

Optimalisasi Algoritma Klasifikasi Ensemble Menggunakan Algoritma Genetika Untuk Prediksi Resiko Diabetes

Aldi Kristiawan Febianto¹, Castaka Agus Sugianto²

^{1,2}Program Studi Teknik Informatika, Politeknik TEDC Bandung, Jawa Barat, Indonesia

E-mail:¹aldikristiawanfebianto@gmail.com, ²Castaka@poltektedc.ac.id

Abstract

Diabetes mellitus (DM) is a global health problem that affects the quality of life and life expectancy of patients. Diabetes risk prediction can help in disease prevention and management. This study aims to optimize ensemble classification for diabetes risk prediction using a Genetic Algorithm (Optimize Selection). The Ensemble methods used are Bagging, Random Forest, and AdaBoost. The genetic algorithm is applied for Ensemble model hyperparameter optimization. The data used is the Pima Indians Diabetes dataset which consists of 768 samples with 8 features. Experimental results show that Ensemble Classification optimized with the Genetic Algorithm (Optimize Selection) produces quite good performance. The accuracy of the Genetic Algorithm (Optimize Selection) + Ensemble Bagging Classification Algorithm got a result of 97.14%, the Genetic Algorithm (Optimize Selection) + Random Forest Ensemble Classification Algorithm got a result of 98.57%, and the Genetic Algorithm (Optimize Selection) + AdaBoost Ensemble Classification Algorithm got a result of 99.87%. These results indicate that this approach can be an effective solution in diabetes risk prediction.

Keywords: Diabetes, Ensemble Classification, Genetic Algorithm, Risk Prediction, Hyperparameter Optimization.

Abstract

Diabetes mellitus (DM) merupakan masalah kesehatan global yang mempengaruhi kualitas hidup dan harapan hidup pasien. Prediksi risiko diabetes dapat membantu dalam pencegahan dan manajemen penyakit. Penelitian ini bertujuan untuk mengoptimalkan klasifikasi ensemble untuk prediksi risiko diabetes menggunakan Algoritma Genetika (Optimize Selection). Metode Ensemble yang digunakan adalah Bagging, Random Forest, dan AdaBoost. Algoritma genetika diterapkan untuk optimasi hiperparameter model Ensemble. Data yang digunakan adalah dataset Pima Indians Diabetes yang terdiri dari 768 sampel dengan 8 fitur. Hasil eksperimen menunjukkan bahwa Klasifikasi Ensemble yang dioptimalkan dengan Algoritma Genetika (Optimize Selection) menghasilkan kinerja yang cukup baik. Akurasi dari Algoritma Genetika (Optimize Selection)+Algoritma Klasifikasi Ensemble Bagging mendapat hasil 97.14%, Algoritma Genetika (Optimize Selection)+Algoritma Klasifikasi Ensemble Random Forest mendapat hasil 98.57%, dan Algoritma Genetika (Optimize Selection)+Algoritma Klasifikasi Ensemble AdaBoost mendapat hasil 99.87%. Hasil ini menunjukkan bahwa pendekatan ini dapat menjadi solusi yang efektif dalam prediksi risiko diabetes.

Keywords: Diabetes, Klasifikasi Ensemble, Algoritma Genetika (Optimize Selection), Prediksi Resiko diabetes, Optimisasi Hiperparameter.

1. Pendahuluan

Diabetes mellitus adalah penyakit metabolik kronis yang ditandai dengan kadar glukosa darah yang tinggi (hiperglikemia) dan gangguan metabolisme karbohidrat, lemak, dan protein akibat kekurangan insulin atau resistensi insulin [1]. Diabetes mellitus merupakan faktor risiko utama untuk penyakit jantung koroner, stroke,

gagal ginjal, kebutaan [2]. Penyakit ini menjadi masalah kesehatan global dengan kasus yang terus meningkat, terutama di negara-negara dengan pendapatan menengah dan rendah [3].

Observasi mendalam terhadap data diabetes yang melibatkan faktor-faktor seperti riwayat keluarga, indeks massa tubuh (BMI), tekanan darah, dan ketebalan kulit telah mengungkapkan pola dan tren tertentu yang berkorelasi dengan kemungkinan pengembangan diabetes. Faktor-faktor genetika juga berkontribusi dalam mempengaruhi risiko diabetes pada individu tertentu.

Self-monitoring of Blood Glucose (SMBG) telah menjadi komponen penting dalam pengobatan diabetes modern [4]. Proses ini memungkinkan pasien untuk melakukan pemeriksaan glukosa darah sendiri dengan menggunakan alat glucometer, yang dapat meningkatkan partisipasi pasien dalam proses terapeutik mereka [5]. Penggunaan model machine learning, seperti algoritma genetika dan klasifikasi ensemble, telah menjadi pendekatan yang menjanjikan dalam memprediksi risiko diabetes [6]. Algoritma genetika membantu dalam pemilihan fitur yang relevan, sementara klasifikasi *Ensemble* menggabungkan kelebihan beberapa model *machine learning* untuk meningkatkan akurasi prediksi [7].

Penelitian ini bertujuan untuk mengoptimalkan model *machine learning* untuk prediksi risiko diabetes menggunakan kombinasi algoritma genetika dan klasifikasi *ensemble*. Dengan memanfaatkan kekuatan algoritma genetika dalam pemilihan fitur dan kemampuan ensemble learning untuk meningkatkan kinerja model, diharapkan dapat diciptakan model prediktif yang akurat dan dapat diandalkan.

Beberapa penelitian yang telah dilakukan menggunakan algoritma *Ensemble* dilakukan oleh Mirqotussa'adah menerapkan teknologi data mining metode *Ensemble* pada Algoritma C4.5 dalam Mendiagnosa Diabetes, hasil eksperimen dengan menambahkan teknik bagging pada algoritma C4.5 didapatkan akurasi sebesar 69,79% [8]. Penelitian lainnya Menggunakan Algoritma Klasifikasi *Random Forest* untuk prediksi kemungkinan diabetes pada tahap awal yang dilakukan oleh Widya Apriliah didapatkan hasil akurasi 97.88% [2]. Yang dilakukan oleh Muflih Ihza Rifatama untuk menerapkan Seleksi Fitur menggunakan *Xgboost* pada data *Breast Cancer Wisconsin (Diagnostic)* yang memiliki 32 fitur dengan 569 record. Penyeleksian fitur dengan metode *XGBoost* berdasarkan feature importance atau fitur yang berpengaruh berdasarkan metode *XGBoost*, seleksi fitur *XGBoost* yang menggunakan 9 fitur mendapatkan akurasi 97.5% [9]. Penelitian Lainnya pernah dilakukan menggunakan Klasifikasi *Random forest* dengan judul "Penerapan Algoritma *Random Forest* Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo" untuk mencari hasil klasifikasi untuk penentuan penjurusan pada SMA Negeri Tugumulyo. untuk dataset yang dibuat berjumlah 340 baris yang terdiri dari 11 kolom, yaitu kolom nomor, nama, jenis kelamin, matematika, bahasa Indonesia, fisika, kima, ekonomi, sejarah, sosiologi, dan kolom label penjurusan. Dan mendapat hasil akurasi 93,14% [10].

2. Metodologi Penelitian

2.1. Menentukan Masalah

Sebelum dilakukan penelitian tahap pertama identifikasi masalah dari kinerja Klasifikasi *Ensemble*, Dari segi akurasi algoritma *Ensemble* yang saat ini digunakan. Untuk lebih mengoptimalkan Klasifikasi *Ensemble* dari pengelolaan fitur, Dilakukan penggabungan Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble* seperti *Bagging*, *Random Forest*, *AdaBoost*.

2.2. Menentukan Tujuan dan Ruang Lingkup

Untuk mengoptimalkan kinerja Algoritma Klasifikasi *Ensemble* dengan memasukkan Algoritma Genetika untuk lebih dioptimasi dalam pengelolaan fitur. Tujuan utama dari penelitian ini adalah untuk meningkatkan akurasi, dan ketepatan prediksi diabetes. Agar

dapat diketahui model mana yang yang lebih baik selain itu dapat mengoptimalkan pengelolaan fitur.

2.3. Mencari Litelaure

Mencari Litelaure atau tujuan Pustaka untuk mendukung kerangka pemahaman yang berkaitan dengan data mining, Algoritma Genetika (Optimize Selection), Algoritma Klasifikasi Ensemble. Data yang dikumpulkan Penulis melalui metode penelitian adalah.

1. Konsep *data mining*
2. algoritma klasifikasi *ensemble*
3. algoritma genetika (*Optimize Selection*)
4. konsep *RapidMiner*
5. hasil penelitian terkait

2.4. Pegumpulan Data

Proses pengumpulan data merupakan langkah penting dalam sebuah penelitian, dimana peneliti memperoleh data yang diperlukan untuk keperluan penelitian. Tanpa data yang memadai, hasil penelitian tidak akan dapat dihasilkan. Dalam penelitian ini, penulis menggunakan metode pengumpulan data sebagai berikut:

1. Data Primer

Data primer adalah data yang diperoleh langsung dari sumber yang terkait dengan penelitian. Dalam penelitian ini, data primer diperoleh dari dataset yang tersedia di platform (*Kaggle, .kaggle.com/datasets/uciml/pima-indians diabetes-database*).

2. Data Sekunder

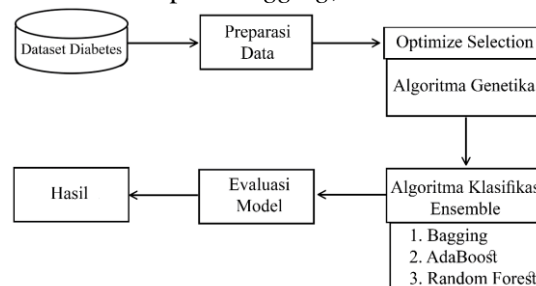
Data sekunder adalah data yang diperoleh melalui pihak kedua yang mengolah data untuk orang lain. Dalam penelitian ini, data sekunder mencakup jurnal, makalah, artikel yang berkaitan dengan data mining, Algoritma Genetika, Algoritma Klasifikasi *Ensemble*, serta data-data terkait penelitian.

3. Observasi

Observasi merupakan cara untuk pengumpulan data secara langsung atau tatap muka, seperti mengumpulkan informasi, dan melakukan wawancara, baik secara langsung maupun melalui media tertentu.

2.5. Metode Yang Digunakan

Pada Gambar tersebut, ada enam langkah yang dijelaskan dalam penelitian ini. Ini termasuk persiapan data diabetes, pengoptimalan pemilihan fitur menggunakan Algoritma Genetika (Optimize Selection)+Klasifikasi Ensemble evaluasi model, dan output yang dihasilkan. Klasifikasi Ensemble Seperti *Bagging, Random Forest* dan *AdaBoost*.



Gambar 1. Metode Yang Digunakan

Tahap pertama penelitian ini mengambil data pada *website (Kaggle, .kaggle.com/datasets/uciml/pima-indians diabetes-database)*. Data yang digunakan termasuk data public yang terdiri dari 786 data, yang mempunyai 8 atribut seperti *Pregnancies, Glucose, BloodPressure, SkinThickness, Insulin,*

BodyMasIndex, DiabetesPedigreeFunction, Age, Outcome. Adapun Data sebelum di preparasi seperti pada Tabel 1:

Tabel 1. Data Diabetes

No	Pregnant	Glucose	Blood Pressure	Skin Thickness	Insulin	Body Mas Index	Pedigree	Age	Outcome
1	6	148	72	35	0	33.6	0.627	50	1
2	1	85	66	29	0	26.6	0.351	31	0
3	8	183	64	0	0	23.3	0.672	32	1
.
.
786	0	137	40	35	168	43.1	2.288	33	1

Dari data pada Tabel 1 terdapat salah satu atribut kelas untuk mediagnosa ada tidaknya diabetes pada seseorang yaitu pada atribut *Outcome*. Kemudian data tersebut dilakukan preparasi data secara manual, Pada atribut *Outcome* yang berisikan angka (1) dan (0) dilakukan perubahan menjadi (1) *Positive* dan (0) menjadi *Negative*. Adapun data yang sudah di preparasi seperti Tabel 2:

Tabel 2. Hasil Preparasi

No	Pregnant	Glucose	Blood Pressure	Skin Thickness	Insulin	Body Mas Index	Pedigree	Age	Outcome
1	6	148	72	35	0	33.6	0.627	50	Positive
2	1	85	66	29	0	26.6	0.351	31	Negative
3	8	183	64	0	0	23.3	0.672	32	Positive
.
.
786	0	137	40	35	168	43.1	2.288	33	Positive

2.6. Proses dan Hasil Penelitian

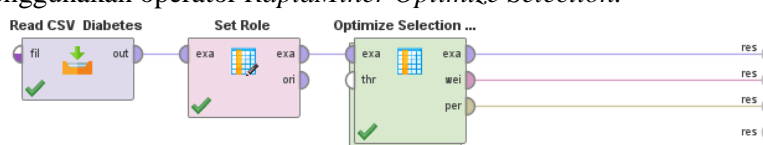
Algoritma Genetika (*Optimize Selection*) + Klasifikasi *Ensemble* Seperti *Random Forest, Bagging, dan AdaBoost*. Untuk mengetahui akurasi mana yang lebih besar dari setiap Klasifikasi *Ensemble* menggunakan Algoritma Genetika (*Optimize Selection*). Algoritma Genetika dipakai untuk mengoptimalkan pemilihan atribut dari Klasifikasi *Ensemble* pada data diabetes. Hasil dari tahapan-tahapan tersebut akan menghasilnya akurasi mana yang lebih besar dan dari segi atribut mana yang akan di pilih atau dioptimalkan dari setiap tahapan Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble*, Apakah dari setiap Klasifikasi *Ensemble Random Forest, Bagging, AdaBoost* pengoptimalnnya dari pemilihan atributnya akan sama atau berbeda.

3. Hasil dan Pembahasan

Pengujian dilakukan mengetahui nilai *Accuracy*, yang keluar dari setiap proses pengujian menggunakan Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble* seperti *Bagging, Random Forest, Adaboost*.

3.1. Proses Pengujian Algoritma Genetika + Klasifikasi *Ensemble*

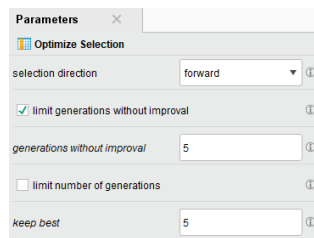
Pengujian ini akan dilakukan memnggunakan aplikasi *RapidMiner*, Disini kita akan memanfaatkan operator pada *RapidMiner* dapat dilihat pada Gambar 2 proses Algoritma Genetika menggunakan operator *RapidMiner Optimize Selection*.



Gambar 2. Proses Algoritma Genetika

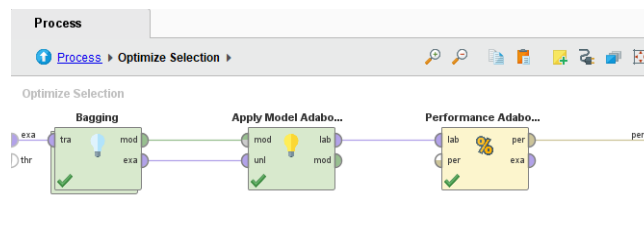
Proses pada Gambar 2 menunjukkan atribut yang digunakan dalam proses *RapidMiner*. Atribut *Read CSV* berfungsi untuk membaca dataset dari file CSV. Operator *Set Role* ini digunakan untuk menentukan peran tertentu seperti atribut kelas. Operator *Optimize Selection* untuk melakukan proses pemilihan atribut secara otomatis dan mengoptimalkan set atribut yang digunakan dalam *data mining*.

Backward Elimination. Pada penelitian penulis menggunakan *Forward Selection* untuk pengoptimalan dalam pemilihan fitur, *Forward Selection* sendiri adalah proses untuk membuat atribut yang optimal dengan menambahkan atribut satu per satu akan terus seperti itu namun akan terus seperti itu jika penambahan atribut tersebut meningkatkan kinerja model. Pada *parameters generations without imroval* dan *keep best* mempunyai nilai default 1 yang berarti algoritma pemilihan fitur standar digunakan. Namun penulis menggunakan nilai 5 agar dapat meningkatkan waktu proses juga dapat lebih mengoptimalkan kinerja model. Dapat dilihat pada Gambar 3 *Configuration parameters*.



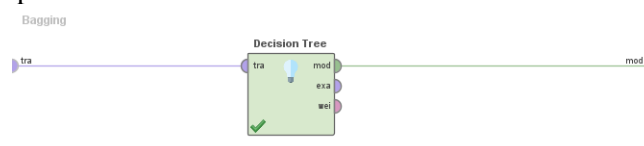
Gambar 3. Configuration parameters

Proses Selanjutnya Detail dari *Optimize Selection* terdapat operator *Klasifikasi Ensemble* yaitu *bagging*, *bagging* sendiri adalah operator termasuk kedalam *Algoritma Ensemble* untuk meningkatkan akurasi model machine learning. Operator ini dapat melatih beberapa model yang berbeda dan dapat menghasilkan akurasi yang baik dari beberapa model yang digabungkan. Dapat dilihat pada Gambar 4 proses *rapid miner* pengujian *Optimize Selection + Bagging*.



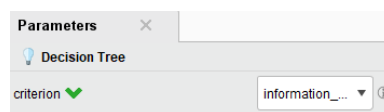
Gambar 4. Proses *Optimize Selection+Bagging*

Dalam detail proses *Bagging* model yang digabungkan itu menggunakan *Decision Tree* dapat dilihat pada Gambar 5:



Gambar 5. Detail Proses *Bagging*

Pada *Parameters Descision Tree* yaitu *criterion* menjadi *information-gain* dapat dilihat pada Gambar 6:



Gambar 6. Detail Parameters *Decision Tree*

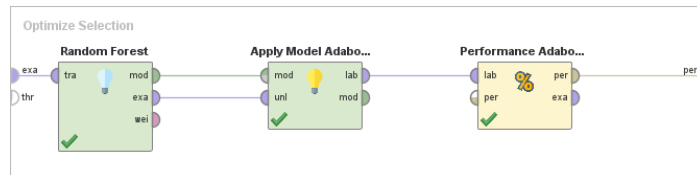
Dalam konteks ini, Bagging digunakan untuk meningkatkan kinerja model *Decision Tree*. Hasil pengujian Algoritma Genetika (*Optimize Selection*)+*Bagging* mendapat akurasi 97.14% dapat dilihat pada Gambar 7 berikut:

accuracy: 97.14%

	true Positive	true Negative	class precision
pred. Positive	252	6	97.67%
pred. Negative	16	494	96.86%
class recall	94.03%	98.80%	

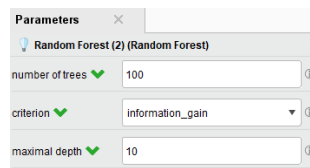
Gambar 7. Tabel Hasil Hasil Pengujian *Optimize Selection*+*Bagging*

Proses Selanjutnya Detail dari *Optimize Selection* terdapat operator Klasifikasi *Ensemble* yaitu *Random Forest*, *Random Forest* memiliki beberapa kelebihan, yaitu dapat meningkatkan hasil akurasi jika terdapat data yang hilang, dan untuk resisting outliers, serta efisien untuk penyimpanan sebuah data. Selain itu, *Random Forest* mempunyai proses seleksi fitur dimana mampu mengambil fitur terbaik sehingga dapat meningkatkan performa terhadap model klasifikasi[10]. Dapat dilihat pada Gambar 8:



Gambar 8. Proses *Optimize Selection*+*Random Forest*

Pada *Parameters Random Forest* yaitu *criterion* menjadi *information-gain* dapat dilihat pada Gambar 9:



Gambar 9. Detail Proses *Random Forest*

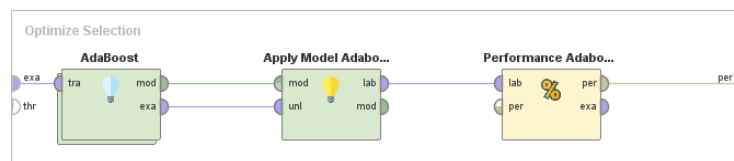
Hasil Pengujian Algoritma Genetika (*Optimize Selection*)+klasifikasi *Ensemble Random Forest* mendapat hasil akurasi 98.57% dapat dilihat pada Gambar 10 berikut:

accuracy: 98.57%

	true Positive	true Negative	class precision
pred. Positive	258	1	99.61%
pred. Negative	10	499	98.04%
class recall	96.27%	99.80%	

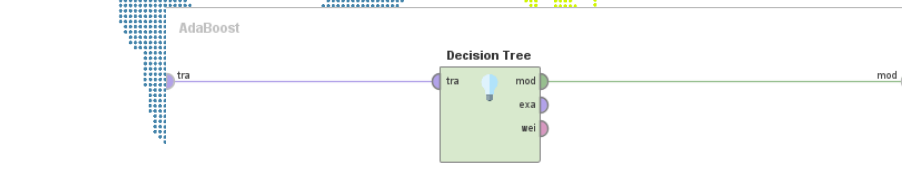
Gambar 10. Tabel Hasil Hasil Pengujian *Optimize Selection*+*Random Forest*

Proses Selanjutnya Detail dari *Optimize Selection* terdapat operator Klasifikasi *Ensemble* yaitu *AdaBoost*, *AdaBoost* sendiri sama seperti Algoritma Klasifikasi *Ensemble Bagging* yang menerapkan beberapa model berbeda juga dapat menghasilkan akurasi yang baik. Dapat dilihat pada Gambar 11:



Gambar 11. Proses *Optimize Selection*+*AdaBoost*

dalam detail proses *AdaBoost* model yang digabungkan itu menggunakan *Decision Tree* dapat dilihat pada Gambar 12:



Gambar 12. Detail Proses *AdaBoost*

Dalam konteks ini, *AdaBoost* juga digunakan untuk meningkatkan kinerja model *Decision Tree* Hasil pengujian Algoritma Genetika (*Optimize Selection*)+*AdaBoost* mendapat akurasi 99.87% dapat dilihat pada Gambar 13:

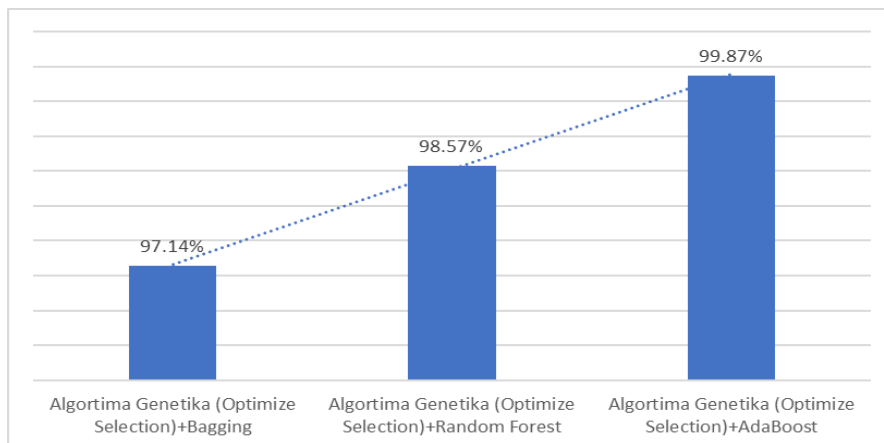
accuracy: 99.87%

	true Positive	true Negative	class precision
pred. Positive	267	0	100.00%
pred. Negative	1	500	99.80%
class recall	99.63%	100.00%	

Gambar 13. Tabel Hasil Hasil Pengujian *Optimize Selection*+*AdaBoost*

3.2. Hasil Pengujian Algoritma Genetika (*Optimize Selection*)+ Klasifikas Ensemble

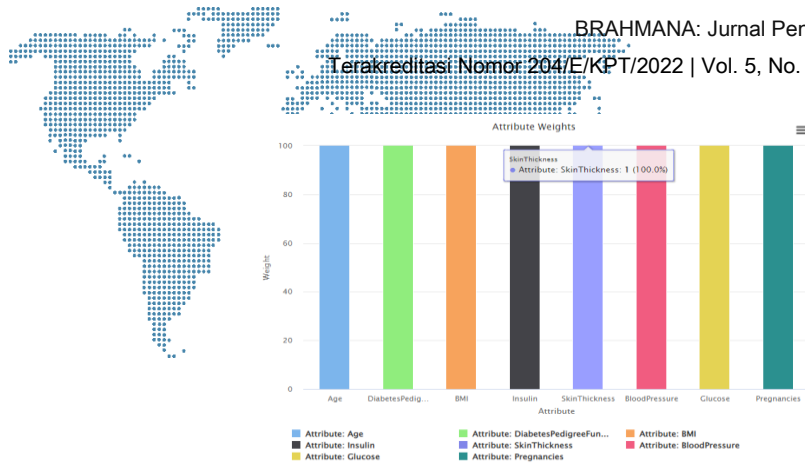
Hasil pengujian Algoritma Genetika (*optimize Selection*)+Algoritma Klasifikasi *Ensemble Bagging*, *Random Forest*, *AdaBoost* dilihat mendapat hasil yang cukup baik., Proses selanjutnya pembahasan untuk mendapatkan hasil penelitian dari hasil pengujian Algoritma Genetika (*optimize Selection*)+Algoritma Klasifikasi *Ensemble*, Hasil dari proses pengujian tersebut dapat dilihat pada Gambar 14:



Gambar 14. Grafik Hasil Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble*

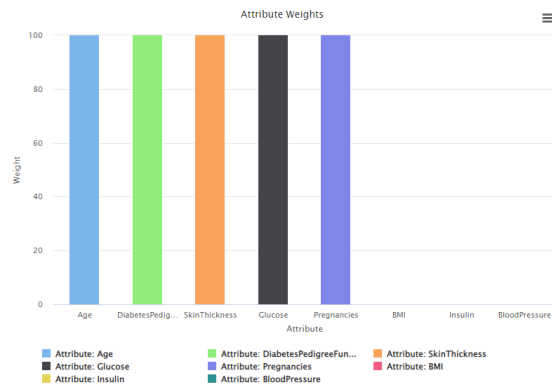
3.3. Hasil Atribut Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble*

Hasil dari pemilihan atribut yang dioptimalkan oleh Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble* berbeda beda tergantung klasifikasi *Ensemble* yang digunakan. Hasil dari pemilih atribut yang dioptimalkan oleh Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble Bagging* adalah semua atribut pada data diabetes dipilih dapat dilihat pada Gambar 15:



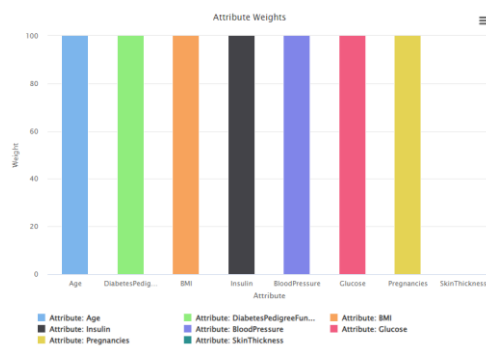
Gambar 15. Grafik Pemilihan Atribut *Optimize Selection*+Klasifikasi *Ensemble Bagging*

Hasil dari pemilih atribut yang dioptimalkan oleh Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble Random Forest* tidak semua dipilih dan di optimalkan namun hanya beberapa seperti Age, DiabetesPedigreeFunction, SkinThickness, Glucose, Pregnancies. Dapat dilihat pada Gambar 16:



Gambar 16. Grafik Pemilihan Atribut *Optimize Selection*+Klasifikasi *Ensemble Random Forest*

Hasil dari pemilih atribut yang dioptimalkan oleh Algoritma Genetika (*Optimize Selection*)+Klasifikasi *Ensemble AdaBoost* hampir semua dipilih dan di optimalkan kecuali SkinThickness tidak termasuk. Dapat dilihat pada Gambar 17:



Gambar 17. Grafik Pemilihan Atribut *Optimize Selection*+Klasifikasi *Ensemble AdaBoost*

4. Kesimpulan

Berdasarkan hasil penelitian yang dilakukan menggunakan Algoritma Genetika (*Optimize Selection*)+Algoritma Klasifikasi *Ensemble* pada data diabetes dapat disimpulkan bahwa Algoritma Genetika (*Optimize Selection*) cocok jika diintegrasikan

dengan Algoritma Klasifikasi *Ensemble* karna dari hasil akurasi bisa dinilai cukup baik, Algoritma Genetika (*Optimize Selection*)+Algoritma Klasifikasi *Ensemble Bagging* mendapat akurasi 97.14%. Algoritma Genetika (*Optimize Selection*)+Algoritma Klasifikasi *Ensemble Random Forest* mendapat akurasi sebesar 98.57%. Algoritma Genetika (*Optimize Selection*)+Algoritma Klasifikasi *Ensemble AdaBoost* mendapat akurasi sebesar 99.87%. Dari hasil yang disebutkan bahwa integrasi Algoritma Genetika (*Optimize Selection*) akan mendapat hasil yang paling tinggi jika di integrasikan dengan algoritma klasifikasi *Ensemble AdaBoost*. Dan dari hasil pemilihan atribut yang di optimalkan akan berbeda-berda tergantung Algoritma Klasifikasi *Ensemble* apa yang digunakan.

Daftar Pustaka

- [1] O. Daanouni, B. Cherradi, and A. Tmiri, "Predicting diabetes diseases using mixed data and supervised machine learning algorithms," *ACM Int. Conf. Proceeding Ser.*, no. October, 2019, doi: 10.1145/3368756.3369072.
- [2] W. Apriliah, I. Kurniawan, M. Baydhowi, and T. Haryati, "Prediksi Kemungkinan Diabetes pada Tahap Awal Menggunakan Algoritma Klasifikasi Random Forest," *J. Sist. Inf.*, vol. 10, no. 1, pp. 163–171, 2021, [Online]. Available: <http://sistemasi.ftik.unisi.ac.id>
- [3] L. M. Cendani and A. Wibowo, "Perbandingan Metode Ensemble Learning pada Klasifikasi Penyakit Diabetes," *J. Masy. Inform.*, vol. 13, no. 1, pp. 33–44, 2022, doi: 10.14710/jmasif.13.1.42912.
- [4] Amelia, "Self-Monitoring of Blood Glucose dalam Mencegah Neropati pada Ektremitas Bawah Pasien Diabetes Melitus Tipe II," *Indones. J. Nurs. Sci. Pract.*, pp. 58–72, 2020.
- [5] S. K. Kermani, A. Khatony, R. Jalali, M. Rezaei, and A. Abdi, "Accuracy and precision of measured blood sugar values by three glucometers compared to the standard technique," *J. Clin. Diagnostic Res.*, vol. 11, no. 4, pp. OC05–OC08, 2017, doi: 10.7860/JCDR/2017/23926.9613.
- [6] V. Agustin, M. M. P. Putra, and A. Husni, "Impact of Enzymatic Hydrolysis on Antioxidant Activity of Snakehead Fish (*Channa striata*) Head Protein Hydrolysate," *J. Ilm. Perikan. dan Kelaut.*, vol. 15, no. 1, pp. 44–56, 2023, doi: 10.20473/jipk.v15i1.38391.
- [7] M. Aryuni, "Penerapan Ensemble Feature Selection dan Klasterisasi Fitur pada Klasifikasi Dokumen Teks," *ComTech Comput. Math. Eng. Appl.*, vol. 4, no. 1, p. 333, 2013, doi: 10.21512/comtech.v4i1.2745.
- [8] M. Mirqotussa'adah, M. A. Muslim, E. Sugiharti, B. Prasetyo, and S. Alimah, "Penerapan Dizcretization dan Teknik Bagging Untuk Meningkatkan Akurasi Klasifikasi Berbasis Ensemble pada Algoritma C4.5 dalam Mendiagnosa Diabetes," *Lontar Komput. J. Ilm. Teknol. Inf.*, vol. 8, no. 2, p. 135, 2017, doi: 10.24843/lkjiti.2017.v08.i02.p07.
- [9] Muflih Ihza Rifatama, Mohammad Reza Faisal, Rudy Herteno, Irwan Budiman, and Muhammad Itqan Mazdadi, "Optimasi Algoritma K-Nearest Neighbor Dengan Seleksi Fitur Menggunakan Xgboost," *J. Inform. dan Rekayasa Elektron.*, vol. 6, no. 1, pp. 64–72, 2023, doi: 10.36595/jire.v6i1.723.
- [10] H. Marlina, A. Zulus, and O. L. W. H. Wijaya, "Penerapan Algoritma Random Forest Dalam Klasifikasi Penjurusan Di SMA Negeri Tugumulyo," *BRAHMANA J. Penerapan Kecerdasan Buatan Terakreditasi*, vol. 4, no. 2, pp. 138–143, 2023, [Online]. Available: <https://tunasbangsa.ac.id/pkm/index.php/brahmana/article/view/188>.