**JURNAL RESPONSIF**, Vol.1 No.1 Agustus 2019, pp. 64~69
E-ISSN**:** 2685-6964 

# Prediksi Harapan Hidup Pasien Kanker Paru Pasca Operasi Bedah Toraks Menggunakan Boosted k-Nearest Neighbor

**Rizki Tri Prasetio[1], Sari Susanti[2]**

[1]Universitas BSI
e-mail: rizki@univbsi.ac.id

[2]Universitas BSI
e-mail: sari.srq@bsi.ac.id

**Abstrak**

Kanker paru-paru menempati peringkat enam dari sepuluh penyakit penyebab kematian terbanyak di Indonesia. Faktor penyebab kanker paru-paru didominasi oleh asap rokok. Operasi bedah toraks menjadi salah satu solusi utama untuk kanker paru-paru. Akan tetapi, terdapat banyak resiko dan komplikasi pasca operasi bedah toraks hingga berujung pada kematian. Pada penelitian ini, akan di prediksi harapan hidup pasien kanker paru-paru setelah menjalani kehidupan satu tahun pasca operasi bedah toraks menggunakan *computer aided diagnosis* (CAD). Prediksi ini dilakukan dengan menganalisa kondisi pasien sebelum dan sesudah operasi. Data yang digunakan pada penelitian ini merupakan data sekunder yang berisi 470 data dengan sebaran 400 data pasien yang hidup (*survival*) dan 70 data pasien yang meninggal (*die*). *Adaptive Boost* digunakan sebagai optimasi level algoritma pada algoritma *k-nearest neighbor*. Hasil penelitian menunjukan bahwa metode yang diusulkan menghasilkan akurasi prediksi harapan hidup sebesar 85.11% menggunakan validasi 10 *fold cross validation* dengan parameter *k* pada algoritma *k-nearest neighbor* bernilai 5.
**Kata kunci:** operasi bedah toraks, harapan hidup pasca operasi, *k-nearest neighbor.*

***Abstract***

*Lung cancer ranks sixth out of ten of the leading causes of death in Indonesia. Factors that cause lung cancer are dominated by cigarette smoke. Thoracic surgery is one of the main solutions for lung cancer. However, there are many risks and complications after thoracic surgery that lead to death. In this study, the life expectancy of lung cancer patients after living one year after thoracic surgery using a computer diagnosis (CAD) will be predicted. This prediction is done by analyzing the patient's condition before and after surgery. The data used in this study are secondary data containing 470 data with 400 data distribution of patients who live (survival) and 70 data of patients who die (die). Adaptive Boost is used as an algorithm level optimization that is applied to the k-nearest neighbor algorithm. The results showed that the proposed method made a prediction accuracy by 85.11% using ten fold cross validation with k parameter of k-nearest neighbor is 5.*
***Keywords****: thoracic surgery, life expectancy, k-nearest neighbor.*

## 1. Pendahuluan

Survei *Sample Registration System* (SRS) Indonesia tahun 2014 menyatakan bahwa terdapat 10 penyakit yang paling banyak menyebabkan kematian di Indonesia. Salah satunya adalah kanker paru-paru (Global Health Data Exchange, 2014). Kanker paru-paru menempati peringkat keenam dengan penderita sebanyak 4,9%. Faktor resiko kanker paru-paru disebabkan oleh asap rokok baik pada perokok aktif maupun perokok pasif, faktor lain yang menyebabkan kanker paru-paru adalah polusi udara dan paparan pada lingkungan kerja (Kementerian Kesehatan Republik Indonesia, 2014).

Penanganan pasien penderita kanker paru-paru salah satunya dapat dilakukan dengan pembedahan toraks. Saluran pernafasan terletak hampir seluruhnya di dalam toraks. Sehingga toraks mempunyai peranan penting dalam pernapasan (Asih & Effendy, 2014). Pasien yang membutuhkan pengobatan atau intervensi dengan operasi

*Naskah diterima 24 Juni 2019; direvisi 8 Juli 2019; diterbitkan 15 Agustus 2019*

untuk penyakit paru-paru harus menjalani operasi toraks (Ferguson, 2007).

Operasi bedah toraks menjadi salah satu solusi utama untuk kanker paru-paru. Akan tetapi, terdapat banyak risiko dan komplikasi dari operasi bedah toraks, seperti gangguan syaraf, infeksi dan komplikasi yang fatal hingga kematian (Shields, Reed, LoCicero, & Feins, 2009). Komplikasi banyak terjadi pada penderita penyakit kardiovaskular seperti gangguan jantung dan pembuluh darah dapat mengakibatkan stroke. Sehingga harapan hidup pasca dilakukannya operasi bedah toraks sangat tipis (Asih & Effendy, 2014).

Masalah utama pada penelitian ini adalah bagaimana memprediksi harapan hidup pasien pasca operasi menggunakan *computer aided diagnosis* (CAD) *System*. Penggunaan CAD dapat membantu memprediksi harapan hidup pasien penderita kanker paru-paru melalui analisa kondisi pasien sebelum dan sesudah operasi. Pengumpulan data pasien kanker paru-paru yang menjalani operasi bedah toraks telah dikumpulkan oleh Maciej Zieba dan Jakub M. Tomczak (Zieba, Tomczak, Lubicz, & Swiatek, 2013). Dataset pasien pasca operasi bedah toraks memiliki dua kelas yaitu, meninggal dalam jangka waktu satu tahun (*die*) dan mampu bertahan hidup (*survival*) dengan jumlah sampel data untuk kelas *die* sebanyak 70 dan untuk kelas *survival* sebanyak 400 sampel (Zieba, Tomczak, Lubicz, & Swiatek, 2013).

Penelitian sebelumnya pada sampel data pasien operasi toraks untuk prediksi harapan hidup pasca operasi dilakukan oleh beberapa peneliti menggunakan berbagai algoritma, antara lain: naïve bayes (Hachesu, Moftian, Dehghani, & Soltani, 2017), *rule based classification* (Koklu, Kahramanli, & Allahverdi, 2015), *support vector machine* (Desuky & El Bakrawy , 2016), *multi layer perceptron*, *semi naïve bayesian* (Hui, Zhou, Jiang, Ji, & Chen, 2017) dan *ensemble svm* (Zieba, Tomczak, Lubicz, & Swiatek, 2013).

Berdasarkan uraian masalah, maka dalam penelitian ini akan dilakukan optimasi pada algoritma *k-nearest neighbor* untuk memprediksi harapan hidup pasien pasca menjalani operasi bedah toraks. Algoritma *k-nearest neighbor* dipilih karena memiliki kemampuan untuk mendeteksi dan menganalisa permasalahan yang sifatnya kompleks dan non-linear (Bourquin, Schmidli, Hoogevest, & Leuenberger, 2018) serta perhitungan dilakukan secara pararel sehingga waktu komputasi lebih cepat (Sharma & Chopra, 2013).

Teknik optimasi yang digunakan pada penelitian ini adalah teknik *boosting*, menggunakan algoritma AdaBoost. AdaBoost digunakan karena sangat mudah diimplementasikan (Sharma & Dey, 2013), tidak perlu mengatur parameter (Zieba, Tomczak, Lubicz, & Swiatek, 2013) dan fleksibel sehingga dapat dikombinasikan dengan berbagai algoritma (Huang, Chen, Liu, Tao, & Li, 2019).

## 2. Metode Penelitian

### 2.1. Desain Penelitian

Penelitian didefinisikan oleh *Higher Education Funding Council for England* (HECFE) sebagai penyelidikan yang dilakukan untuk mendapatkan pengetahuan dan pemahaman (Dawson, 2009), serta merujuk pada aktifitas penyelidikan sistematis atau investigasi di suatu bidang, dengan tujuan menemukan atau merevisi fakta, teori, aplikasi dan sebagainya (Berndtsson, Hansson, Olsson, & Lundell, 2008).

Menurut (Dawson, 2009) ada empat metode penelitian yang paling umum digunakan yaitu: *action research*, *experiment*, *case study*, dan *survey*. Pada penelitian ini menggunakan metode *experiment*, yaitu penelitian yang melibatkan penyelidikan kepada beberapa variable menggunakan tes tertentu yang dikendalikan sendiri oleh peneliti. Pada penelitian ini dilakukan beberapa tahapan penelitian sebagai berikut:

1. Pengumpulan Data (*Data Gathering*)
   Pada tahapan ini dijelaskan tentang bagaimana dan darimana data dalam penelitian ini didapatkan. Pada tahap ini juga ditentukan data yang akan diproses.
2. Pengolahan Data Awal (*Data Pre-processing*)
   Pengolahan data awal meliputi pembersihan data, pentransformasian data ke dalam bentuk yang dibutuhkan serta pengelompokan dan penentuan atribut data.
3. Metode yang Diusulkan (*Proposed Method*)
   Setelah pengolahan data awal, lalu dibuatkan model yang sesuai dengan jenis data. Pembagian data ke dalam data pelatihan (training dataset) dan data pengujian (testing dataset) juga diperlukan untuk pembuatan model.

4. Eksperimen dan Pengujian Model (*Model Test and Experiment*)
   Pada tahapan ini, dilakukan eksperimen dan pengujian model terhadap data yang sebelumnya sudah diolah. Perhitungan dengan masing-masing algoritma akan diulang beberapa kali untuk mendapatkan besaran parameter terbaik.
5. Evaluasi dan Validasi Hasil (*Result Evaluation and Validation*)
   Tahap evaluasi merupakan tahap yang terakhir dari kegiatan penelitian, dimana dalam tahap ini hasil dari tahapan eksperimen dan pengujian model sebagai evaluasi.

### 2.2. Pengumpulan Data

Data yang digunakan pada penelitian ini merupakan data sekunder. Data sekunder adalah data yang tidak diperoleh langsung dari obyek penelitian, melainkan telah dikumpulkan oleh pihak lain. Data sekunder yang digunakan pada penelitian ini merupakan kumpulan data dari Wroclaw Thoracic Surgery Centre , data tersebut merupakan data pasien penderita kanker paru-paru yang menjalani operasi bedah thoraks dari tahun 2009 hingga 2014. Dataset ini diambil dari UCI Machine Learning Repository yang diunduh melalui http://archive.ics.uci.edu/ml/datasets/Thoracic+Surgery+Data.

Dataset ini berisi informasi setiap pasien yang diwakili dalam data yang ditetapkan oleh 16 atribut yang merupakan kondisi sebelum dan sesudah pasien menjalani operasi bedah thoraks. 16 atribut tersebut berupa data nominal, numeric dan binary. Dataset pasien pasca operasi bedah toraks memiliki dua kelas yaitu, meninggal dalam jangka waktu satu tahun (*die*) dan mampu bertahan hidup (*survival*) dengan jumlah sampel data untuk kelas *die* sebanyak 70 dan untuk kelas *survival* sebanyak 400 sampel.

Tabel 1. Atribut Dataset

| Atribut | Deskripsi | Tipe Data |
|---|---|---|
| DGN | Diagnosis - kombinasi spesifik kode ICD-10 untuk tumor primer dan sekunder serta lebih dari satu tumor, jika ada | Nominal |
| PRE4 | Jumlah udara yang bisa dihembuskan secara paksa dari paru-paru setelah mengambil nafas sedalam mungkin (FVC) | Numerik |
| PRE5 | Jumlah udara yang telah dihembuskan pada akhir detik pertama dari FVC (FEV1) | Numerik |
| PRE6 | Ukuran kemampuan umum pasien kanker dalam aktivitas sehari-hari (Zubrod Scale) | Nominal |
| PRE7 | Rasa sakit sebelum operasi | Binary |
| PRE8 | *Haemoptysis* sebelum operasi | Binary |
| PRE9 | *Dyspnoea* sebelum operasi | Binary |
| PRE10 | Batuk sebelum operasi | Binary |
| PRE11 | Kondisi lemah sebelum operasi | Binary |
| PRE14 | Ukuran tumor (TNM) | Nominal |
| PRE17 | Diabetes | Binary |
| PRE19 | *myocardial infarction* (MI) hingga 6 bulan | Binary |
| PRE25 | Penyakit yang menyerang arteri/aliran darah (PAD) | Binary |
| PRE30 | Merokok | Binary |
| PRE32 | Asma | Binary |
| AGE | Usia saat operasi | Numerik |
| RISK | Mampu bertahan hidup setelah 1 tahun \| bernilai T jika meninggal | Binary |

### 2.3. Pengolahan Data Awal

Pengolahan data awal merupakan tindak lanjut dari pengumpulan data. Pengolahan data awal dimulai dari melakukan transformasi data dari nominal dan biner secara keseluruhan menjadi numerik agar dapat diproses oleh algoritma *k-nearest neighbor.*

Setelah proses transformasi data, dilakukan normalisasi data untuk menskalakan dataset agar mudah dalam proses perhitungan.

Tahap selanjutnya adalah membagi data menjadi *data training* dan *data testing* menggunakan *cross validation*.

### 2.4. Metode yang Diusulkan

Metode yang diusulkan yaitu menerapkan algoritma *k-nearest neighbor* sebagai algoritma *classifier* yang telah dioptimasi oleh algoritma *boosting AdaBoost*. Hasil klasifikasi kemudian di evaluasi akurasinya menggunakan *confussion matrix*.

Pengolahan data awal dimulai dari transformasi data dari tipe data nominal dan biner ke numerik agar dapat dihitung oleh algoritma *k-nearest neighbor.*

Tahap berikutnya melakukan normalisasi menggunakan *z-transformation* untuk lebih memudahkan dalam perhitungan algoritma. Selanjutnya data akan dibagi menjadi *data training* dan *data testing* menggunakan *cross validation.*

Setelah itu data training akan dilatih dengan algoritma *k-nearest neighbor* yang dioptimasi dengan algoritma *AdaBoost* dengan iterasi sebanyak 10. Setelah didapatkan model yang paling optimal. Model tersebut akan diterapkan untuk menguji model terhadap data *testing*.

Setelah menguji model pada data *testing*, dilakukan evaluasi terhadap model dengan cara menghitung akurasi yang dihasilkan dengan *confussion matrix*. Proses ini akan berulang seterusnya sebanyak 10 kali hingga mendapatkan hasil akurasi yang paling optimal.

### 2.5. Eksperimen dan Pengujian Model

Penelitian yang dilakukan dalam eksperimen ini menggunakan komputer untuk melakukan proses perhitungan terhadap model yang diusulkan. Tahapan eksperimen pada penelitian ini adalah:

1. Menyiapkan 2 dataset untuk eksperimen, dataset *training* dan dataset *testing*.
2. Mendesain arsitektur *neural network* dengan optimasi *AdaBoost*
3. Melakukan *training* dan *testing* terhadap model *k-nearest neighbor* dan mencatat hasil akurasinya.
4. Mengembangkan aplikasi.

### 2.6. Evaluasi dan Validasi Hasil

Pada tahap ini akan dilakukan proses pengujian model yang dihasilkan oleh tool Rapidminer dengan mengevaluasi perbandingan hasil akurasi seluruh eksperimen yaitu eksperimen *k-nearest neighbor*. Sementara itu, evaluasi yang digunakan adalah *confussion matrix*.

## 3. Hasil dan Pembahasan

Hasil dalam penelitian dilakukan dalam dua eksperimen yaitu eksperimen terhadap algoritma *k-nearest neighbor* tanpa optimasi, serta eksperimen terhadap algoritma *k-nearest neighbor* dengan menggunakan AdaBoost.

### 3.1. Hasil Eksperimen dan Pengujian

Eksperimen pada dataset *thoracic surgery* menggunakan algoritma *k-nearest neighbor* tanpa menggunakan optimasi apapun. Eksperimen dilakukan sebanyak satu kali menggunakan *cross* validation. Indikator untuk mengetahui hasil terbaik ditunjukan oleh besarnya nilai akurasi untuk masing-masing eksperimen.

Tabel 2. Hasil Eksperimen dengan *k-Nearest Neighbor*

| Validasi | Akurasi |
|---|---|
| *Cross Validation* | 77.66% |

Hasil eksperimen pertama menunjukan akurasi prediksi harapan hidup pasien pasca operasi bedah toraks menggunakan algoritma *k-nearest neighbor* sebesar 77.66%.

Eksperimen selanjutnya menggunakan algoritma *k-nearest neighbor* dengan optimasi *AdaBoost*. Eksperimen dilakukan sebanyak satu kali menggunakan *cross* validation. Indikator untuk mengetahui hasil terbaik ditunjukan oleh besarnya nilai akurasi untuk masing-masing eksperimen.

Tabel 3. Hasil Eksperimen dengan *k-Nearest Neighbor + AdaBoost*

| Validasi | Akurasi |
|---|---|
| *Cross Validation* | 85.11% |

Hasil eksperimen kedua menunjukan akurasi prediksi harapan hidup pasien pasca operasi bedah toraks menggunakan algoritma *k-nearest neighbor* dan *AdaBoost* sebesar 85.11%.

### 3.2. Pembahasan

Berdasarkan hasil eksperimen yang telah dilakukan maka dapat terlihat bahwa metode yang diusulkan dapat meningkatkan kinerja algoritma *k-nearest neighbor* sebesar 7.45% dari 77.66% menjadi 85.11%. perbandingan hasil eksperimen dapat dilihat pada tabel 4.

Tabel 4. Perbandingan Hasil Eksperimen

| Validasi | *k-Nearest Neighbor* | *k-Nearest Neighbor + AdaBoost* |
|---|---|---|
| *Cross Validation* | 77.66% | 85.11% |

**4. Kesimpulan**

Penelitian ini mengkombinasikan teknik *boosting* AdaBoost sebagai optimasi level algoritma pada algoritma *neural network* untuk prediksi harapan hidup pasien kanker paru-paru pasca operasi bedah thoraks. Berdasarkan hasil eksperimen pada penelitian ini, maka dapat ditarik kesimpulan bahwa algoritma *AdaBoost* dapat meningkatkan performa algoritma *k-nearest neighbor* dalam memprediksi harapan hidup pasien pasca operasi bedah toraks sebesar 7.45% dari 77.66% menjadi 85.11%.

Pada penelitian berikutnya, dapat diterapkan optimasi pada level data untuk mengatasi ketidakseimbangan data dikarenakan sampel dataset yang tidak seimbang antara kelas *survival* dan *die*. Selain itu juga dapat diterapkan optimasi parameter pada parameter *k* untuk lebih mengoptimalkan kinerja algoritma.

**Referensi**


Asih, N. G., & Effendy, C. (2014). *Keperawatan Medical Bedah.* Jakarta: Buku Kedokteran EGC.

Berndtsson, M., Hansson, J., Olsson, B., & Lundell, B. (2008). *Thesis Projects : A Guide for Students in Computer Science and Information Systems 2nd Edition.* London: Springer.

Bourquin, J., Schmidli, H., Hoogevest, P. v., & Leuenberger, H. (2018). Advantages of Artificial Neural Networks (ANNs) as alternative modelling technique for data sets showing non-linear relationships using data from a galenical study on a solid dosage form. *European Journal of Pharmaceutical Sciences*, 5-16.

Dawson, C. W. (2009). *Projects in Computing and Information Systems : A Student's Guide 2nd Edition.* London: Pearson Education Limited.

Desuky, A. S., & El Bakrawy , L. M. (2016). Improved prediction of post-operative life expectancy after Thoracic Surgery. *Advance in System Science Application*, 70-80.

Ferguson, M. K. (2007). *Thoracic Surgery Atlas.* Elsevier.

Global Health Data Exchange. (2014). *Indonesia Sample Registration System - Deaths.* Washington: GHDx.

Hachesu, P. R., Moftian, N., Dehghani, M., & Soltani, T. S. (2017). Analyzing a Lung Cancer Patient Dataset with the Focus on Predicting Survival Rate One Year after Thoracic Surgery. *Asian Pacific Journal of Cancer Prevention*, 1531-1536.

Huang, Q., Chen, Y., Liu, L., Tao, D., & Li, X. (2019). On Combining Biclustering Mining and AdaBoost for Breast Tumor Classification. *IEEE Transactions on Knowledge and Data Engineering*, 1.

Hui, B., Zhou, H., Jiang, Y., Ji, L., & Chen, J. (2017). The Research of Postoperative Life Expectancy of Lung Cancer Based on Semi Naive Bayesian. *Computer Science and Artificial Intelligence(CSAI)*, 17-19.

Kementerian Kesehatan Republik Indonesia. (2014). *InfoDATIN Kanker.* Jakarta: Pusat Data dan Informasi Kementerian Kesehatan Republik Indonesia.

Koklu, M., Kahramanli, H., & Allahverdi, N. (2015). APPLICATIONS OF RULE BASED CLASSIFICATION TECHNIQUES FOR THORACIC SURGERY. *Managing Intellectual Capital and Innovation for Sustainable and Inclusive Society Management, Knowledge and Learning Joint International Conference* (pp. 1991-1998). Bari: Technology, Innovation and Industrial Management.

Sharma, A., & Chopra, A. (2013). Artificial Neural Networks: Applications In Management. *IOSR Journal of Business and Management (IOSR-JBM)*, 32-40.

Sharma, A., & Dey, S. (2013). A boosted SVM based sentiment analysis approach for online opinionated text. *Proceedings of the 2013 Research in Adaptive and Convergent Systems* (pp. 28-34). Montreal: ACM.

Shields, T. W., Reed, C. E., LoCicero, J., & Feins, R. H. (2009). *General*

*Thoracic Surgery*. Philadelphia: Wolters Kluwer Business.

Zieba, M., Tomczak, J. M., Lubicz, M., & Swiatek, J. (2013). Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied Soft Computing*, 99-108.

Zieba, M., Tomczak, J. M., Lubicz, M., & Swiatek, J. (2014). Boosted SVM for extracting rules from imbalanced data in application to prediction of the post-operative life expectancy in the lung cancer patients. *Applied Soft Computing*, 99-108.