %, *Principal Component Analysis* berbasis *Bagging* sebesar 71.5517 %, sedangkan *Principal Component Analysis* berbasis *Random Forest* sebesar 79.3103 %. Hal ini membuktikan bahwa hasil evaluasi penelitian *Principal Component Analysis* berbasis *Random Forest* mampu memberikan perbaikan tingkat kesalahan klasifikasi sebesar 18.33%.

Dengan metode *Principal Component Analysis* berbasis *Random Forest* menunjukkan akurasi yang tidak terlalu tinggi. Dari penelitian yang dilakukan diketahui bahwa metode seleksi dapat meningkatkan hasil diagnosis klasifikasi kanker payudara secara signifikan dengan jumlah fitur yang lebih kecil.

REFERENSI

- Adiwijaya, A. (2018). Deteksi Kanker Berdasarkan Klasifikasi Microarray Data. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 2, 181–186.
- Arifin, T., Riana, D., & Hapsari, G. I. (2014). Klasifikasi Statistik Tekstur Sel Pap Smear Dengan Decesion Tree. *Jurnal Informatika*, 1, 1–7.
- Arifin, T., Riana, D., & Hapsari, G. I. (2014). Klasifikasi Statistik Tekstur Sel Pap Smear Dengan Decesion Tree. *Jurnal Informatika*, (1), 1–7.
- Assiri, A., Kamel, H. F., & Hassanien, M. F. (2015). Resistin, visfatin, adiponectin, and leptin: risk of breast cancer in pre-and postmenopausal saudi females and their possible diagnostic and predictive implications as novel biomarkers. *Disease Markers*.
- Assiri, A. M., & Kamel, H. F. (2016). Evaluation of diagnostic and predictive value of serum adipokines: Leptin, resistin and visfatin in postmenopausal breast cancer. *Obesity Research & Clinical Practice*, 10, 442–453.
- Cole, K. D., He, H. J., & Wang, L. (2013). Breast cancer biomarker measurements and standards. *PROTEOMICS–Clinical Applications*, 7(1–2), 17–29.
- Crisóstomo, J., Matafome, P., Santos-Silva, D., Gomes, A. L., Gomes, M., Patrício, M., ... & Seça, R. (2016). Hyperresistinemia and metabolic dysregulation: a risky crosstalk in obese breast cancer. *Endocrine*, 433–442.
- Dalamaga, M., Sotiropoulos, G., Karmaniolas, K., Pelekanos, N., Papadavid, E., & Lekka, A. (2013). Serum resistin: a biomarker of breast cancer in postmenopausal women? Association with clinicopathological characteristics, tumor markers, inflammatory and metabolic parameters. *Clinical Biochemistry*, 46, 584–590.
- Ghaderyan, P., Abbasi, A., & Sedaaghi, M. H. (2014). An efficient seizure prediction method using KNN-based undersampling and linear frequency measures. *Journal of Neuroscience Methods*, 232, 134–142.
- Hasan, H., & Tahir, N. M. (2010). Feature selection of breast cancer based on principal component analysis. In *2010 6th International Colloquium on Signal Processing & Its Applications*, 1–4.
- Hwa, H. L., Kuo, W. H., Chang, L. Y., Wang, M. Y., Tung, T. H., Chang, K. J., & Hsieh, F. J. (2008). Prediction of breast cancer and lymph node metastatic status with tumour markers using logistic regression models. *Journal of Evaluation in Clinical Practice*, 14(2), 275–280.
- Kamarudin, M. H., & Maple, C. (2017). *A LogitBoost-Based Algorithm for Detecting Known and Unknown Web Attacks*. 5.
- KEMENTERIAN KESEHATAN REPUBLIK INDONESIA
PANDUAN
PENATALAKSANAAN KANKER
PAYUDARA. (n.d.). *Kanker payudara*.
- Lin, L., Wang, F., Xie, X., & Zhong, S. (2017). Random forests-based extreme learning machine ensemble for multi-regime time series prediction. *Expert Systems with Applications*, 83, 164–176.
- Nasution, M. Z. (2016). PENERAPAN PRINCIPAL COMPONENT ANALYSIS (PCA) DALAM PENENTUAN FAKTOR DOMINAN YANG MEMPENGARUHI PENGIDAP KANKER SERVIKS (Studi Kasus: Cervical Cancer Dataset). *Jurnal Mantik*, 10, 442–453.
- RI, K. (2015). *Panduan penatalaksanaan kanker payudara. Kementerian Kesehatan Republik Indonesia. Komite Penanggulangan Kanker Nasional*.
- Santillán-Benítez, J. G., Mendieta-Zerón, H., Gómez-Oliván, L. M., Torres-Juárez, J. J., González-Bañales, J. M., Hernández-Peña, L. V., & Ordóñez-Quiroz, A. (2013). The tetrad BMI, leptin, leptin/adiponectin (L/a) ratio and CA 15-3 are reliable biomarkers of breast cancer. *Journal of Clinical Laboratory Analysis*, 27, 12–20.
- Saputra, W. S. J., Sujatmika, A. R., & Arifin, A. Z. (2011). *Seleksi Fitur Menggunakan Random Forest Dan Neural Network*. 2011(Ies), 978–979.
- Wahyu SJ, S., Arif Rahman, S., & Agus Zainal, A. (2016). Penerapan metode seleksi fitur untuk meningkatkan hasil diagnosis kanker payudara. *Simetris. Jurnal Teknik Mesin, Elektro Dan Ilmu Komputer*, 283–294.
- Wibawa, M. S., & Novianti, K. D. P. (2017). Reduksi Fitur Untuk Optimalisasi Klasifikasi Tumor Payudara Berdasarkan Data Citra FNA. *E-Proceedings KNS&I STIKOM Bali*, 73–78.