



Ukuran Akurasi Klasifikasi Penyakit Mesothelioma Menggunakan Algoritma *K-Nearest Neighbor* dan *Backward Elimination*

Maxsi Ary

Universitas Bina Sarana Informatika
maxsiary@gmail.com

Dyah Ayu Feby Rismiati

Universitas Bina Sarana Informatika
dyahayu223@gmail.com

Abstrak

Tujuan dari penelitian ini adalah mengukur tingkat akurasi hasil klasifikasi objek menggunakan algoritma *K-Nearest Neighbor* dan *Backward Elimination*. Pengukuran tingkat akurasi diperlukan untuk menentukan tindakan selanjutnya, misalnya dalam menentukan deteksi awal suatu penyakit mesothelioma. Mesothelioma adalah kanker langka yang mempengaruhi dinding sel tipis dari organ dan struktur internal tubuh manusia yang dapat ditemukan di pleura, peritoneum, dan jantung. Pengklasifikasian suatu objek dapat digunakan dengan beberapa metode. Proses klasifikasi data dari suatu objek dapat mempermudah dalam menentukan tindakan selanjutnya. Nilai akurasi pengukuran algoritma *K-Nearest Neighbor* digunakan sebagai nilai awal penentuan tingkat akurasi setelah dilakukan seleksi fitur *backward elimination*. Algoritma *K-Nearest Neighbor* digunakan untuk klasifikasi pada objek. *Backward Elimination* digunakan untuk memilih atribut yang paling relevan pada proses klasifikasi. Proses seleksi fitur menggunakan *Backward Elimination* dilakukan bersamaan dengan proses pemodelan menggunakan *K-Nearest Neighbor* untuk menemukan subset fitur (set atribut) yang paling relevan. Objek penelitian diperoleh dari machine learning repository dengan nama dataset penyakit mesothelioma. Transformasi data dikelompokkan kedalam data training dan data testing. Hasil yang menarik pada penelitian adalah nilai tingkat akurasi lebih besar dari nilai awal dan set atribut terbaik setelah dilakukan seleksi fitur *backward elimination*.

Kata Kunci: akurasi hasil klasifikasi, penyakit mesothelioma, algoritma *K-Nearest Neighbor*, *Backward Elimination*.

Abstract

The purpose of this study is to measure the level of accuracy of the classification of objects using the *K-Nearest Neighbor* and *Backward Elimination* algorithms. Measurement of the level of accuracy is needed to determine the next action, for example in determining the initial detection of a mesothelioma disease. Mesothelioma is a rare cancer that affects the thin cell walls of the organs and internal structures of the human body which can be found in the pleura, peritoneum and heart. Classifying an object can be used with several methods. The process of classification of data from an object can make it easier to determine the next action. The accuracy of measurement of the *K-Nearest Neighbor* algorithm is used as the initial value of determining the level of accuracy after the *backward elimination* feature selection was done. *K-Nearest Neighbor* algorithm is used to classify objects. *Backward Elimination* is used to select the most relevant attributes in the classification process. The feature selection process uses *backward elimination* done together with the modeling process using *K-Nearest Neighbor* to find the most relevant feature subset (set of attributes). The object of the research was obtained from a machine learning repository with the name dataset mesothelioma disease. Data transformation is grouped into training data and testing data. An interesting result in the study is that the value of the accuracy level is greater than the initial value and the best set of

attributes after the backward elimination feature selection is done.

Keywords: accuracy of classification results, mesothelioma disease, K-Nearest Neighbor algorithm, Backward Elimination.

1. Pendahuluan

Algoritma K-Nearest Neighbor adalah pendekatan yang digunakan dalam pengklasifikasian suatu data ke data yang sudah ada secara mudah dan efisien (Mustakim & Oktaviani F, 2016). Klasifikasi data menggunakan Algoritma K-Nearest Neighbor termasuk kedalam algoritma supervised (Han & Kamber, 2006). Konsep mendasar dari algoritma K-Nearest Neighbor yaitu mencari jarak paling terdekat diantara data yang terevaluasi dengan sejumlah K (*neighbor*) paling dekat dengan data uji (Sonatha, 2013). K-Nearest Neighbor memiliki kelebihan pada data training yang memiliki banyak *noisy* serta efektif terhadap jumlah data *training* tinggi atau besar. Namun kekurangan dari K-Nearest Neighbor masih perlu penentuan nilai K dan pemilihan atribut terbaik (Bahri & Maliki, 2012).

Pemilihan atribut terbaik dapat dilakukan dengan menggunakan metode seleksi fitur. Tujuan seleksi fitur adalah untuk mengidentifikasi fitur atau atribut dalam kumpulan data yang sama pentingnya, dan membuang semua atribut lain yang tidak relevan (Yunita, 2017). Backward Elimination adalah metode seleksi fitur yang dilakukan dengan cara pemilihan kedepan yakni menguji semua atribut kemudian menghapus atribut-atribut yang dianggap tidak relevan. Semua atribut diproses satu per satu, jika atribut dianggap tidak berpengaruh atau tidak relevan dalam model maka akan dihapus dari model (Bode, 2017).

Penyakit Mesothelioma telah dipelajari oleh peneliti yang berbeda dalam beberapa tahun terakhir. (Er. & Temurtas, 2012) menggunakan *Mesothelioma's Disease Dataset* untuk mengklasifikasikan data berdasarkan kesamaan dan prototipenya menggunakan metode *Probabilistic Neural Network* dan diperoleh akurasi sebesar 96,30%. (Nilashi, Roudbaraki, Farahmand, Branch, & Branch, 2017) menggunakan *Expectation Maximization* dan *Naive Bayes* untuk melakukan prediksi penyakit Mesothelioma dengan akurasi sebesar 93,21%, sedangkan (Khan, Sikander, Anwar, & Khan, 2018) melakukan klasifikasi penyakit Mesothelioma dengan menggunakan metode *Support Vector Machine* (SVM) memiliki akurasi yang lebih besar yaitu 98%. Namun penelitian lebih lanjut diperlukan untuk menemukan algoritma yang akan selalu memiliki rasio klasifikasi tertinggi.

Mesothelioma adalah kanker langka yang mempengaruhi dinding sel tipis dari organ dan struktur

internal tubuh manusia yang dapat ditemukan di pleura, peritoneum, dan jantung. Kanker ini tergolong agresif dan banyak penderitanya yang tidak berhasil terobati (Marianti, 2018). Penyebab utama kanker Mesothelioma adalah paparan mineral asbes, radiasi, dan disposisi genetika (Zervos, Bizekis, & Pass, 2008).

Penyakit Mesothelioma jarang terjadi, sehingga sulit bagi konsultan medis untuk menganalisis perawatan yang tepat. Identifikasi Mesothelioma melibatkan pengujian *CT-scan*, *MRI*, *PET-scan*, tes darah *MESOMark*, *SOMAmer*, dan biopsi untuk menyelidiki pertumbuhan kanker pada jaringan (Khan, Sikandar, Anwar, & Khan, 2018). Mesothelioma akan terdeteksi ketika seorang pasien mengunjungi dokter untuk memeriksakan gejala yang terasa seperti sesak napas, nyeri di dada atau punggung, batuk terus-menerus atau gejala lainnya (Molinari, 2019).

Untuk mengurangi kesalahan dan mempermudah dalam deteksi penyakit Mesothelioma, dapat dilakukan penerapan dan pemanfaatan teknik data mining. Data mining telah digunakan dalam menganalisis data untuk mendapatkan informasi yang berguna dan telah menjadi fokus perhatian di bidang akademik, industri, ekonomi, medis dan kalangan bisnis (Tutuncu & Cataltas, 2017). Penggunaan teknik data mining diharapkan dapat memberikan pengetahuan yang sebelumnya tersembunyi didalam gudang data sehingga menjadi informasi yang berharga (Huda, 2010).

Teknik data mining dibagi berdasarkan tugas yang dapat dilakukan, salah satunya adalah klasifikasi (Nofriansyah, 2014). Klasifikasi memiliki tujuan untuk mengelompokkan data menjadi beberapa kelompok. Proses pengelompokan menggunakan acuan data yang telah diketahui kelompoknya. Beberapa algoritma yang termasuk dalam klasifikasi seperti *Decision Tree*, *Naive Bayes*, *K-Nearest Neighbor*, *Neural Network*, *Support Vector Machine* (Adinugroho & Sari, 2018)

2. Kajian Pustaka Dan Perumusan Hipotesis

2.1. Klasifikasi K-Nearest Neighbor

Klasifikasi adalah contoh *task machine learning* tipe *supervised* untuk melakukan prediksi. Dalam klasifikasi kita dapat menentukan orang atau objek kedalam suatu kategori tertentu. Informasi tentang objek sebelumnya digunakan sebagai bahan algoritma untuk mendapatkan *rule* (Alfisahrin, 2014).

Pada klasifikasi yang menjadi target *feature* untuk diprediksi adalah *feature*, istilah kelas (*class*) dan dapat dibagi menjadi kategori yang disebut dengan istilah level. Saat ini telah banyak teknik/metode klasifikasi yang dikembangkan untuk menyelesaikan masalah klasifikasi 2 class. Sehingga untuk menyelesaikan masalah klasifikasi multiclass dapat digunakan

kombinasi teknik/metode yang biasa digunakan untuk menyelesaikan klasifikasi 2 class. Berikut ini adalah beberapa nama algoritma tipe supervised learning, yaitu (Faisal, 2017):

- a. *K-Nearest Neighbor*
- b. *Naive Bayes*
- c. *Support Vector Machine*
- d. *Decision Trees*
- e. *Linear Regression*
- f. *Neural Network*

Menurut (Prasetyo, 2014) metode *Nearest Neighbor* termasuk dalam klasifikasi yang *lazy learner* karena menunda proses pelatihan (atau bahkan tidak melakukan sama sekali) sampai ada data uji yang ingin diketahui label kelasnya, maka metode baru akan menjalankan algoritmanya. *K-Nearest Neighbor* (KNN) adalah metode klasifikasi berdasarkan data pembelajaran yang jaraknya paling dekat dengan objek tersebut. Tujuan dari algoritma *Nearest Neighbor* adalah mengklasifikasikan obyek baru berdasarkan atribut dan *training sample*. Diberikan suatu titik *query*, selanjutnya akan ditemukan sejumlah K obyek atau (titik *training*) yang paling dekat dengan titik *query*. Nilai prediksi dari *query instance* akan ditentukan berdasarkan klasifikasi ketetanggaan (Jumadi & Winarko, 2015).

2.2. Seleksi Fitur Backward Elimination

Feature Selection adalah suatu proses yang dilakukan untuk menemukan subhimpunan dari himpunan fitur yang tersedia untuk meningkatkan aplikasi dari suatu algoritma pembelajaran (Kesuma, 2011:61). Menurut (Chandrashekar & Sahin, 2014:17), *Feature Selection* dibedakan kedalam 3 kelompok, yaitu:

1. *Feature Selection Tipe Wrapper*

Feature Selection tipe *wrapper* adalah *feature selection* yang melakukan pemilihan bersamaan dengan pelaksanaan pemodelan. *Selection* tipe ini menggunakan suatu *criterion* yang memanfaatkan *classification rate* atau perkiraan akurasi yang dihasilkan dari pemodelan. Metode *wrapper* memiliki hasil yang lebih baik dibandingkan metode filter karena metode *wrapper* mempertimbangkan bias dari algoritma pembelajaran yang digunakan. Namun, metode *wrapper* memiliki biaya komputasi yang mahal karena menjalankan algoritma klasifikasi berkali-kali. Contoh dari metode *wrapper* antara lain *Backward Elimination* dan *Forward Selection* (Chandani, Wahono, & Purwanto, 2015:57).

2. *Feature Selection Tipe Filter*

Feature selection tipe filter adalah metode seleksi fitur yang dilakukan dengan cara pencarian independen yaitu mengeluarkan atribut yang tidak relevan sebelum melakukan proses klasifikasi.

3. *Feature Selection Tipe Embedded*

Berbeda dengan pendekatan *wrapper*, memperlakukan seleksi fitur sebagai pembungkus di sekitar proses klasifikasi, pendekatan *embedded* (tertanam) menyematkan seleksi dalam algoritma klasifikasi.

Backward Elimination merupakan suatu metode yang termasuk kedalam *feature selection* tipe *wrapper* yang memiliki fungsi untuk mengoptimalkan kinerja suatu model. Pemilihan variabel dilakukan dengan cara pemilihan kedepan yakni menguji semua variabel kemudian menghapus variabel-variabel yang dianggap tidak signifikan. Variabel diproses satu per satu, jika variabel dianggap tidak berpengaruh atau tidak signifikan dalam model maka akan dihapus dari model. Iterasi diulangi hingga tidak ada perbaikan yang diamati pada penghapusan fitur (Bode, 2017).

2.3. Evaluasi Dan Validasi

Menurut (Widodo, Handayanto, & Herlawati, 2013) hingga saat ini permasalahan mengenai prediksi terhadap performa suatu mesin pembelajaran (*machine learning*) merupakan topik yang menarik dan masih kontroversial. Membandingkan performa antara beragam metode pada mesin pembelajaran merupakan permasalahan yang tidak mudah. Untuk saat ini secara sederhana diasumsikan bahwa performa suatu metode diukur dari sejauh mana kemampuan metode tersebut dalam melakukan klasifikasi secara akurat terhadap data tertentu yang diuji.

Confussion Matrix merupakan alat yang berguna untuk menganalisis seberapa baik *classifier* dapat mengenali tupel dari kelas yang berbeda. *Confusion matrix* memberikan rincian klasifikasi, kelas yang diprediksi akan ditampilkan dibagian atas matrix dan kelas yang diobservasi ditampilkan dibagian kiri (Sulaehani, 2016).

Untuk menghitung nilai akurasi dapat dilakukan dengan rumus (1) perhitungan sebagai berikut:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\% \quad (1)$$

Keterangan:

TP = Proporsi sampel bernilai true yang diprediksi secara benar.

FN = Proporsi sampel bernilai false yang diprediksi sebagai sampel bernilai true.

FP = Proporsi sampel bernilai false yang diprediksi secara benar.

TN = Proporsi sampel bernilai true yang diprediksi sebagai sampel bernilai false.

Untuk klasifikasi data mining, nilai AUC dapat dibagi menjadi beberapa kelompok (Sulaehani, 2016):

0.90 – 1.00 = *Excellent classification*

0.80 – 0.90 = *Good classification*

0.70 – 0.80 = *Fair classification*
 0.60 – 0.70 = *Poor classification*
 0.50 – 0.60 = *Failure*

2.4. Perumusan Hipotesis

Hipotesis pada penelitian ini adalah sebagai berikut:

H0 : Seleksi fitur Backward Elimination tidak dapat meningkatkan nilai akurasi algoritma K-Nearest Neighbor untuk klasifikasi penyakit mesothelioma

H1 : Seleksi fitur Backward Elimination dapat meningkatkan nilai akurasi algoritma K-Nearest Neighbor untuk klasifikasi penyakit mesothelioma

3. Metode Penelitian

Mesothelioma's Disease Dataset adalah objek hasil penelitian dari Orhan Er, Abdullah Cetin Tanrikulu, Abdurrahman Abakay dan Feyzullah Temurtas. Dataset memiliki 34 atribut dan 1 label (atribut target) yang merupakan data pemeriksaan penyakit Mesothelioma. Pengumpulan data ini melibatkan 324 pasien di rumah sakit Dicle University, Fakultas Kedokteran. *Dataset* diambil langsung melalui halaman *UCI Machine Learning Repository* dengan nama *dataset Mesothelioma's Disease*.

Mesothelioma's Disease Dataset merupakan data sekunder yang sudah siap untuk proses *data mining*. Metode yang akan digunakan pada penelitian ini adalah metode klasifikasi K-Nearest Neighbor dan seleksi fitur *Backward Elimination*.

Tahap *preprocessing* pada *dataset* dengan menerapkan normalisasi yang bertujuan untuk membuat setiap atribut memiliki data dalam skala yang sama. Setelah itu data dibagi menjadi dua bagian yaitu *data training* sebanyak 80% dan *data testing* sebanyak 20%. Untuk menghilangkan atribut yang tidak relevan dan meningkatkan nilai akurasi yang dihasilkan diterapkan seleksi fitur *Backward Elimination*.

Pembagian data training dan data testing ditentukan dari jenis klasifikasi. Terdapat dua jenis klasifikasi yang ditentukan, yaitu *health* dan *mesothelioma*. Tabel 1 merupakan jumlah data yang digunakan untuk data training dan data testing.

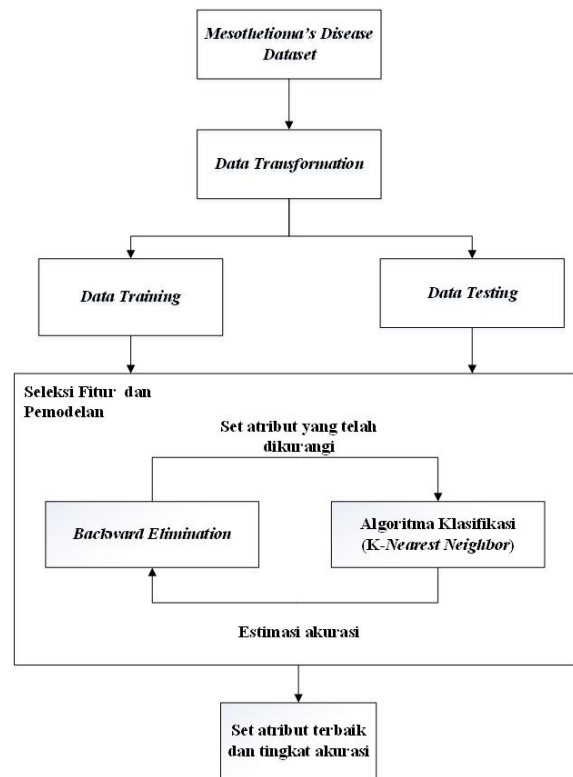
Tabel 1. Data Training dan Data Testing

No	Jenis Klasifikasi	Jumlah Record	Data Training	Data Testing
1	Health	228	182	46
2	Mesothelioma	96	77	19
Jumlah		324	259	65

Penggunaan sampel untuk pengujian terhadap model yang dihasilkan atau dapat dikatakan sebagai

data testing, sedangkan *data training* berfungsi sebagai bahan pelatihan suatu model. *Mesothelioma's Disease Dataset* memiliki total *record* sebanyak 324, kemudian telah dibagi menjadi dua bagian, yaitu 80% untuk *data training* dan 20% untuk *data testing*, sehingga jumlah *training data* sebanyak 259 *record* dan *testing data* sebanyak 65 *record*.

Berikut adalah metodologi penelitian yang dilakukan seperti diperlihatkan pada gambar 1.



Gambar 1. Metodologi Penelitian

4. Analisa Dan Perancangan

Algoritma K-Nearest Neighbor dan seleksi fitur backward elimination pada objek penyakit mesothelioma digunakan untuk meningkatkan hasil tingkat akurasi pada klasifikasi penyakit tersebut.

Peningkatan hasil akurasi klasifikasi tersebut akan diperoleh atau tidak diperoleh, hasilnya akan terlihat pada hasil penelitian. Penelitian dilakukan dalam dua tahap. Tahap pertama adalah implementasi penggunaan algoritma K-Nearest Neighbor tanpa penerapan seleksi fitur Backward Elimination. Tahap kedua adalah implementasi penggunaan algoritma k-nearest neighbor dengan penerapan seleksi fitur Backward Elimination.

Untuk menghilangkan atribut yang tidak relevan dan meningkatkan hasil akurasi dalam proses klasifikasi penyakit mesothelioma, digunakan penerapan seleksi fitur. Pada penelitian ini akan diterapkan seleksi fitur *Backward Elimination* pada

model yang diuji. Seleksi fitur dilakukan bersamaan dengan pelaksanaan pemodelan dengan cara pemilihan selanjutnya dan menguji kembali semua fitur. Fitur terpilih yang berpengaruh akan dikeluarkan dari model satu per satu. Jika fitur yang terpilih dianggap tidak berpengaruh atau tidak signifikan pada model, maka fitur tidak akan dikeluarkan dari pemodelan.

Pemodelan data menggunakan *Software Rapidminer Studio*. Tahapan pemodelan adalah sebagai berikut:

- a. Pilih data Access-Files-Read-Read Excel
- b. Pilih Import Configuration Wizard
- c. Input dataset Mesothelioma's Disease
- d. Input data training dan data testing
- e. Penggabungan seleksi fitur menggunakan Backward Elimination, dan menentukan jumlah max atribut yang dikeluarkan

4.1. Implementasi Algoritma K-Nearest Neighbor

Algoritma k-nearest neighbor menggunakan 34 atribut dan beberapa nilai *k* yang berbeda. Tabel 2 adalah hasil akurasi klasifikasi algoritma k-nearest neighbor.

Tabel 2. Nilai Akurasi pada Jumlah K yang berbeda

Jumlah <i>k</i>	Akurasi
1	92.31%
3	92.31%
5	93.85%
7	90.77%

Hasil percobaan dengan jumlah *k* yang berbeda, terlihat bahwa nilai tertinggi akurasi adalah *k*=5 sebesar 93.85%.

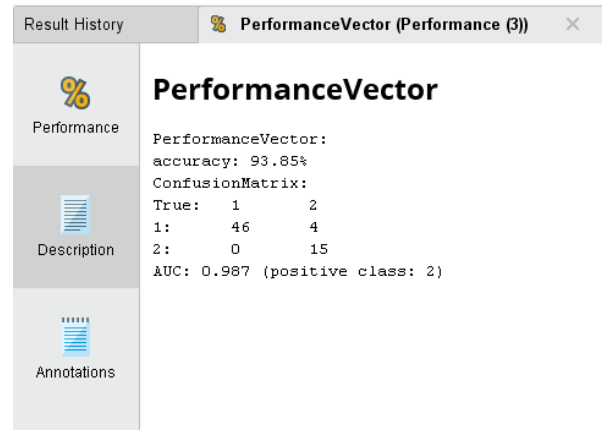
Prediksi 1 untuk *health* dan prediksi 2 untuk mesothelioma, merupakan prediksi yang dikeluarkan dari perhitungan K-Nearest Neighbor untuk memprediksi data testing. Sedangkan untuk True 1 (*Health*) dan True 2 (*Mesothelioma*) adalah nilai klasifikasi yang sebenarnya dari data testing. Tabel 3 memperlihatkan confusion matrix algoritma K-Nearest Neighbor untuk prediksi 1 dan prediksi 2.

Tabel 3. Confusion Matrix Algoritma K-Nearest Neighbor

	True 1 (<i>Health</i>)	True 1 (<i>Mesothelioma</i>)	Class Precision
Prediksi 1 (<i>Health</i>)	46	4	92%
Prediksi 2 (<i>Mesothelioma</i>)	0	15	100%

Class Recall	100%	78.95%	
--------------	------	--------	--

Hasil evaluasi yang diperoleh menggunakan *Software Rapidminer* akan terlihat seperti pada gambar 2 berikut.



Gambar 2. Tampilan Performance Vector Menggunakan Rapidminer

Hasil dari *Software Rapidminer Studio* yang disajikan pada tabel 3, menghasilkan proposal sampel bernilai *true* yang diprediksi secara benar adalah 46. Proporsi sampel bernilai *false* yang diprediksi sebagai sampel bernilai *true* adalah 4. Proposal sampel bernilai *false* yang diprediksi secara benar adalah 0. Proporsi sampel bernilai *true* yang diprediksi sebagai sampel bernilai *false* adalah 15.

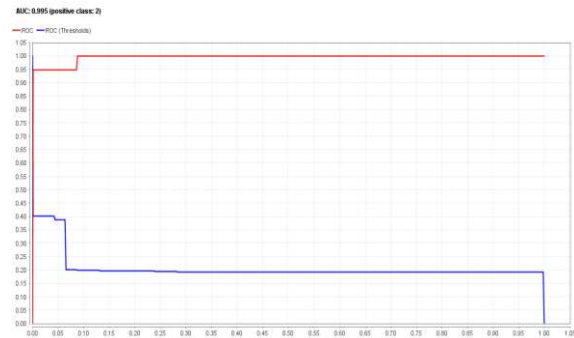
Hasil perhitungan persentase nilai akurasi adalah sebagai berikut:

$$A = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\%$$

$$A = \frac{46 + 15}{46 + 15 + 0 + 4} \cdot 100\% = 93.85\%$$

Nilai akurasi *confusion matrix* hasil klasifikasi K-Nearest Neighbor menggunakan seleksi fitur Backward Elimination diperoleh 93.85%.

Berikut adalah Grafik ROC pada implementasi algoritma K-Nearest Neighbor (gambar 3).



Gambar 3. Grafik ROC implementasi algoritma k-nearest neighbor

Pengujian tersebut menghasilkan kurva ROC (*Receiver Operating Characteristic*) yang dapat digunakan untuk menganalisis seberapa baik classifier dapat mengenali data dari kelas yang berbeda, seperti yang terlihat pada gambar 3 menghasilkan nilai AUC (*Area Under the ROC Curve*) sebesar 0.995 (kategori *Excellent Classification*).

4.2. Implementasi K-Nearest Neighbor Dan Backward Elimination

Seleksi fitur digunakan untuk menghilangkan atribut-atribut yang tidak relevan pada dataset, dan seleksi fitur juga digunakan untuk meningkatkan akurasi dari klasifikasi penyakit mesothelioma. Pada tabel 3 menampilkan hasil penelitian terhadap dataset yang telah melalui proses seleksi fitur menggunakan backward elimination. Jumlah atribut sebelum dilakukan seleksi fitur sebanyak 34 atribut, sedangkan setelah dilakukan seleksi fitur jumlah atribut menjadi 31 atribut.

Seleksi fitur dilakukan dengan cara mengeluarkan satu per satu atribut bersamaan dengan melakukan pemodelan menggunakan algoritma *K-Nearest Neighbor*. Setelah itu *Backward Elimination* akan mengeluarkan atribut yang tidak berpengaruh atau jika suatu atribut dikeluarkan dari model maka hasil akurasi dari algoritma *K-Nearest Neighbor* akan menunjukkan peningkatan dari hasil akurasi sebelum menerapkan metode seleksi fitur. Pada setiap iterasi, *Backward Elimination* akan mengeluarkan satu atribut yang tidak relevan dan menghasilkan subset atribut baru. Atribut-atribut yang dikeluarkan oleh *Backward Elimination* antara lain; *City*, *Ache on Chest* dan *Habit of Cigarette*.

Backward Elimination dilakukan dengan cara menguji semua variabel kemudian menghapus variabel-variabel yang dianggap tidak berpengaruh terhadap proses klasifikasi data.

Pada iterasi pertama, atribut yang dikeluarkan dari pemodelan adalah *City* dengan persentase akurasi untuk dikeluarkan sebesar 96.92%. Pada iterasi kedua, atribut yang dikeluarkan dari pemodelan adalah *Ache*

on Chest dengan persentase akurasi untuk dikeluarkan sebesar 98.46%. Pada iterasi ketiga, atribut yang dikeluarkan dari pemodelan adalah *Habit of Cigarette* dengan persentase akurasi untuk dikeluarkan sebesar 98.46%.

Berikut tabel 4 untuk prediksi 1 dan 2 algoritma k-nearest neighbor dan seleksi fitur backward elimination.

Tabel 4. Confussion Matrix Algoritma K-Nearest Neighbor dan Seleksi Fitur Backward Elimination

	True 1 (Health)	True 1 (Mesothelioma)	Class Precision
Prediksi 1 (Health)	46	1	97.8%
Prediksi 2 (Mesothelioma)	0	18	100%
Class Recall	100%	84.74%	

Hasil dari *Software Rapidminer Studio* yang disajikan pada tabel 4, menghasilkan proposal sampel bernilai *true* yang diprediksi secara benar adalah 46. Proporsi sampel bernilai *false* yang diprediksi sebagai sampel bernilai *true* adalah 1. Proposal sampel bernilai *false* yang diprediksi secara benar adalah 0. Proporsi sampel bernilai *true* yang diprediksi sebagai sampel bernilai *false* adalah 18.

Hasil perhitungan persentase nilai akurasi adalah sebagai berikut:

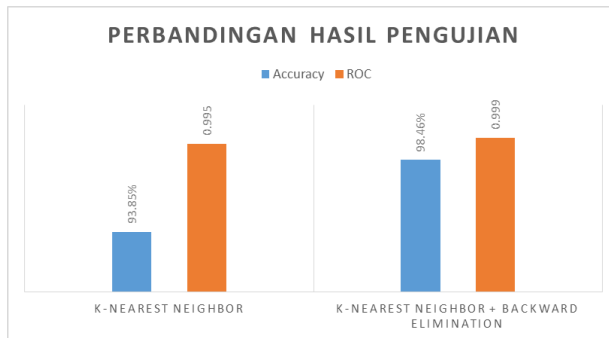
$$A = \frac{TP + TN}{TP + TN + FP + FN} \cdot 100\%$$

$$A = \frac{46 + 18}{46 + 18 + 0 + 1} \cdot 100\% = 98.46\%$$

Nilai akurasi *confussion matrix* hasil klasifikasi *K-Nearest Neighbor* menggunakan seleksi fitur *Backward Elimination* diperoleh 98.46%.

Pengujian kurva ROC (*Receiver Operating Characteristic*) menghasilkan nilai AUC sebesar 0.995 (kategori *Excellent Classification*).

Tingkat akurasi dari hasil pengujian seleksi fitur menggunakan *Backward Elimination* dengan cara menghitung jumlah data testing yang dapat terklasifikasi secara benar atau tingkat akurasi dan melalui grafik ROC (*Receiver Operating Characteristic*). Gambar 4 adalah perbandingan nilai akurasi dan AUC (*Area Under the ROC Curve*) pada *K-Nearest Neighbor* sebelum dan sesudah menggunakan seleksi fitur *Backward Elimination*.



Gambar 4. Perbandingan Hasil Pengujian

Perbandingan hasil pengujian pada gambar 4 menunjukkan bahwa algoritma K-Nearest Neighbor mengalami peningkatan sebesar 4,61% setelah menggunakan seleksi fitur Backward Elimination. Implementasi seleksi fitur Backward Elimination dapat meningkatkan akurasi klasifikasi pada penyakit Mesothelioma. Selain itu *atribut City, Ache on Chest dan Habit of Cigaretete* merupakan atribut yang tidak relevan dalam proses klasifikasi penyakit Mesothelioma.

5. Simpulan

Simpulan dari penelitian ini adalah tingkat akurasi yang dihasilkan oleh algoritma K-Nearest Neighbor dalam klasifikasi penyakit mesothelioma sebesar 93.85% dan nilai AUC sebesar 0.995 (termasuk dalam kategori *Excellent Classification*). Sedangkan untuk hasil klasifikasi algoritma K-Nearest Neighbor setelah dilakukan seleksi fitur menggunakan Backward Elimination menunjukkan peningkatan akurasi sebesar 4,61%, sehingga tingkat akurasi yang dihasilkan sebesar 98.46% dan nilai AUC sebesar 0.999 (termasuk dalam kategori *Excellent Classification*).

6. Referensi

- Adinugroho, S., & Sari, Y. A. (2018). *Implementasi Data Mining Menggunakan Weka*. Malang: UB Press.
- Bahri, R. S., & Maliki, I. (2012). Feature Extraction Pada Optical Character Recognition. *Jurnal Komputer dan Informatika (Komputa)*, 29-35.
- Bode, A. (2017). K-Nearest Neighbor Dengan Feature Selection Menggunakan Backward Elimination Untuk Prediksi Harga Komoditi Kopi Arabika. *ILKOM*, 188-195.
- Ditjen Dikti Kemdikbud. (2012, February 2). *Surat Dirjen Dikti No. 152/E/T/2012 : Wajib Publikasi Ilmiah Bagi S1/S2/S3*. Retrieved from Kementrian Riset, Teknologi dan Pendidikan Tinggi Kopertis Wilayah XII Maluku Utara: <http://www.kopertis12.or.id/2012/02/01/surat-dirjen-dikti-no-152et2012-tentang-wajib-publikasi-ilmiah-bagi-s1s2s3.html>
- Er., O. C., & Temurtas, F. (2012). An Approach Based On Probabilistic Neural Network For Diagnosis Of

Mesothelioma's Disease. *Computer and Electrical Engineering*, 75-81.

- Faisal, M. R. (2017). *Seri Belajar Data Science Klasifikasi dengan Bahasa Pemrograman R*. Indonesia: Net Developer Community.
- Han, J., & Kamber, M. (2006). *Data Mining: Concept and Techniques*. New York: Morgan Kaufmann Publisher.
- Huda, N. M. (2010). *Aplikasi Data Mining Untuk Menampilkan Informasi Tingkat Kelulusan Mahasiswa*. Semarang: Universitas Diponegoro.
- Jumadi, & Winarko, E. (2015). Penggunaan Knn (K-Nearest Neighbor) Untuk Klasifikasi Teks Berita yang Tak-Terkelompokkan Pada Saat Pengklasteran Oleh Stc (Suffix Tree Clustering). *Jurnal ISTEK, IX(1)*, 50-81.
- Khan, S. N., Sikandar, G., Anwar, S., & Khan, M. T. (2018). Classification of Malignant Mesothelioma Cancer Using Support Vector Machine. *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)* (pp. 12-16). Sukkur Sindh, Pakistan: Sukkur IBA University.
- Khan, S. N., Sikander, G., Anwar, S., & Khan, M. T. (2018). Classification Of Mesothelioma Cancer Using Support Vectore Machine. 12-16.
- Marianti. (2018, November 6). *alodokter*. Retrieved Februari 21, 2019, from Mesothelioma: <https://www.alodokter.com/mesothelioma>
- Molinari, L. (2019). *Mesothelioma.com*. Retrieved Februari 21, 2019, from Pleural Mesothelioma: <https://www.mesothelioma.com/mesothelioma/types/pleural.htm>
- Mustakim, & Oktaviani F, G. (2016). Algoritma K-Nearest Neighbor Classification Sebagai Sistem Prediksi Predikat Prestasi Mahasiswa. *Jurnal Sains, Teknologi dan Industri, Vol. 13, No.2, ISSN 1693-2390 print/ISSN 2407-0939 online*, pp.195 -202.
- Nilashi, M., Roudbaraki, M. Z., Farahmand, M., Branch, L., & Branch, A. (2017). A Predictive Method For Mesothelioma Disease Classification Using Naive Bayes Classifier. *Journal of Soft Computing*, 8-14.
- Nofriansyah, D. (2014). *Konsep Data Mining Vs Sistem Pendukung Keputusan*. Yogyakarta: Deepublish.
- Prasetyo, E. (2014). *Data Mining - Mengolah Data Menjadi Informasi Menggunakan Matlab*. Yogyakarta: Penerbit ANDI.
- Sonatha, Y. (2013). Churn Prediction Pelanggan Menggunakan Crisp-DM. *Jurnal Ilmiah Elektron*, 53-62.
- Sulaehani, R. (2016). Prediksi Keputusan Klien Telemarketing untuk Deposito Pada Bank Menggunakan Algoritma Naive Bayes Berbasis Backward Elimination. *Jurnal Ilmiah ILKOM*, 182-189.
- Tutuncu, K., & Catalas, O. (2017). Intelligent Systems and Applications in Engineering Diagnosis of Mesothelioma Disease Using Different Classification Techniques. *Intelligent Systems and Applications in Engineering*, 7-11.
- Widodo, P., Handayanto, R. T., & Herlawati. (2013). *Penerapan Data Mining Dengan Matlab*. Bandung: Penerbit Rekayasa Sains.
- Yunita. (2017). Seleksi Fitur Menggunakan Backward Elimination Pada Prediksi Cuaca Dengan Neural

Network. *Indonesia Journal on Computer and Information Tchnology (IJCIT)*, 26-37.

Zervos, M. D., Bizekis, C., & Pass, H. I. (2008). Malignant Mesothelioma 2008. *Current Opinion in Pulmonary Medicine*, 303-309.