

Klasifikasi Penurunan Fungsi Kognitif Pasien Stroke Menggunakan Metode Klasifikasi *Random Forest*

Muhammad Shidqi Fadlilah¹, Randy Cahya Wihandika², Bayu Rahayudi³

Program Studi Teknik Informatika, Fakultas Ilmu Komputer, Universitas Brawijaya
Email: ¹mshidqifadlilah@gmail.com, ²rendicahya@ub.ac.id, ³ubay1@ub.ac.id

Abstrak

Stroke merupakan salah satu penyakit yang menyerang segala golongan manusia tanpa memandang ras, jenis kelamin, dan usia. Salah satu dampak dari penyakit stroke adalah penurunan fungsi kognitif. Otak manusia memiliki banyak saraf yang beberapa diantaranya adalah mengatur kerja fungsi kognitif manusia. Berdasarkan penelitian yang dilakukan oleh Wardani (2015), faktor yang memengaruhi penurunan fungsi kognitif terdiri dari tiga belas faktor. Dengan jumlah yang tidak sedikit tersebut, maka dibutuhkan proses klasifikasi yang dapat mendeteksi penurunan fungsi kognitif pada pasien stroke. Maka dalam penelitian ini, dibuat sistem yang dapat mengklasifikasi penurunan fungsi kognitif pada pasien stroke dengan menggunakan metode klasifikasi *random forest* berdasarkan data yang didapatkan dari Rumah Sakit Saiful Anwar pada bulan Juli sampai dengan Oktober tahun 2014. Metode *random forest* dipilih karena metode ini baik untuk data kategorikal seperti pada kasus faktor-faktor pengaruh penurunan fungsi kognitif pada pasien stroke ini. Berdasarkan hasil pengujian dengan lima kombinasi data yang terdiri dari 19 data uji dan 76 data latih, jumlah *tree* terbaik yang didapatkan adalah 100 *tree*. Hasil rata-rata akurasi yang didapatkan dari semua percobaan adalah 53,094%. Hal ini disebabkan oleh beberapa faktor yang menjadi *error* pada proses klasifikasi tersebut. Salah satu faktor yang menyebabkan ketidaksempurnaan klasifikasi ini adalah penyebaran data latih yang tidak merata kelasnya.

Kata kunci: *random forest, Decision Tree, Fungsi Kognitif, Stroke*

Abstract

Stroke is a disease that attacks all human, regardless of race, gender, and age. One of the effects of stroke is a decreased cognitive function. A human brain has many nerves, one of them is regulating the work of the human's cognitive function. Based on research by Wardhani (2015), factors that decreasing cognitive function consist of thirteen factors. So, a system that can detect a decreased cognitive function on a stroke patient is needed. So in this research, we make a system that could be used to classify the decreasing cognitive function using a random forest method. the random forest was chosen because this method is good for categorical data. Based on the testing result, the best tree that builds in this system was 100 trees. The average result of the accuracy obtained from all experiments were 53.094%. That number means that the system is still far from perfect. One of the factors that caused this system's imperfection was the distribution of training classes were not evenly distributed.

Keywords: *Random Forest, Decision Tree, Cognitive Function, Stroke*

1. LATAR BELAKANG

Penyakit stroke adalah penyakit yang penderitanya mengalami gangguan vaskular. Hal tersebut disebabkan oleh pecahnya pembuluh darah di otak yang menimbulkan gangguan kognitif berdasarkan lokasi pecahnya pembuluh darah di otak penderita tersebut. Gangguan kognitif terdiri dari penurunan kesadaran,

gangguan visuospatial, gangguan pembelajaran non-verbal, gangguan berkomunikasi, dan penurunan tingkat atensi penderita (Hanas, et al., 2015).

Dari beberapa penyakit di dunia yang dapat menyerang manusia dan menyebabkan kematian, stroke merupakan salah satu penyakit yang menyerang segala golongan manusia tanpa memandang ras, jenis kelamin, dan usia.

Menurut organisasi kesehatan dunia, World Health Organization (WHO), stroke atau Cerebrovascular disease ini adalah tanda klinis yang berkembang akibat gangguan fungsi otak karena adanya penyumbatan atau terjadi pecah pembuluh darah pada otak penderita dengan gejala yang berlangsung lebih dari 24 jam (Nugraha, et al., 2017).

WHO menyebutkan ada sekitar 795.000 orang di Amerika Serikat terserang penyakit stroke setiap tahunnya. Dari 795.000 orang tersebut, 610.000 di antaranya merupakan stroke perdananya atau bukan status kekambuhan stroke. Sedangkan 185.000 orang lainnya merupakan kasus kekambuhan stroke yang berulang. Dari empat juta orang di Amerika Serikat yang mengalami stroke, 15-30% pasien pasca stroke menderita cacat yang permanen. Di Indonesia, data Riset Kesehatan Dasar tahun 2013 menyebutkan bahwa tujuh dari seribu orang terkena stroke. Hal tersebut dapat dikatakan dari setiap tujuh orang meninggal, satu di antaranya meninggal dikarenakan stroke (Malik & Maulina, 2015).

Salah satu dampak dari penyakit stroke adalah penurunan fungsi kognitif. Otak manusia memiliki banyak saraf yang beberapa diantaranya adalah mengatur kerja fungsi kognitif manusia. Penurunan fungsi kognisi manusia dimulai pada usia 30 tahun dan berlangsung secara bertahap atau progresif. Penyebab penurunan fungsi kognitif ini bisa karena vaskular, hal ini dapat diartikan sebagai gangguan yang berlanjut dari mild cognitive impairment sampai diemisia. Hal ini dikarenakan adanya keterlibatan kerusakan vaskular yang berguna untuk memori, kognisi, dan kebiasaan. Kelainan kognitif akibat vaskular dapat dideskripsikan sebagai kelainan heterogen dengan berbagai tipe lesi cerebrovascular yang berbeda dan akan berkembang sampai dengan tahap akhirnya yaitu demensia (Jellinger, 2013).

Pendeteksian penurunan fungsi kognitif pada pasien stroke ini perlu dilakukan dengan menggunakan komputer. Hal ini ditujukan untuk mempersingkat waktu dalam proses pendeteksian. Seperti penelitian yang dilakukan oleh Nugraha (2017), menggunakan 203 data rekam medik pasien yang menderita penyakit stroke yang ada di Rumah Sakit Umum Daerah Udata Palu. Penelitian tersebut mendapatkan tingkat keberhasilan dari klasifikasi yang dilakukan tersebut sebesar 89,65% tingkat akurasi. Berdasarkan penelitian yang dilakukan oleh Wardhani (2015), faktor yang

memengaruhi penurunan fungsi kognitif terdiri dari 13 faktor. Dengan jumlah yang tidak sedikit tersebut, maka dibutuhkan sistem yang dapat mendeteksi penurunan fungsi kognitif pada pasien stroke. Meskipun sebuah sistem tidak akan mendapatkan hasil yang sempurna dalam tingkat akurasi, pengguna sistem tersebut dalam hal ini adalah penderita stroke tetap dapat menggunakannya untuk mengetahui deteksi awal sebelum memeriksakannya ke dokter lebih lanjut untuk hasil yang lebih maksimal.

Dari penjabaran tentang stroke dan faktor penurunan fungsi kognitif di atas, maka dalam penelitian kali ini dilakukan pembuatan sistem yang dapat digunakan untuk mengklasifikasi penurunan fungsi kognitif pada pasien stroke dengan menggunakan metode klasifikasi random forest berdasarkan data yang didapatkan. Metode random forest dipilih karena metode ini baik untuk data kategorikal seperti pada kasus faktor-faktor pengaruh penurunan fungsi kognitif pada pasien stroke ini. Metode ini menggunakan voting dari sejumlah decision tree yang dibangun dari data latih yang ada. Penelitian mengenai klasifikasi menggunakan random forest terlebih dahulu dilakukan oleh Fitriani (2007) untuk mengklasifikasi penerimaan tenaga kerja baru dan oleh Rukmigayatri (2015) untuk klasifikasi kemunculan titik panas pada lahan gambut.

Metode *random forest* merupakan sebuah perkembangan lebih lanjut dari metode *Classification and Regression Trees (CART)*. Metode ini dapat melakukan klasifikasi dan regresi. Yang membedakan dari CART adalah jumlah *tree* yang digunakan dalam metode ini lebih dari satu. Hasil klasifikasi atau regresi dari setiap *tree* akan dilakukan voting mencari kelas terbanyak yang dihasilkan. Kata random pada nama metode ini diartikan sebagai data latih yang digunakan dalam pembangunan *tree* dalam metode ini adalah hasil pengacakan secara tidak beraturan (Binarwati, et al., 2017). Metode melakukan tiga proses utama yaitu pengacakan data latih, pembangunan pohon keputusan, dan *voting*.

Pengacakan data dilakukan untuk mendapatkan data pada data latih secara acak untuk membedakan antara satu *tree* dengan *tree* yang lainnya. Tidak jarang terdapat redundansi data pada hasil pengacakan data latih. Data tersebut digunakan untuk menentukan kelas menggunakan pohon CART. Sehingga hasil klasifikasi dari setiap *tree* yang terbangun memiliki kelas yang berbeda.

Langkah selanjutnya setelah proses pengacakan data berdasarkan data latih adalah menumbuhkan pohon CART dari data random yang dihasilkan. Pemilihan fitur pemisah dalam CART adalah menggunakan nilai *gain* yang dihasilkan. Untuk mendapatkan nilai *gain* pada setiap fitur pada data latih, terlebih dahulu dilakukan perhitungan entropi kelas dan entropi fitur tersebut seperti pada persamaan 1 dan 2.

$$Ent(S) = \sum_{j=1}^k -p_j \log_2 p_j \quad (1)$$

Keterangan:

Ent: Entropi

S : Himpunan

K : Jumlah Kelas

Pj: Probabilitas Kelas dalam Himpunan

$$Gain = Ent(S) - \sum_{j=1}^k Ent(A) \quad (2)$$

Keterangan:

Ent: Entropi

S : Himpunan Kelas

A : Himpunan Fitur

Tahap evaluasi dilakukan setelah semua proses pada proses klasifikasi ini telah selesai. Evaluasi bertujuan untuk mengetahui seberapa baik klasifikasi yang telah dilakukan tersebut. Pada proses evaluasi pada Klasifikasi Penurunan Fungsi Kognitif Pasien Stroke Menggunakan Metode Klasifikasi *random forest* dilakukan dengan beberapa tahap yaitu evaluasi parameter yang optimal dan evaluasi akhir dengan parameter optimal yang telah didapatkan sebelumnya. Proses tersebut akan dijelaskan lebih lanjut pada perancangan pengujian pada Bab Perancangan.

Kedua proses tersebut dievaluasi dengan mencari nilai akurasi dari setiap pengujianinya. Perhitungan akurasi dilakukan dengan rumus seperti pada persamaan 3.

$$Akurasi = \frac{B}{N} \times 100\% \quad (3)$$

Keterangan:

B: Jumlah Klasifikasi Sesuai

N: Jumlah Data Uji

2. METODOLOGI PENELITIAN

2.1 Pengambilan Data

Tahapan pengambilan data ini dilakukan menggunakan data yang diambil dari data

penelitian Fakultas Kedokteran Universitas Brawijaya pada Rumah Sakit Saiful Anwar Malang pada bulan Juli sampai dengan Oktober pada tahun 2014 yang di dalamnya terdapat data pasien stroke thrombosis.

Data yang didapatkan adalah data klasifikasi yang kelas akhirnya adalah Normal, Kemungkinan Gangguan, dan Gangguan. Serta fitur yang digunakan untuk mendapatkan kelas tersebut berjumlah 15 fitur yang diantaranya adalah usia, jenis kelamin, pendidikan, riwayat stroke, nilai BMI (Body Mass Index), riwayat hipertensi, diabetes melitus, riwayat hiperkolestrol, riwayat merokok, letak lesi, atrial fibrilasi, dan gagal ginjal.

Data data yang didapatkan dalam penelitian sebelumnya tersebut berjumlah 95 data. Dari 95 data tersebut, 76 akan digunakan sebagai data latih, 19 sisanya akan digunakan sebagai data uji. Penyebaran dari 95 data latih yang dimiliki adalah 32 normal, 31 kemungkinan, dan 7 gangguan.

2.2 Implementasi

Implementasi adalah penerapan klasifikasi sesuai dengan perancangan yang telah dibuat terlebih dahulu yaitu manualisasi data latih yang didapat. Implementasi pada penelitian kali ini dilakukan dengan menggunakan bahasa pemrograman Python. Pemilihan bahasa pemrograman Python sendiri dikarenakan teknik penulisan bahasa pemrograman Python tidak sulit tanpa harus menuliskan tipe data sebelumnya dan variabelnya yang dinamis.

Pada tahap implementasi dilakukan klasifikasi dari data latih yang didapatkan menggunakan metode *random forest*. Metode *random forest* dipilih berdasarkan hasil dari penelitian yang dilakukan Rukmigayatri (2015), metode ini menghasilkan tingkat akurasi sebesar 87,40%.

Hasil yang disajikan dalam sistem yang dibangun penulis dalam penelitian kali ini adalah tanpa user interface. Proses input untuk melakukan pengujian di langkah berikutnya. Proses input yang dilakukan adalah angka untuk menentukan jumlah pembangunan *decision tree* yang dilakukan dalam metode *random forest*.

2.3 Evaluasi

Evaluasi yang dilakukan pada penelitian kali ini adalah pengujian jumlah *tree* yang dibangun pada metode *random forest*. Pengujian dilakukan dengan metode *K-fold cross*

validation. Metode pengujian ini adalah salah satu metode yang dilakukan untuk mengevaluasi kinerja dari kombinasi data latih dan data uji. Seluruh dataset yang dimiliki dibagi menjadi lima *fold*, yaitu F1, F2, F3, F4, dan F5. Setiap *fold* tersebut berisikan 19 data. Dari 5 subset data tersebut akan dibuat menjadi lima kombinasi yang terdiri dari empat *fold* data latih dan satu *fold* data uji.

3. HASIL DAN ANALISIS

3.1 Pengujian *K-fold cross validation* Jumlah Fitur

Pengujian seleksi fitur adalah pengujian untuk menentukan jumlah fitur yang digunakan dalam proses klasifikasi. Seleksi fitur dilakukan dengan menggunakan metode *information gain* dengan menggunakan rumus yang sama dengan proses pemisahan pada pembangunan *tree*. Rumus yang dimaksud telah dijabarkan pada Persamaan 1 dan Persamaan 2.

Pengujian seleksi fitur ini dilakukan dengan metode *k-fold cross validation*. Metode pengujian ini adalah salah satu metode yang dilakukan untuk mengevaluasi kinerja dari kombinasi data latih dan data uji. Seluruh dataset yang dimiliki dibagi menjadi lima *fold*, yaitu F1, F2, F3, F4, dan F5. Setiap *fold* tersebut berisikan 19 data. Dari 5 subset data tersebut dibuat menjadi lima kombinasi yang terdiri dari empat *folds* data latih dan satu *fold* data uji.

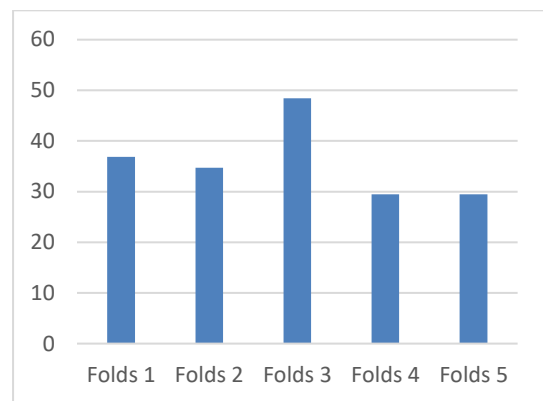
Masing-masing dari kombinasi yang terbentuk, dilakukan pengujian jumlah fitur dari mulai 8 fitur sampai dengan 13. Pengujian yang telah dirancang tersebut dilakukan sebanyak lima kali dikarenakan data yang dilatih terlebih dahulu diacak secara random pada proses *training*. Dari pengujian tersebut didapatkan jumlah fitur berdasarkan hasil nilai akurasi yang didapatkan. Nilai akurasi tersebut adalah rata-rata dari setiap eksekusi pengujian.

Jika pada setiap percobaan tersebut menghasilkan nilai akurasi nol atau *error*, maka dilakukan kembali hingga maksimal 10 kali. Jika setelah 10 kali dijalankan tetap menghasilkan nilai nol, maka nilai akurasi pada percobaan tersebut adalah nol. Jumlah *tree* yang digunakan pada pengujian seleksi fitur menggunakan *k-fold cross validation* ini adalah 50 *tree*.

3.1.1 Hasil Pengujian 13 Fitur

Berikut adalah hasil pengujian seleksi fitur dengan menggunakan jumlah fitur adalah 13

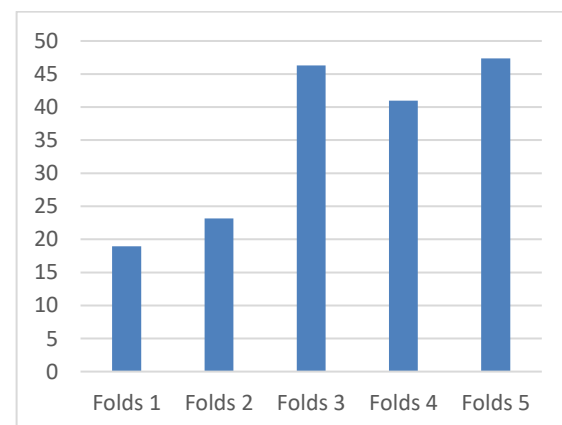
fitur yang digambarkan pada Gambar 1. Berdasarkan hasil pengujian dengan menggunakan 13 fitur pada lima kombinasi *folds*, didapatkan rata-rata akurasi adalah 35,78%. Kombinasi terbaik terdapat pada kombinasi 3 dengan F3 sebagai data uji dan *fold* lainnya sebagai data latih. Pada gambar, sumbu x adalah *folds* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



Gambar 1. Hasil Pengujian 13 Fitur

3.1.2 Hasil Pengujian 12 Fitur

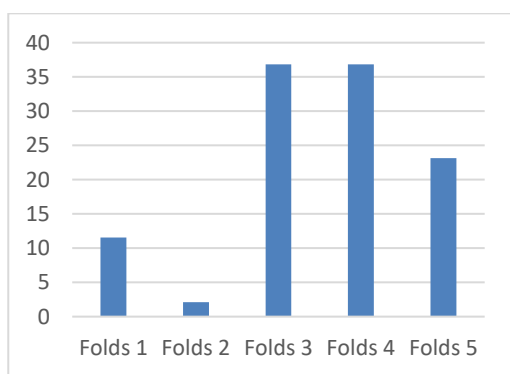
Berikut adalah hasil pengujian seleksi fitur dengan menggunakan jumlah fitur adalah 12 fitur yang digambarkan pada Gambar 2. Berdasarkan hasil pengujian dengan menggunakan 12 fitur pada lima kombinasi *folds*, didapatkan rata-rata akurasi adalah 35,36%. Kombinasi terbaik terdapat pada kombinasi 5 dengan F5 sebagai data uji dan *fold* lainnya sebagai data latih. Pada gambar, sumbu x adalah *folds* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



Gambar 1. Hasil Pengujian 12 Fitur

3.1.3 Hasil Pengujian 11 Fitur

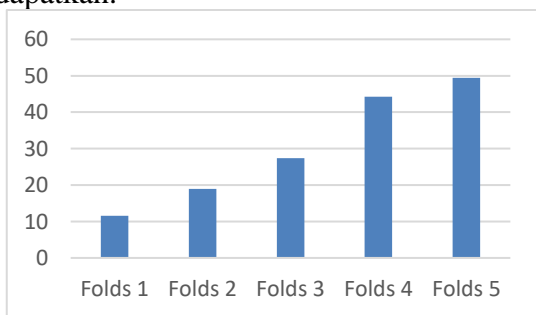
Berikut adalah hasil pengujian seleksi fitur dengan menggunakan jumlah fitur adalah 11 fitur yang digambarkan pada Gambar 3. Berdasarkan hasil pengujian dengan menggunakan 11 fitur pada lima kombinasi *folds*, didapatkan rata-rata akurasi adalah 22,1%. Kombinasi terbaik terdapat pada kombinasi 3 dengan F3 sebagai data uji dan *fold* lainnya sebagai data latih. Pada gambar, sumbu x adalah *folds* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



Gambar 1. Hasil Pengujian 11 Fitur

3.1.4 Hasil Pengujian 10 Fitur

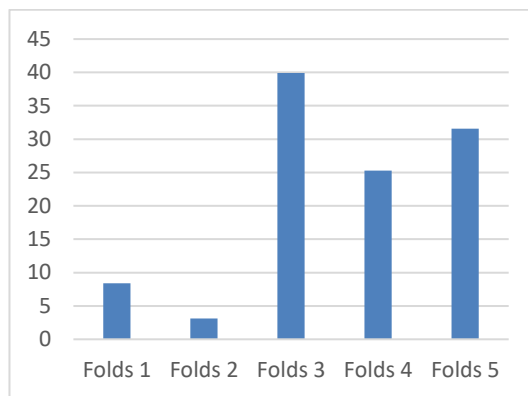
Berikut adalah hasil pengujian seleksi fitur dengan menggunakan jumlah fitur adalah 10 fitur yang digambarkan pada Gambar 4. Berdasarkan hasil pengujian dengan menggunakan 10 fitur pada lima kombinasi *folds*, didapatkan rata-rata akurasi adalah 30,31%. Kombinasi terbaik terdapat pada kombinasi 5 dengan F5 sebagai data uji dan *fold* lainnya sebagai data latih. Pada gambar, sumbu x adalah *folds* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



Gambar 1. Hasil Pengujian 10 Fitur

3.1.5 Hasil Pengujian 9 Fitur

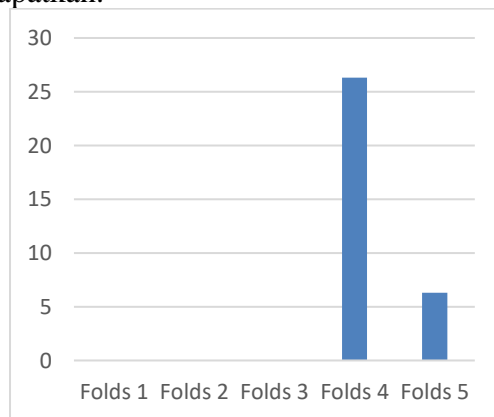
Berikut adalah hasil pengujian seleksi fitur dengan menggunakan jumlah fitur adalah 9 fitur yang digambarkan pada Gambar 5. Berdasarkan hasil pengujian dengan menggunakan 9 fitur pada lima kombinasi *folds*, didapatkan rata-rata akurasi adalah 21,68%. Kombinasi terbaik terdapat pada kombinasi 3 dengan F3 sebagai data uji dan *fold* lainnya sebagai data latih. Pada gambar, sumbu x adalah *folds* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



Gambar 1. Hasil Pengujian 9 Fitur

3.1.6 Hasil Pengujian 8 Fitur

Berikut adalah hasil pengujian seleksi fitur dengan menggunakan jumlah fitur adalah 8 fitur yang digambarkan pada Gambar 6. Berdasarkan hasil pengujian dengan menggunakan 8 fitur pada lima kombinasi *folds*, didapatkan rata-rata akurasi adalah 6,5%. Kombinasi terbaik terdapat pada kombinasi 4 dengan F4 sebagai data uji dan *fold* lainnya sebagai data latih. Pada gambar, sumbu x adalah *folds* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.

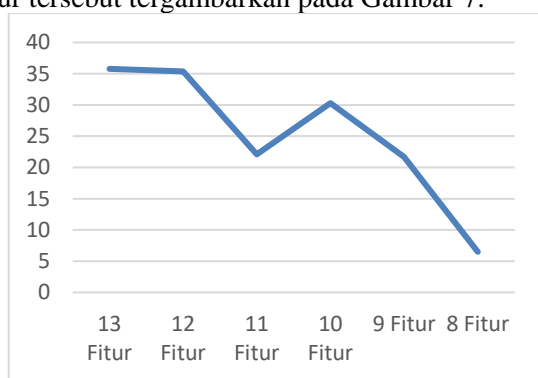


Gambar 1. Hasil Pengujian 8 Fitur

3.1.7 Hasil Penujian Seleksi Fitur

Pada hasil pengujian seleksi fitur di atas, didapatkan jumlah fitur terbaik untuk klasifikasi ini adalah 13 fitur. Seperti yang tergambar pada grafik, semakin sedikit fitur yang digunakan, semakin kecil pula nilai akurasi yang dihasilkan. Hal tersebut terjadi pada kasus masih terdapat lebih dari satu kelas pada data latih yang sedang *ditrain*, namun, semua fitur yang dimiliki sudah memiliki *value* yang sama yang menghasilkan nilai *gain* nol.

Grafik rata-rata akurasi pengujian seleksi fitur tersebut tergambar pada Gambar 7.



Gambar 1. Grafik Nilai Rata-rata Akurasi Pada Enam Jumlah Fitur

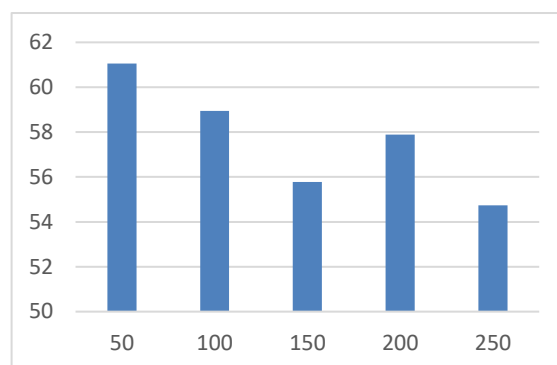
3.2 Pengujian K-fold cross validation Jumlah tree

Pada pengujian sebelumnya didapatkan jumlah fitur yang terbaik adalah 13 fitur. Maka dari itu pada pengujian ini digunakan 13 fitur untuk proses klasifikasi. Pengujian ini dilakukan untuk mengetahui kombinasi data latih dan data uji terbaik dari dataset yang dimiliki. Jumlah keseluruhan dataset adalah 95. Dari data tersebut dibagi menjadi lima *folds* yang masing-masing *fold* terdiri dari 19 data. Kombinasi data latih dan data uji dari *folds* yang ada telah tergambar pada perancangan pengujian. Berikut adalah hasil pengujian dengan menggunakan metode *K-fold cross validation*.

3.2.1 Hasil Pengujian Kombinasi Folds 1

Pada pengujian ini dilakukan klasifikasi menggunakan *random forest* dengan kombinasi F1 sebagai data uji, F2, F3, F4, dan F5 sebagai data latih. Pengujian dilakukan dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*. Grafik nilai akurasi berbanding jumlah *tree* tersebut tergambar pada Gambar 8.

Pada gambar, sumbu x adalah jumlah *tree* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.

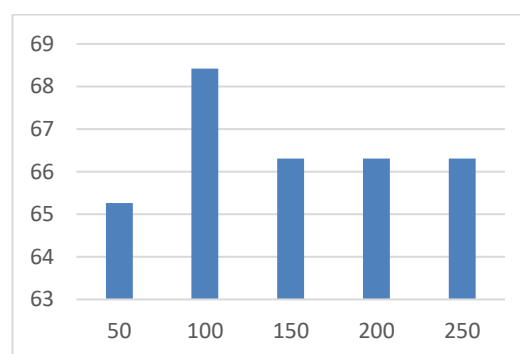


Gambar 8. Hasil Akurasi Pengujian Folds 1

Dari hasil pengujian pada kombinasi *folds* 1 dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*, didapatkan rata-rata akurasi sebesar 58,68%.

3.2.2 Hasil Pengujian Kombinasi Folds 2

Pada pengujian ini dilakukan klasifikasi menggunakan *random forest* dengan kombinasi F2 sebagai data uji, F1, F3, F4, dan F5 sebagai data latih. Pengujian dilakukan dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*. Grafik nilai akurasi berbanding jumlah *tree* tersebut tergambar pada Gambar 9. Pada gambar, sumbu x adalah jumlah *tree* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



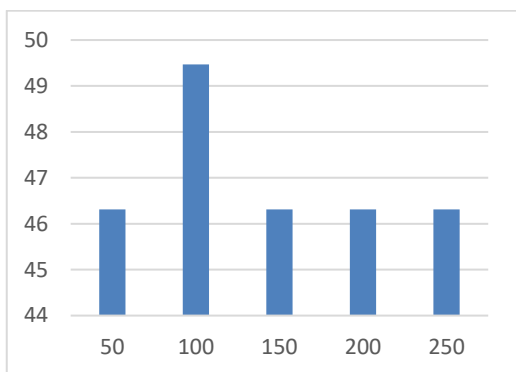
Gambar 9. Hasil Akurasi Pengujian Folds 2

Dari hasil pengujian pada kombinasi *folds* 2 dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*, didapatkan rata-rata akurasi sebesar 66,52%.

3.2.3 Hasil Pengujian Kombinasi Folds 3

Pada pengujian ini dilakukan klasifikasi menggunakan *random forest* dengan kombinasi

F3 sebagai data uji, F2, F1, F4, dan F5 sebagai data latih. Pengujian dilakukan dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*. Grafik nilai akurasi berbanding jumlah *tree* tersebut tergambarkan pada Gambar 10. Pada gambar, sumbu x adalah jumlah *tree* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.

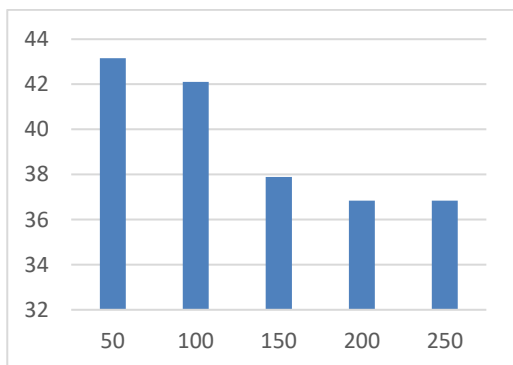


Gambar 10. Hasil Akurasi Pengujian Folds 3

Dari hasil pengujian pada kombinasi *folds* 3 dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*, didapatkan rata-rata akurasi sebesar 46,94%.

3.2.4 Hasil Pengujian Kombinasi Folds 4

Pada pengujian ini dilakukan klasifikasi menggunakan *random forest* dengan kombinasi F4 sebagai data uji, F2, F3, F1, dan F5 sebagai data latih. Pengujian dilakukan dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*. Grafik nilai akurasi berbanding jumlah *tree* tersebut tergambarkan pada Gambar 11. Pada gambar, sumbu x adalah jumlah *tree* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.



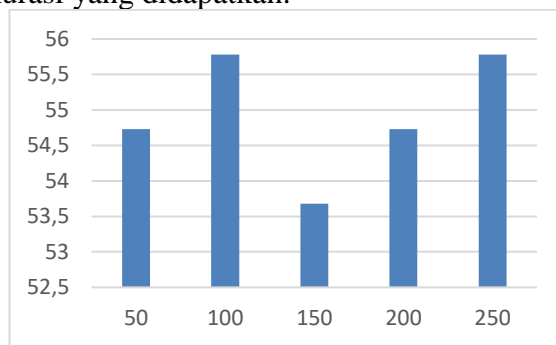
Gambar 11. Hasil Akurasi Pengujian Folds 4

Dari hasil pengujian pada kombinasi *folds* 4 dengan lima jumlah *tree* yang berbeda, 50, 100,

150, 200, dan 250 *tree*, didapatkan rata-rata akurasi sebesar 39,36%.

3.2.5 Hasil Pengujian Kombinasi Folds 5

Pada pengujian ini dilakukan klasifikasi menggunakan *random forest* dengan kombinasi F5 sebagai data uji, F2, F3, F4, dan F1 sebagai data latih. Pengujian dilakukan dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*. Grafik nilai akurasi berbanding jumlah *tree* tersebut tergambarkan pada Gambar 12. Pada gambar, sumbu x adalah jumlah *tree* yang digunakan dan sumbu y adalah tingkat akurasi yang didapatkan.

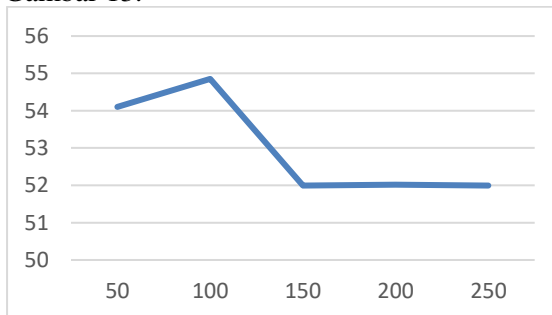


Gambar 12. Hasil Akurasi Pengujian Folds 5

Dari hasil pengujian pada kombinasi *folds* 5 dengan lima jumlah *tree* yang berbeda, 50, 100, 150, 200, dan 250 *tree*, didapatkan rata-rata akurasi sebesar 54,94%.

3.2.6 Hasil Pengujian Jumlah Tree

Pada hasil pengujian dengan lima jumlah *tree* yang berbeda, didapatkan jumlah *tree* yang paling baik untuk sistem ini adalah 100 *tree*. Seperti yang tergambar pada grafik, jumlah *tree* 50 ke 100 mengalami kenaikan, namun saat jumlah *tree* ditambah menjadi 150, nilai akurasi sistem berkurang. Grafik akurasi sistem berdasarkan jumlah *tree* tergambarkan pada Gambar 13.



Gambar 13. Grafik Nilai Rata-rata Akurasi Pada Lima Jumlah Tree

3.3 Analisis Hasil Pengujian

3.3.1 Analisis Hasil Pengujian Jumlah Fitur

Dari pengujian jumlah fitur dengan menyeleksi fitur pada dataset yang dimiliki dengan metode *information gain*, didapatkan bahwa fitur yang terbaik untuk klasifikasi ini adalah berjumlah 13 fitur. Dilihat dari grafik nilai rata-rata seleksi fitur, semakin sedikit fitur yang digunakan, nilai akurasi yang dihasilkan semakin kecil. Hal ini disebabkan beberapa dari percobaan yang dilakukan dengan *k-fold cross validation* mengalami *error* atau memiliki nilai akurasi sama dengan nol.

Pada kasus percobaan yang mendapatkan nilai akurasi sama dengan nol, hal ini terjadi ketika pada data latih yang dilakukan pemecahan pada proses pembangunan *tree* masih memiliki dua atau lebih kelas namun semua fiturnya telah memiliki nilai yang sama. Hal tersebut menyebabkan data yang masih memiliki kelas lebih dari satu tersebut tidak dapat dipisah sehingga pembangunan *tree* tersebut dianggap gagal dan mendapatkan nilai akurasi sebesar nol.

Dari Gambar 1 sampai dengan Gambar 6, terlihat bahwa setiap kombinasi *folds* yang dijalankan, menghasilkan nilai akurasi yang berbeda. Ada dua faktor utama yang menyebabkan perbedaan tingkat akurasi dari setiap kombinasi *folds*. Faktor yang pertama adalah penyebaran data latih pada setiap kombinasi *folds*. Penyebaran kelas pada data latih yang semakin merata, semakin tinggi nilai akurasi yang dihasilkan. Faktor kedua adalah pengaruh dari proses pengacakan data pada proses *training*. Apabila pada proses pengacakan data latih tersebut menghasilkan data baru yang memiliki kelas yang merata, maka data *random* tersebut menghasilkan nilai akurasi yang tinggi.

3.3.2 Analisis Hasil Pengujian Jumlah Tree

Setelah dilakukan pengujian dengan menggunakan metode *K-fold cross validation*, didapatkan kombinasi terbaik adalah kombinasi 2 dengan F2 sebagai data uji dan sisanya menjadi data latih. Jumlah *tree* terbaik yang didapatkan adalah 100 *tree*. Hasil rata-rata akurasi yang didapatkan dari semua percobaan adalah 53,094%.

Pada penelusuran penulis terhadap kekurangan dari proses klasifikasi ini adalah dikarenakan penyebaran kelas pada dataset tidak

merata. Data dengan kelas “Gangguan” memiliki jumlah yang tidak seimbang dengan kedua kelas lainnya. Hal ini berpengaruh saat pengacakan secara random pada proses *training*. Hal tersebut menyebabkan salah satu dari 150 *tree* yang terbangun, tidak terdapat kelas “Gangguan”, sehingga data uji dengan kelas asli “Gangguan” tidak dapat diklasifikasikan sebagai kelas “Gangguan”.

Masalah lain ditemukan ketika data uji memiliki *value* dari salah satu fitur yang tidak terdapat pada *value* fitur yang ada pada data latih yang telah diacak secara random. Seperti contoh data uji memiliki *value* “4” untuk fitur usia, namun pada *tree* yang terbangun tidak memiliki *value* “4” pada fitur usianya. Hal ini menyebabkan data uji tersebut tidak dapat diklasifikasikan atau pada implementasi kelasnya menjadi “X”.

4. KESIMPULAN

Berdasarkan hasil penelitian dan pengujian yang telah dilakukan pada tahap sebelumnya, maka dapat ditarik kesimpulan adalah penerapan metode random forest untuk klasifikasi penurunan fungsi kognitif pasien stroke ini dapat berjalan dengan rata-rata nilai ketepatan sebesar 53.094% berdasarkan pengujian *k-fold cross validation*.

Yang kedua, jumlah fitur yang digunakan dalam proses klasifikasi sangat berpengaruh. Jumlah fitur yang optimal pada penelitian ini adalah berjumlah 13 fitur.

Yang ketiga, jumlah *tree* memengaruhi hasil dari klasifikasi penurunan fungsi kognitif pasien stroke menggunakan metode random forest. Jumlah *tree* yang optimal pada penelitian ini adalah 100 *trees*.

Nilai akurasi setiap pengujian dapat berubah-ubah dikarenakan data yang digunakan dalam proses *training* terlebih dahulu diacak secara random.

DAFTAR PUSTAKA

- Binarwati, L., Mukhlash, I. & Soetrisno, 2017. Implementasi Algoritma Genetika untuk Optimalisasi *random forest* Dalam Proses Klasifikasi Penerimaan Tenaga Kerja Baru: Studi Kasus PT.XYZ. *Jurnal Sains dan Seni ITS*.
- Fitriani Ab, R., Bijaksana, M. A. & Adiwijaya, 2007. Analisis dan Implementasi Algoritma *random forest* Sebagai Sebuah Classifier

dalam Data Mining. *Tugas Akhir Telkom University.*

Malik, A. & Maulina, M., 2015. Fungsi Kognitif Pasien Stroke Berdasarkan Mini Mental State Examination (MMSE) di Rumah Sakit Umum Cut Meutia Kabupaten Aceh Utara. *Fakultas Kedokteran Universitas Malikussaleh.*

Rukmigayatri, R., 2015. Klasifikasi Kemunculan Titik Panas pada Lahan Gambut di Sumatera dan Kalimantan Menggunakan Algoritme *random forest.* *Fakultas Matematika dan Ilmu Pengetahuan Alam Institut Pertanian Bogor.*

Wardhani, Y. I., 2015. Faktor-Faktor yang Mempengaruhi Penurunan Fungsi Kognitif Pasien Stroke di Rumah Sakit Saiful Anwar Malang.