

Klasifikasi Diabetes Suku Indian Pima Menggunakan Kombinasi Metode *Random Forest* dan *Naive Bayes*

Diajukan untuk Menyusun Skripsi
di Jurusan Teknik Informatika Fakultas Ilmu Komputer UNSRI



Oleh :

Muhammad Gheddi Vijaya Liandra
NIM : 09021181520013

JURUSAN TEKNIK INFORMATIKA
FAKULTAS ILMU KOMPUTER UNIVERSITAS SRIWIJAYA
2020

LEMBAR PENGESAHAN TUGAS AKHIR

KLASIFIKASI DIABETES SUKU INDIAN PIMA MENGUNAKAN KOMBINASI METODE RANDOM FOREST DAN NAIVE BAYES

Oleh:

M. GHEDDI V L
NIM : 09021181520013

Palembang, Agustus 2020

Pembimbing I,



Samsuryadi, M.Kom., Ph.D
NIP. 197102041997021003

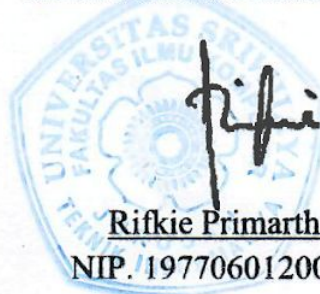
Pembimbing II,



Kanda Januar, S.Kom., MT.
NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

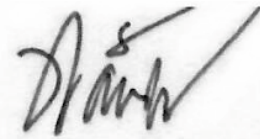
TANDA LULUS UJIAN SIDANG TUGAS AKHIR

Pada hari Selasa, tanggal 28 Juli 2020 telah dilaksanakan ujian sidang tugas akhir oleh Jurusan Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

Nama : M. Gheddi Vijaya L
NIM : 09021181520013
Judul : Klasifikasi Diabetes Suku Indian Pima menggunakan Kombinasi Metode *Random Forest* dan *Naive Bayes*.

1. Pembimbing I

Samsuryadi, M.Kom., Ph.D
NIP. 197102041997021003



2. Pembimbing II

Kanda Januar, S.Kom., MT.
NIP. 199001092019031012



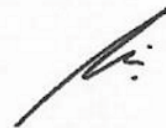
3. Penguji I

Julian Supardi, S.Pd., M.T
NIP. 197207102010121001



4. Penguji II

Rizki Kurniati, M.T
NIP. 199107122019032016



Mengetahui,

Ketua Jurusan Teknik Informatika



Rifkie Primartha, M.T
NIP. 197706012009121004

HALAMAN PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Muhammad Gheddi V L
NIM : 09021181520013
Program Studi : Teknik Informatika
Judul Skripsi : Klasifikasi Diabetes Suku Indian Pima
Menggunakan Kombinasi Metode *Random Forest*
dan *Naive Bayes*.

Hasil Pengecekan Software
iThenticate/Turnitin : 10 %

Menyatakan bahwa Laporan Projek saya merupakan hasil karya sendiri dan bukan hasil penjiplakan/plagiat. Apabila ditemukan unsur penjiplakan/plagiat dalam laporan projek ini, maka saya bersedia menerima sanksi akademik dari Universitas Sriwijaya sesuai dengan ketentuan yang berlaku.

Demikian, pernyataan ini saya buat dengan sebenarnya dan tidak ada paksaan oleh siapapun.



Palembang, Agustus 2020



MOTTO DAN PERSEMBAHAN

MOTTO:

- Menuntut ilmu adalah taqwa. Menyampaikan ilmu adalah ibadah. Mengulang-ulang ilmu adalah zikir. Mencari ilmu adalah jihad (Imam Al Ghazali).
- Orang berani jauh lebih langka dari orang jenius (Peter Thiel).

Kupersembahkan karya tulis ini kepada :

- *Kedua orang tuaku tercinta dan adikku tersayang*
- *Keluarga besarku*
- *Teman - teman Informatika 2015*
- *Fakultas Ilmu Komputer
Universitas Sriwijaya*

CLASSIFICATION OF PIMA INDIAN DIABETES USING COMBINED RANDOM FOREST AND NAIVE BAYES METHOD

By:

Muhammad Gheddi Vijaya L
09021181520013

ABSTRACT

Combined *Random Forest* and *Naive Bayes* method can produce better prediction accuracy. This study uses the combined Random Forest and Naive Bayes method to see whether the combined method will always produce better results than the individual method, or not. Based on the result of 280 test cases with Pima Indian diabetes dataset, the combined method only outperformed the accuracy of both the individual Random Forest and Naive Bayes method for 96 times, losing to Random Forest that outperformed the combined method and individual Naive Bayes method for 127 times.

Keywords: Data Mining, Random Forest, Naive Bayes, Diabetes.

Supervisor I,



Samsuryadi, M.Kom., Ph.D.
NIP. 197102041997021003

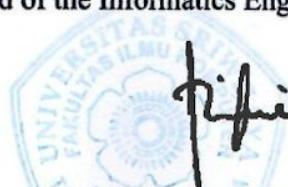
Palembang, August 2020
Supervisor II,



Kanda Januar, S.Kom., MT.
NIP. 199001092019031012

Approve,

Head of the Informatics Engineering Department



Rifkie Primartha, M.T
NIP. 197706012009121004

KLASIFIKASI DIABETES SUKU INDIAN PIMA MENGGUNAKAN
KOMBINASI METODE RANDOM FOREST DAN NAIVE BAYES

Oleh:

Muhammad Gheddi Vijaya L
09021181520013

ABSTRAK

Metode *Data Mining* ketika dikombinasikan dapat menghasilkan akurasi prediksi yang lebih baik. Pada penelitian ini menggunakan kombinasi metode *Random Forest* dan *Naive Bayes* untuk melihat apakah metode kombinasi selalu dapat mengungguli metode pembentuknya. Berdasarkan hasil pengujian dengan dataset suku Indian Pima, didapatkan bahwa akurasi dari kombinasi kedua metode hanya mengungguli metode *Random Forest* tunggal dan *Naive Bayes* tunggal sebanyak 96 kali, tertinggal dengan metode *Random Forest* tunggal yang unggul sebanyak 127 kali.

Keywords: *Data Mining*, *Random Forest*, *Naive Bayes*, Diabetes.

Pembimbing I,



Samsuryadi, M.Kom., Ph.D.
NIP. 197102041997021003


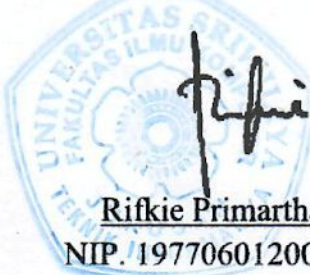
Palembang, August 2020
Pembimbing II,



Kanda Januar, S.Kom., MT.
NIP. 199001092019031012

Mengetahui,

Ketua Jurusan Teknik Informatika

Rifkie Primartha, M.T
NIP. 197706012009121004

KATA PENGANTAR

Puji syukur kepada Allah atas berkat dan rahmat-Nya yang telah diberikan kepada Penulis sehingga dapat menyelesaikan Tugas Akhir ini dengan baik. Tugas akhir ini disusun untuk memenuhi salah satu syarat guna menyelesaikan pendidikan program Strata-1 pada Fakultas Ilmu Komputer Program Studi Teknik Informatika di Universitas Sriwijaya.

Dalam menyelesaikan Tugas Akhir ini banyak pihak yang telah memberikan bantuan dan dukungan baik secara langsung maupun secara tidak langsung. Penulis ingin menyampaikan rasa terima kasih kepada:

1. Kedua orang tuaku dan seluruh keluarga besarku yang selalu mendokan serta memberikan dukungan baik moril maupun materil.
2. Bapak Jaidan Jauhari, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas Sriwijaya, Bapak Rifkie Primartha, M.T. selaku Ketua Jurusan Teknik Informatika dan Ibu Alvi Syahrini Utami, Ph.D. selaku Sekretaris Jurusan Teknik Informatika
3. Ibu Mastura Diana Marieska, M.T. selaku dosen pembimbing akademik penulis sekarang, yang keduanya telah membimbing, mengarahkan dan memberikan motivasi penulis dalam proses perkuliahan.
4. Bapak Samsuryadi, Ph.D. selaku dosen pembimbing I dan Bapak Kanda Januar, M.T. selaku pembimbing II yang telah membimbing, mengarahkan, dan memberikan motivasi penulis dalam proses pengerjaan Tugas Akhir.
5. Bapak Julian Supardi, M.T. selaku dosen penguji I, dan Ibu Rizki Kurniati, M.T. selaku dosen penguji II yang telah memberikan masukan dan dorongan dalam proses pengerjaan Tugas Akhir.
6. Seluruh dosen Program Studi Teknik Informatika Fakultas Ilmu Komputer Universitas Sriwijaya.

7. Mbak Wiwin, Pak Toni, Mbak Winda, Mbak Titi dan seluruh staf tata usaha yang telah membantu dalam kelancaran proses administrasi dan akademik selama masa perkuliahan.
8. Sahabat penulis (Abiyyu, Adi, Agus, Aji, Alex, Ajrul, Hanif, Ilham, Imam, Opan, Ricko, Rusdi, Sena, Winto, Yusuf) dan Kelompok “Budak Cewek Selalu Kompak” (Dhea, Decha, Fitrah, Limar, Oktef, Prilly, Qurrota, Saniyah), serta seluruh teman-teman jurusan Teknik Informatika yang telah berbagi keluh kesah, motivasi, semangat, dan canda tawa selama masa perkuliahan.

Penulis menyadari dalam penyusunan Tugas Akhir ini masih terdapat banyak kekurangan disebabkan keterbatasan pengetahuan dan pengalaman, oleh karena itu kritik dan saran yang membangun sangat diharapkan untuk kemajuan penelitian selanjutnya. Akhir kata semoga Tugas Akhir ini dapat berguna dan bermanfaat bagi kita semua.

Palembang, Agustus 2020



M Gheddi Vijaya L

DAFTAR ISI

	Halaman
HALAMAN JUDUL	i
HALAMAN LEMBAR PENGESAHAN	ii
HALAMAN TANDA LULUS UJIAN SIDANG TUGAS AKHIR	iii
HALAMAN PERNYATAAN	iv
MOTTO DAN PERSEMBAHAN	v
ABSTRACT	vi
ABSTRAK	vii
KATA PENGANTAR	viii
DAFTAR ISI	ix
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiv
DAFTAR LAMPIRAN	xv
BAB I PENDAHULUAN	I-1
1.1 Pendahuluan	I-1
1.2 Latar Belakang	I-1
1.3 Rumusan Masalah	I-3
1.4 Tujuan Penelitian	I-3
1.5 Manfaat Penelitian	I-4
1.6 Batasan Masalah	I-4
1.7 Sistematika Penulisan	I-5
1.8 Kesimpulan	I-6
BAB II KAJIAN LITERATUR	II-1

2.1	Pendahuluan	II-1
2.2	Landasan Teori	II-1
2.2.1	Diabetes Mellitus	II-1
2.2.2	Klasifikasi	II-2
2.2.3	Naive Bayes	II-3
2.2.4	Random Forest	II-4
2.2.5	Agregasi	II-7
2.2.6	Feature Selection	II-8
2.2.7	Confusion Matrix	II-10
2.2.8	K-Fold Cross Validation	II-11
2.3	Penelitian Lain yang Relevan	II-12
2.4	Kesimpulan	II-17
BAB III METODOLOGI PENELITIAN		III-1
3.1	Pendahuluan	III-1
3.2	Unit Penelitian	III-1
3.3	Pengumpulan Data	III-2
3.4	Tahapan Penelitian	III-2
3.4.1	Kerangka Kerja	III-3
3.4.2	Kriteria Pengujian	III-5
3.4.3	Format Data Pengujian	III-6
3.4.4	Alat Yang Digunakan Dalam Pelaksanaan Penelitian	III-6
3.4.5	Pengujian Penelitian	III-7
3.4.6	Analisis Hasil Pengujian dan Membuat Kesimpulan	III-8
3.5	Metode Pengembangan Perangkat Lunak	III-8
3.5.1	Fase Insepsi	III-8
3.5.2	Fase Elaborasi	III-9
3.5.3	Fase Kontruksi	III-9
3.5.4	Fase Transisi	III-10
3.6	Manajemen Proyek Penelitian	III-10
BAB IV PENGEMBANGAN PERANGKAT LUNAK		III-1
4.1	Pendahuluan	IV-1

4.2	Rational Unified Process	IV-1
4.2.1	Fase Insepsi	IV-1
4.2.2	Fase Elaborasi	IV-13
4.2.3	Fase Konstruksi	IV-24
4.2.4	Fase Transisi	IV-27
4.3	Kesimpulan	IV-41
BAB V ANALISIS HASIL PENGUJIAN		III-1
5.1	Pendahuluan	V-1
5.2	Hasil Percobaan Penelitian	V-1
5.2.1	Skenario Percobaan Kesatu	V-2
5.2.2	Skenario Percobaan Kedua	V-7
5.2.3	Skenario Percobaan Ketiga	V-12
5.2.4	Skenario Percobaan Keempat	V-17
5.3	Analisis Hasil Pengujian	V-24
5.4	Kesimpulan	V-29
BAB VI KESIMPULAN DAN SARAN		III-1
6.1	Pendahuluan	VI-1
6.2	Kesimpulan	VI-1
6.3	Saran	VI-2
DAFTAR PUSTAKA		

DAFTAR TABEL

	Halaman
II-1. Model Confusion Matrix	II-10
II-2. Penelitian Terkait	II-15
III-1. Rancangan Tabel Confusion Matrix untuk Hasil Klasifikasi	III-6
III-2. Work Breakdown Structure	III-11
IV-1. Kebutuhan Fungsional	IV-3
IV-2. Kebutuhan Non Fungsional	IV-4
IV-3. Definisi Aktor Use Case	IV-6
IV-4. Definisi Use Case	IV-6
IV-5. Skenario Use Case 01	IV-7
IV-6. Skenario Use Case 02	IV-8
IV-7. Skenario Use Case 03	IV-9
IV-8. Skenario Use Case 04	
IV-11	
IV-9. Implementasi Kelas	
IV-26	
IV-10. Rencana Pengujian Use Case Select Data	
IV-28	
IV-11. Rencana Pengujian Use Case Cross Validate NB	
IV-29	
IV-12. Rencana Pengujian Use Case Cross Validate RF	
IV-29	
IV-13. Rencana Pengujian Use Case Cross Validate NBRF	
IV-30	
V-1. Hasil Pengujian Skenario Pertama dengan 768 sampel	V-3
V-2. Hasil Pengujian Skenario Pertama dengan 600 sampel	V-3
V-3. Hasil Pengujian Skenario Pertama dengan 500 sampel	V-4
V-4. Hasil Pengujian Skenario Pertama dengan 400 sampel	V-5

V-5.	Hasil Pengujian Skenario Pertama dengan 300 sampel	V-5
V-6.	Hasil Pengujian Skenario Pertama dengan 200 sampel	V-6
V-7.	Hasil Pengujian Skenario Pertama dengan 100 sampel	V-6
V-8.	Hasil Pengujian Skenario Kedua dengan 768 sampel	V-7
V-9.	Hasil Pengujian Skenario Kedua dengan 600 sampel	V-8
V-10.	Hasil Pengujian Skenario Kedua dengan 500 sampel	V-9
V-11.	Hasil Pengujian Skenario Kedua dengan 400 sampel	V-9
V-12.	Hasil Pengujian Skenario Kedua dengan 300 sampel	V-10
V-13.	Hasil Pengujian Skenario Kedua dengan 200 sampel	V-11
V-14.	Hasil Pengujian Skenario Kedua dengan 100 sampel	V-11
V-15.	Hasil Pengujian Skenario Ketiga dengan 768 sampel	V-12
V-16.	Hasil Pengujian Skenario Ketiga dengan 600 sampel	V-13
V-17.	Hasil Pengujian Skenario Ketiga dengan 500 sampel	V-13
V-18.	Hasil Pengujian Skenario Ketiga dengan 400 sampel	V-14
V-19.	Hasil Pengujian Skenario Ketiga dengan 300 sampel	V-15
V-20.	Hasil Pengujian Skenario Ketiga dengan 200 sampel	V-16
V-21.	Hasil Pengujian Skenario Ketiga dengan 100 sampel	V-17
V-22.	Hasil Pengujian Skenario Keempat dengan 768 sampel	V-18
V-23.	Hasil Pengujian Skenario Keempat dengan 600 sampel	V-19
V-24.	Hasil Pengujian Skenario Keempat dengan 500 sampel	V-19
V-25.	Hasil Pengujian Skenario Keempat dengan 400 sampel	V-20
V-26.	Hasil Pengujian Skenario Keempat dengan 300 sampel	V-21
V-27.	Hasil Pengujian Skenario Keempat dengan 200 sampel	V-22
V-28.	Hasil Pengujian Skenario Keempat dengan 100 sampel	V-23
V-29.	Perbandingan Prediksi Tiap Classifier	V-27

DAFTAR GAMBAR

	Halaman
II-1. Taksonomi Data Mining	II-3
II-2. Arsitektur Umum <i>Random Forest</i>	II-5
II-3. Feature Selection Process	II-9
II-4. K-Fold Cross Validation	II-12
III-1. Kerangka Kerja Penelitian	III-4
IV-1. Diagram Use Case	IV-5
IV-2. Interface	IV-14
IV-3. Activity Diagram	IV-16
IV-4. Sequence Diagram	IV-20
IV-5. Class Diagram	IV-25
IV-6. Antarmuka Perangkat Lunak	IV-27

DAFTAR LAMPIRAN

Lampiran I. Source Code Program

L-1

Bab I

PENDAHULUAN

1.1 Pendahuluan

Pada bab ini membahas tentang latar belakang masalah, rumusan masalah, tujuan penelitian, manfaat penelitian, dan batasan masalah. Bab ini juga menjelaskan secara umum tentang keseluruhan penelitian.

Pendahuluan dimulai dengan penjelasan mengenai latar belakang masalah dimana algoritma yang digunakan dapat menyelesaikan kasus klasifikasi penderita diabetes dari suku Indian Pima. Setelah mengetahui permasalahan, maka penelitian tersebut perlu untuk dilakukan dengan menggunakan kombinasi algoritma *Random Forest* dan *Naive bayes* untuk prediksi penyakit diabetes suku Indian Pima.

1.2. Latar Belakang

Kemampuan untuk mendeteksi dan memprediksi sangatlah penting dalam banyak bidang kesehatan. *Diabetes Mellitus* merupakan salah satu dari beberapa masalah kesehatan yang paling menantang baik di negara berkembang maupun negara maju (Dewangan & Agrawal, 2015). Seperti yang dilansir dari *Tempo.com* pada tahun 2014 penderita diabetes terkhusus di Indonesia melonjak 500 ribu orang dari tahun sebelumnya. Akibat lonjakan tersebut, Indonesia mendapatkan peringkat kelima dunia penderita diabetes terbanyak.

Banyak penyandang *Diabetes Mellitus* yang terdiagnosis setelah mengalami komplikasi. Padahal, apabila dilakukan diagnosa secara dini, maka penanganan bisa dilakukan lebih cepat dan komplikasi yang membahayakan dapat dihindari (Lesmana, 2012). Untuk menekan angka kematian akibat diabetes, praktisi kesehatan harus melakukan diagnosa diabetes sedini mungkin.

Beberapa penelitian telah dilakukan terkait klasifikasi penyakit diabetes. Salah satunya adalah penelitian oleh (Tamilvanan & Bhaskaran, 2017) yang membandingkan beberapa metode *data mining* untuk mengukur performanya pada prediksi penyakit diabetes.. Pada penelitian ini membandingkan tiga metode yaitu *NB-Tree*, *Naive Bayes* dan *Random Forest*. Hasil dari penelitian ini menunjukkan bahwa metode *Naive bayes* memiliki akurasi tertinggi yaitu sebesar 76.3% dibanding kedua metode lainnya.

Penelitian lain juga telah dilakukan dalam klasifikasi diabetes oleh Chang et al. (2018) menggunakan lima metode, yaitu *Random Forest*, *C4.5*, *Multi Layer Perceptron*, *Logistic Regression*, dan *Support Vector Machine*. Hasil penelitian tersebut menunjukkan bahwa *Random Forest* memiliki akurasi tertinggi dalam prediksi diabetes dengan akurasi sebesar 82,7% dibandingkan 4 metode lainnya.

Penelitian lain menggunakan kombinasi metode *data mining* untuk klasifikasi diabetes dilakukan oleh (Devi, 2015). Dalam penelitian ini membandingkan 3 metode yaitu KNN, *Naive Bayes*, dan *Random Forest* beserta kombinasi dari ketiga metode tersebut. Hasil penelitian menunjukkan bahwa kombinasi dari *Naive Bayes* dan *Random Forest* memiliki akurasi prediksi lebih

tinggi dibandingkan kombinasi KNN dan *Naive Bayes*, kombinasi KNN, *Naive Bayes*, *Random Forest*, dan ketiga metode individual yaitu sebesar 99.5%.

Berdasarkan penelitian yang telah dilakukan sebelumnya, penelitian ini menggunakan kombinasi metode *Naive Bayes* dan *Random Forest* untuk prediksi diabetes suku Indian Pima karena kombinasi dari kedua metode memiliki akurasi yang tinggi dan *Random Forest* memiliki kelebihan dalam menghindari *overfitting*.

1.3. Rumusan Masalah

Berdasarkan dari latar belakang tersebut, maka rumusan masalah pada penelitian ini adalah sebagai berikut :

1. Bagaimana proses klasifikasi diabetes suku Indian Pima menggunakan kombinasi metode *Random Forest* dan *Naive Bayes* classifier?
2. Apakah tingkat akurasi kombinasi classifier selalu lebih baik dari salah satu classifier pembentuknya?

1.4 Tujuan Penelitian

Tujuan dari penelitian ini adalah sebagai berikut :

- a. Membangun kombinasi *Random Forest classifier* dengan *Naive Bayes classifier* untuk klasifikasi penyakit diabetes suku Indian Pima

- b. Mengembangkan perangkat lunak menggunakan classifier *Random Forest* dan *Naive Bayes* untuk klasifikasi penyakit diabetes suku Indian Pima.
- c. Menguji apakah akurasi dari kombinasi *Naive Bayes* dan *Random Forest classifier* lebih baik daripada classifier tunggal pembentuknya.

1.5 Manfaat Penelitian

Manfaat dari penelitian ini yaitu :

- a. Classifier pada perangkat lunak yang dihasilkan dapat membantu memprediksi penyakit diabetes.
- b. Kombinasi *Naive Bayes* dan *Random Forest classifier* dapat menjadi sebuah alternatif bagi masyarakat untuk mengenali penyakit diabetes secara dini.

1.6 Batasan Masalah

Batasan permasalahan dalam penelitian ini adalah sebagai berikut:

- a. Data yang digunakan adalah dataset suku Indian Pima.
- b. Output dari perangkat lunak adalah akurasi dari proses klasifikasi.
- c. Perangkat lunak hanya dapat menampilkan 1 buah output pada setiap eksekusi.
- d. Pengumpulan data akurasi hasil klasifikasi menggunakan Anaconda Navigator.

1.7 Sistematika penulisan

Sistematika penulisan skripsi ini adalah sebagai berikut

BAB 1. PENDAHULUAN

Pada bab ini diuraikan mengenai latar belakang, perumusan masalah, tujuan dan manfaat penelitian, bahasan masalah serta sistematika penulisan

BAB II. KAJIAN LITERATUR

Pada bab ini akan menjelaskan dasar- dasar teori yang digunakan dalam penelitian. Pembahasan dalam bab ini meliputi hasil penelitian terdahulu yang relevan dengan penelitian ini, penjelasan mengenai metode *Random Forest*, dan penjelasan mengenai algoritma *Naive Bayes* serta penjelasan lainnya yang berkaitan dengan penelitian ini.

BAB III. METODOLOGI PENELITIAN

Pada bab ini akan dibahas mengenai tahapan yang akan dilaksanakan pada penelitian ini. Masing- masing rencana tahapan penelitian dideskripsikan dengan rinci dengan mengacu pada suatu kerangka kerja. Di akhir bab ini berisi perancangan manajemen proyek pada pelaksanaan penelitian

1.8 Kesimpulan

Berdasarkan uraian di atas, penelitian ini dilakukan klasifikasi kemungkinan penyakit diabetes dari data medis pasien menggunakan metode agregasi *Naive Bayes* dan *Random Forest*. Metode agregasi adalah metode sederhana yang dapat menghasilkan akurasi lebih baik dibanding metode tunggal. Data yang digunakan pada penelitian ini adalah data dari suku Indian Pima.

Daftar Pustaka

- Alfarobi, I., Tutupoly, T. A., & Ade, S. (2018). KOMPARASI ALGORITMA C4.5, Naive bayes, dan Random Forest untuk klasifikasi data kelulusan mahasiswa Jakarta. *Jurnal Mitra Dan Teknologi Pendidikan, IV*(1), 1–14. Diakses 17 agustus 2019.
- American Diabetes Association (2016). Classification and Diagnosis of Diabetes, *39*(January), 13–22. <https://doi.org/10.2337/dc16-S005>. Diakses 14 agustus 2019.
- An, Y., Sun, S., & Wang, S. (2017). Naive Bayes Classifiers for Music Emotion Classification Based on Lyrics, (1), 635–638. Diakses 12 september 2019.
- Breiman, L. (2001). Random Forest, 1–33. Diakses 2 agustus 2019.
- Chang, H., & Chang, C. (2018). A Comparative Analysis of Data Mining Techniques for Prediction of Postprandial Blood Glucose : A Cohort Study, *7*(Icmhi 2017), 132–141. Diakses 2 agustus 2019.
- Devi, R. G., & Sumanjani, P. (2015). Improved classification techniques by combining KNN and Random Forest with Naive Bayesian Classifier, (March), 1–4. Dewangan, A., & Agrawal, P. (2015). Classification of Diabetes Mellitus Using Machine Learning Techniques, (5), 145–148. Diakses 16 mei 2019.

Dewangan, A., & Agrawal, P. (2015). Classification of Diabetes Mellitus Using Machine Learning Techniques, (5), 145–148.

Gorunescu, F. (2011). *Data Mining Concepts, Models and Techniques*. Springer. <https://doi.org/10.1007/978-3-642-19721-5>.

Diakses 14 september 2019.

Guyon, I. (2003). An Introduction to Variable and Feature Selection 1 Introduction, 3, 1157–1182. Diakses 12 juli 2019

Kruchten, P. (2002). Tutorial : Introduction to the Rational Unified Process *Proceedings of the 24th International Conference on Software Engineering. ICSE 2002*, 58113. Diakses 14 agustus 2019.

Kumar, V., & Minz, S. (2014). Feature Selection : A literature Review, 4(3). <https://doi.org/10.6029/smarter.2014.03.007>. Diakses 1 agustus 2019

Lesmana, I. P. D. (2012). Pengembangan Decision Tree J48 Untuk Diagnosis Penyakit Diabetes Mellitus, 2012(Semantik), 189–193. Diakses 30 juli 2019.

Nurhayati, Hadihardaja, I. K., Cahyono, M., & Soekarno, I. (2014). A Study of Hold-Out and K-Fold Cross Validation for Accuracy of Groundwater Modeling in Tidal Lowland Reclamation Using Extreme Learning Machine. *2nd International Conference on Technology, Informatics, Management, Engineering & Environment*, 228–233. Diakses 3 september 2019.

- Rahman, M. F., Darmawidjadja, M. I., & Alamsah, D. (2017). Klasifikasi Untuk Diagnosa Diabetes Menggunakan Metode Bayesian Regularization Neural Network (RBNN). *Jurnal Informatika*, 11(1), 36–45. Diakses 12 juli 2019
- Tamilvanan, B., & Bhaskaran, D. V. M. (2017). An Experimental Study of Diabetes Disease Prediction System Using Classification Techniques. *IOSR Journal of Computer Engineering*, 19(01), 39–44.
<https://doi.org/10.9790/0661-1901043944>. Diakses 10 juli 2019
- Vigneswari, D., Kumar, N. K., Raj, V. G., Gugan, A., & Vikash, S. R. (2019). Machine Learning Tree Classifiers in Predicting Diabetes Mellitus. *2019 5th International Conference on Advanced Computing & Communication Systems (ICACCS)*, 84–87. Diakses 16 september 2019
- Walia, H., Rana, A., & Kansal, V. (2017). A Naïve Bayes Approach for working on Gurmukhi Word Sense Disambiguation. *6th International Conference on Reliability, Infocom Technologies and Optimization*.
<https://doi.org/10.1109/ICRITO.2017.8342465>. Diakses 17 september 2019
- Zhao, Y. (2013). An example of calculating gini gain in CART, (3), 3–5.
Diakses 28 september 2019.