

**PENERAPAN *ENSEMBLE VOTING* PADA *DECISION TREE C4.5*,
K-NEAREST NEIGHBOR DAN *BACKPROPAGATION* UNTUK
KLASIFIKASI MAHASISWA *DROP OUT***

SKRIPSI



Disusun Oleh
DAFFA NUR CHOLIS
H92219044

**PROGRAM STUDI MATEMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL
SURABAYA**

2023

PERNYATAAN KEASLIAN

Saya yang bertanda tangan di bawah ini,

Nama : DAFFA NUR CHOLIS

NIM : H92219044

Program Studi : Matematika

Angkatan : 2019

Menyatakan bahwa saya tidak melakukan plagiat dalam penulisan skripsi saya yang berjudul "PENERAPAN *ENSEMBLE VOTING* PADA *DECISION TREE C4.5*, *K-NEAREST NEIGHBOR* DAN *BACKPROPAGATION* UNTUK KLASIFIKASI MAHASISWA *DROP OUT*". Apabila suatu saat nanti terbukti saya melakukan tindakan plagiat, maka saya bersedia menerima sanksi yang telah ditetapkan.

Demikian pernyataan keaslian ini saya buat dengan sebenar-benarnya.

Surabaya, 13-01-2023

Yang menyatakan,



DAFFA NUR CHOLIS

NIM. H92219044

LEMBAR PERSETUJUAN PEMBIMBING

Skripsi oleh

Nama : DAFFA NUR CHOLIS
NIM : H92219044
Judul Skripsi : PENERAPAN *ENSEMBLE VOTING* PADA *DECISION TREE C4.5*, *K-NEAREST NEIGHBOR* DAN *BACKPROPAGATION* UNTUK KLASIFIKASI MAHASISWA *DROP OUT*

telah diperiksa dan disetujui untuk diujikan.

Pembimbing I


Nurissaidah Ulinnuha, M.Kom
NIP. 199011022014032004

Pembimbing II


Dian Yuliati, M.Si
NIP. 198707142020122015

Mengetahui,
Ketua Program Studi Matematika
UIN Sunan Ampel Surabaya


Yuniar Farida, M.T
NIP. 197905272014032002

PENGESAHAN TIM PENGUJI SKRIPSI


Skripsi oleh

Nama : DAFFA NUR CHOLIS
NIM : H92219044
Judul Skripsi : PENERAPAN *ENSEMBLE VOTING* PADA *DECISION TREE C4.5*, *K-NEAREST NEIGHBOR* DAN *BACKPROPAGATION* UNTUK KLASIFIKASI MAHASISWA *DROP OUT*


Telah dipertahankan di depan Tim Penguji pada tanggal 13-01-2023

Mengesahkan,
Tim Penguji

Penguji I


Putroue Keumala Intan, M.Si
NIP. 198805282018012001


Penguji II


Dr. Abdulloh Hamid, M. Pd
NIP. 198506282014031003

Penguji III


Nurissaidah Uliyulha, M.Kom
NIP. 1990111022014032004

Penguji IV


Dian Yuliati, M.Si
NIP. 198707142020122015

Mengetahui,
Dekan Fakultas Sains dan Teknologi
Universitas Jember Ampel Surabaya


Dr. A. Saepul Hamdani, M.Pd
NIP. 196303312000031002



UIN SUNAN AMPEL
SURABAYA

KEMENTERIAN AGAMA
UNIVERSITAS ISLAM NEGERI SUNAN AMPEL SURABAYA
PERPUSTAKAAN

Jl. Jend. A. Yani 117 Surabaya 60237 Telp. 031-8431972 Fax.031-8413300
E-Mail: perpus@uinsby.ac.id

LEMBAR PERNYATAAN PERSETUJUAN PUBLIKASI
KARYA ILMIAH UNTUK KEPENTINGAN AKADEMIS

Sebagai sivitas akademika UIN Sunan Ampel Surabaya, yang bertanda tangan di bawah ini, saya:

Nama : DAFFA NUR CHOLIS
NIM : H92219044
Fakultas/Jurusan : SAINIS DAN TEKNOLOGI / MATEMATIKA
E-mail address : daffonurcholis28@gmail.com

Demi pengembangan ilmu pengetahuan, menyetujui untuk memberikan kepada Perpustakaan UIN Sunan Ampel Surabaya, Hak Bebas Royalti Non-Eksklusif atas karya ilmiah :

☒ Skripsi ☐ Tesis ☐ Disertasi ☐ Lain-lain (.....)*
yang berjudul :

PENERAPAN ENSEMBLE VOTING PADA DECISION TREE
C4.5, K-NEAREST NEIGHBOR DAN BACKPROPAGATION
UNTUK KLASIFIKASI MAHASISWA DROP OUT

beserta perangkat yang diperlukan (bila ada). Dengan Hak Bebas Royalti Non-Eksklusif ini Perpustakaan UIN Sunan Ampel Surabaya berhak menyimpan, mengalih-media/format-kan, mengelolanya dalam bentuk pangkalan data (database), mendistribusikannya, dan menampilkan/mempublikasikannya di Internet atau media lain secara *fulltext* untuk kepentingan akademis tanpa perlu meminta ijin dari saya selama tetap mencantumkan nama saya sebagai penulis/pencipta dan atau penerbit yang bersangkutan.

Saya bersedia untuk menanggung secara pribadi, tanpa melibatkan pihak Perpustakaan UIN Sunan Ampel Surabaya, segala bentuk tuntutan hukum yang timbul atas pelanggaran Hak Cipta dalam karya ilmiah saya ini.

Demikian pernyataan ini yang saya buat dengan sebenarnya.

Surabaya, 13 Januari 2023

Penulis

(Daffa Nur Choliz)
nama terang dan tanda tangan

ABSTRAK

PENERAPAN *ENSEMBLE VOTING* PADA *DECISION TREE C4.5*, *K-NEAREST NEIGHBOR* DAN *BACKPROPAGATION* UNTUK KLASIFIKASI MAHASISWA *DROP OUT*

Banyaknya mahasiswa yang mengalami *drop out* merupakan salah satu indikator yang mempengaruhi akreditasi perguruan tinggi. Akreditasi sangat penting bagi keberlangsungan sebuah lembaga. Ada atau tidaknya mahasiswa yang mengalami *drop out* menjamin kualitas dan mutu lulusan dari perguruan tinggi. Banyak faktor yang menyebabkan terjadinya *drop out* pada mahasiswa dan pada masing-masing perguruan tinggi memiliki peraturan yang berbeda-beda. Penelitian ini melakukan klasifikasi mahasiswa aktif dan mahasiswa *drop out* menggunakan 1092 data mahasiswa yang terdiri dari 557 data mahasiswa aktif dan 535 data mahasiswa *drop out*. Peneliti menggunakan data semester, Satuan Kredit Semester (SKS), Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), jalur masuk dan Uang Kuliah Tunggal (UKT) sebagai variabel independen. Peneliti melakukan klasifikasi menggunakan *Ensemble Voting* pada metode *Decision Tree C4.5*, KNN dan *Backpropagation* sebagai metode tunggal. Selain untuk mengetahui klasifikasi mahasiswa aktif dan mahasiswa *drop out* penelitian ini bertujuan untuk membuktikan apakah metode *Ensemble Voting* mampu mendapatkan hasil akurasi, presisi dan *recall* lebih baik dari pada metode tunggal. Peneliti menggunakan perbandingan data latih dan data uji sebesar 90:10. Peneliti memasukan hasil klasifikasi model tunggal ke dalam metode *Ensemble Voting*. Metode *Decision Tree C4.5* mendapat akurasi 95.45%, presisi 98.03% dan *recall* 92.59%. KNN mendapat akurasi 96.36%, presisi 100% dan *recall* 92.59%. *Backpropagation* mendapat akurasi 90.90%, presisi 95.83% dan *recall* 95.18%. Sedangkan pada *Ensemble Voting rule* yang digunakan adalah *Ensemble Soft Voting* dengan bobot (2,1,1). *Ensemble Voting* mampu memperbaiki akurasi, presisi dan *recall* dari masing-masing metode tunggal. *Ensemble Voting* mendapat akurasi 98.18%, presisi 100% dan *recall* 96.29%.

Kata kunci: *Drop Out*, *Ensemble Learning*, *Ensemble Voting*, Klasifikasi, Mahasiswa.

ABSTRACT

IMPLEMENTATION OF ENSEMBLE VOTING ON DECISION TREE C4.5, K-NEAREST NEIGHBOR AND BACKPROPAGATION FOR DROP OUT STUDENTS CLASSIFICATION

The number of students who have dropped out is an indicator that affects higher education accreditation. Accreditation is very important for the sustainability of an institution. The presence or absence of students who have dropped out guarantees the quality and quality of graduates from tertiary institutions. Many factors cause student drop out and each university has different regulations. This study classified active students and drop out students using 1092 student data consisting of 557 active student data and 535 drop out student data. Researchers use semester data, Semester Credit Unit, Semester Achievement Index, grade-point average, college entrance and tuition fees as independent variables. Researchers classify using Ensemble Voting on the Decision Tree C4.5 method, KNN and Backpropagation as a single method. In addition to knowing the classification of active students and drop out students, this study aims to prove whether the Ensemble Voting method is able to get better accuracy, precision and recall results than the single method. Researchers used a comparison of training data and test data of 90:10. Researchers enter the results of a single model classification into the Ensemble Voting method. The Decision Tree C4.5 method gets 95.45% accuracy, 98.03% precision and 92.59% recall. KNN gets 96.36% accuracy, 100% precision and 92.59% recall. Backpropagation gets 90.90% accuracy, 95.83% precision and 95.18% recall. Meanwhile, the Ensemble Voting rule used is Ensemble Soft Voting with a weight of (2,1,1). Ensemble Voting is able to improve the accuracy, precision and recall of each single method. Ensemble Voting got 98.18% accuracy, 100% precision and 96.29% recall.

Keywords: *Classification, Drop Out, Ensemble Learning, Ensemble Voting, Student.*

DAFTAR ISI

HALAMAN JUDUL	i
LEMBAR PERSETUJUAN PEMBIMBING	ii
PENGESAHAN TIM PENGUJI SKRIPSI	iii
HALAMAN PERNYATAAN KEASLIAN	iv
MOTTO	v
HALAMAN PERSEMBAHAN	vi
KATA PENGANTAR	vii
ABSTRAK	ix
ABSTRACT	x
DAFTAR ISI	1
DAFTAR TABEL	4
DAFTAR GAMBAR	5
I PENDAHULUAN	6
1.1. Latar Belakang Masalah	6
1.2. Rumusan Masalah	14
1.3. Tujuan Penelitian	15
1.4. Manfaat Penelitian	15
1.5. Batasan Masalah	16
1.6. Sistematika Penulisan	16
II TINJAUAN PUSTAKA	18
2.1. <i>Data Mining</i>	18
2.2. Klasifikasi	18
2.3. Mahasiswa	19
2.4. <i>Drop Out</i>	19
2.5. Semester	20
2.6. Satuan Kredit Semester (SKS)	20
2.7. Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK)	21

2.8. Jalur Masuk	21
2.9. Uang Kuliah Tunggal (UKT)	22
2.10. Normalisasi	22
2.11. <i>Decision Tree C4.5</i>	23
2.12. <i>K-Nearest Neighbors</i> (KNN)	26
2.13. <i>Backpropagation</i>	28
2.14. <i>Ensemble Voting</i>	34
2.15. <i>Confusion Matrix</i>	38
2.16. Integrasi Keislaman	40
2.16.1. Bermanfaat dalam Perspektif Islam	40
2.16.2. Menuntut Ilmu dalam Perspektif Islam	41
2.16.3. Mengulang Perkuliahan dalam Perspektif Islam	42
III METODE PENELITIAN	45
3.1. Jenis Penelitian	45
3.2. Sumber Data	45
3.3. Variabel Penelitian	45
3.4. Tahapan Penelitian	46
3.5. Tahapan Metode Penelitian	47
3.5.1. Tahapan Metode <i>Decision Tree C4.5</i>	48
3.5.2. Tahapan Metode KNN	48
3.5.3. Tahapan Metode <i>Backpropagation</i>	49
3.5.4. Tahapan Metode <i>Ensemble Voting</i>	50
IV HASIL DAN PEMBAHASAN	51
4.1. Deskripsi Data	51
4.2. <i>Pre-processing Data</i>	54
4.3. Perhitungan Metode <i>Decision Tree C4.5</i>	56
4.3.1. Membangun Model <i>Decision Tree C4.5</i>	56
4.3.2. Hasil dan Evaluasi <i>Decision Tree C4.5</i>	57
4.4. Perhitungan Metode KNN	60
4.4.1. Membangun Model KNN	60

4.4.2. Hasil dan Evaluasi KNN	62
4.5. Perhitungan Metode <i>Backpropagation</i>	64
4.5.1. Membangun Model <i>Backpropagation</i>	64
4.5.2. Hasil dan Evaluasi <i>Backpropagation</i>	65
4.6. Perhitungan Metode <i>Ensemble Voting</i>	68
4.6.1. Membangun Model <i>Ensemble Voting</i>	68
4.6.2. Hasil dan Evaluasi <i>Ensemble Voting</i>	74
4.7. Evaluasi Metode <i>Decision Tree C4.5</i> , KNN, <i>Backpropagation</i> dan <i>Ensemble Voting</i>	76
4.8. Integrasi Keilmuan	78
4.8.1. Kajian Keislaman Terkait Hasil Klasifikasi	78
4.8.2. Kajian Keislaman Terkait Analisa Studi Kasus	81
4.8.3. Kajian Keislaman Terkait Keberhasilan Klasifikasi	86
V PENUTUP	90
5.1. Kesimpulan	90
5.2. Saran	91
DAFTAR PUSTAKA	91
A Akar Pohon keputusan <i>Decision Tree C4.5</i>	103
B Pohon Keputusan <i>Decision Tree C4.5</i> Kiri	104
C Pohon Keputusan <i>Decision Tree C4.5</i> Kanan	105

UIN SUNAN AMPEL
S U R A B A Y A

DAFTAR TABEL

2.1 <i>Confusion Matrix</i>	39
2.2 Arti Nilai <i>Confusion Matrix</i> (Dinata et al., 2020)	39
4.1 Distribusi Data	54
4.2 Sampel Hasil Klasifikasi <i>Decision Tree C4.5</i>	58
4.3 Evaluasi <i>Decision Tree C4.5</i>	59
4.4 Sampel Data Latih Dan Data Uji	60
4.5 Uji Coba Parameter K	62
4.6 Sampel Hasil Klasifikasi KNN	62
4.7 Uji Coba Parameter <i>Hidden Layer</i>	65
4.8 Sampel Hasil Klasifikasi <i>Backpropagation</i>	66
4.9 Sampel Hasil Klasifikasi Model Tunggal	68
4.10 Sampel Hasil Klasifikasi Model Tunggal Dan <i>Ensemble Voting</i>	73
4.11 Uji Coba <i>Rule</i> Dan Bobot <i>Ensemble Voting</i>	74
4.12 Sampel Hasil Klasifikasi <i>Ensemble Voting</i>	74
4.13 Perbandingan Evaluasi Metode Klasifikasi	77
4.14 Uji Coba Perbandingan Data Latih dan Data Uji	78

DAFTAR GAMBAR

1.1 Hubungan kompleksitas model dan nilai error (Dong and Yu, 2019)	11
2.1 Model <i>Decision Tree C4.5</i> (Damanik et al., 2019)	24
2.2 Klasifikasi <i>K - Nearest Neighbor</i> (Zhang et al., 2018)	27
2.3 Jarak Dua Titik	28
2.4 Arsitektur <i>Backpropagation</i> (Pujianto et al., 2018)	29
2.5 Ensemble Voting (Assiri et al., 2020)	36
3.1 Tahapan Penelitian	46
3.2 Tahapan Metode Penelitian	47
3.3 Tahapan Metode <i>Decision Tree C4.5</i>	48
3.4 Tahapan Metode KNN	48
3.5 Tahapan Metode <i>Backpropagation</i>	49
3.6 Tahapan Metode <i>Ensemble Voting</i>	50
4.1 Data Mahasiswa	51
4.2 Sampel Data Penelitian	52
4.3 Sampel Data Penelitian	52
4.4 Sebaran Data	53
4.5 Normalisasi Data Penelitian	55
4.6 <i>Confusion Matrix Decision Tree C4.5</i>	59
4.7 <i>Confusion Matrix KNN</i>	63
4.8 <i>Confusion Matrix Backpropagation</i>	67
4.9 <i>Confusion Matrix Ensemble Voting</i>	75
4.10 Sampel Perbandingan Hasil Metode Klasifikasi	76

BAB I

PENDAHULUAN

1.1. Latar Belakang Masalah

Tidak sedikit dampak yang ditimbulkan dari perkembangan teknologi informasi. Masyarakat dapat menggunakan internet untuk mengakses informasi yang dibutuhkan. Manusia berlomba menciptakan sebuah sistem untuk memberikan informasi secara akurat. Banyak perusahaan atau instansi yang berusaha memberikan informasi yang berasal dari *database*. Banyak cara untuk mendapatkan informasi dari *database* salah satunya menggunakan teknik *data mining* (Sutoyo and Almaarif, 2020; Mughal et al., 2018). *Data mining* melibatkan strategi kombinasional dari teknologi basis data, analisis statistik dan *machine learning* (Derisma, 2020). Salah satu instansi yang dapat memberikan informasi penting adalah perguruan tinggi. Salah satu keuntungan dari informasi yang berasal dari perguruan tinggi adalah memahami informasi mahasiswa yang berpotensi *drop out*. Data mahasiswa *drop out* adalah salah satu indikator penurunan akreditasi perguruan tinggi (Armansyah, 2021; Alban and Mauricio, 2019; Manrique et al., 2019; Aesy et al., 2021). Salah satu poin penting ketika mengevaluasi standar akreditasi nasional untuk pendidikan tinggi adalah data mahasiswa *drop out*, informasi ini didapatkan dari BAN-PT tahun 2015. Informasi mahasiswa yang berpotensi *drop out* dapat digunakan sebagai bentuk evaluasi perguruan tinggi dan mengantisipasi adanya dampak buruk dari kasus mahasiswa *drop out*. Perguruan tinggi harus melakukan introspeksi jika mendapatkan

informasi banyak mahasiswa yang berpotensi *drop out*. Islam juga mengajarkan umatnya untuk selalu introspeksi diri dan berusaha memperbaiki diri dalam al-Qur'an surat Al-Hasyr ayat 18 berikut:

يَا أَيُّهَا الَّذِينَ آمَنُوا اتَّقُوا اللَّهَ وَلْتَنْظُرْ نَفْسٌ مَّا قَدَّمَتْ لِغَدٍ وَاتَّقُوا اللَّهَ إِنَّ اللَّهَ خَبِيرٌ بِمَا تَعْمَلُونَ

Artinya: “Wahai orang-orang yang beriman, bertakwalah kepada Allah dan hendaklah setiap orang memperhatikan apa yang telah diperbuatnya untuk hari esok (akhirat). Bertakwalah kepada Allah. Sesungguhnya Allah Maha Teliti terhadap apa yang kamu kerjakan.”

Ayat di atas menjelaskan mengenai perintah kepada orang-orang yang beriman untuk bertakwa kepada Allah SWT dan perintah untuk memperbaiki diri atau bermuhasabah diri. Perguruan tinggi harus memperhatikan setiap mahasiswanya terutama yang berpotensi *drop out*. Perlu adanya evaluasi yang dilakukan oleh perguruan tinggi jika mengetahui terdapat cukup banyak mahasiswa yang berpotensi mengalami *drop out*. *Data mining* bisa digunakan untuk membantu orang lain yang membutuhkan sebuah informasi dari *database*. Membantu orang yang sedang membutuhkan telah diajarkan dalam agama Islam yakni dalam hadits riwayat Ahmad nomor 7601 sebagai berikut:

حَدَّثَنَا يَزِيدُ أَخْبَرَنَا هِشَامُ بْنُ حَسَّانَ عَنْ مُحَمَّدِ بْنِ وَاسِعٍ عَنْ أَبِي هُرَيْرَةَ قَالَ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ مَنْ سَتَرَ أَخَاهُ الْمُسْلِمَ فِي الدُّنْيَا سَتَرَهُ اللَّهُ فِي الْآخِرَةِ وَمَنْ نَفَسَ عَنْ أَخِيهِ كُرْبَةً مِنْ كُرْبِ الدُّنْيَا نَفَسَ اللَّهُ عَنْهُ كُرْبَةً يَوْمَ الْقِيَامَةِ وَاللَّهُ فِي عَوْنِ الْعَبْدِ مَا كَانَ الْعَبْدُ فِي عَوْنِ أَخِيهِ

Artinya: “Telah menceritakan kepada kami Yazid telah mengabarkan

kepada kami Hisyam bin Hassaan dari Muhammad bin Wasi' dari Abu Hurairah, dia berkata; Rasulullah Shallallahu Alaihi Wassalam bersabda: "Barangsiapa menutupi aib saudaranya sesama muslim di dunia, maka Allah akan menutupi aibnya kelak pada hari kiamat. Dan barangsiapa melapangkan kesulitan saudaranya di dunia maka Allah akan melapangkan kesulitannya kelak pada hari kiamat, dan Allah senantiasa menolong hamba-Nya selama ia mau menolong sesamanya.""

Data mining melibatkan strategi kombinasional dari teknologi basis data, analisis statistik dan *machine learning* (Derisma, 2020). Salah satu instansi yang dapat memberikan informasi penting adalah perguruan tinggi. Salah satu keuntungan dari informasi yang berasal dari perguruan tinggi adalah memahami informasi mahasiswa yang berpotensi *drop out*. Mahasiswa yang mengalami *drop out* akan menyebabkan kerugian baik bagi diri sendiri dan pihak perguruan tinggi. Hukum Islam dapat dibagi menjadi hukum ibadah dan hukum mu'amalah. Saat ini, hukum mu'amalah berbentuk berbagai disiplin ilmu yang melibatkan subjek hukum yang berbeda. Dengan demikian, bidang pembahasan dan persoalan hukum Islam sangat luas. Solusi alternatif yang dapat dilakukan untuk mengatasinya adalah dengan merumuskan prinsip-prinsip fiqh yang merupakan generalisasi dari permasalahan fiqh tersebut. Salah satu kaidah fiqh (As-Suyuthi, 57) yang menjelaskan mudharat atau merugikan orang lain adalah sebagai berikut:

الضَّرَرُ يُزَالُ

Artinya: "Kemudharatan itu hendaklah dihilangkan."

Mahasiswa harus mengantisipasi terjadinya *drop out* dikarenakan supaya tidak merugikan orang lain. Pada tahun 2019, mahasiswa *drop out* di Indonesia

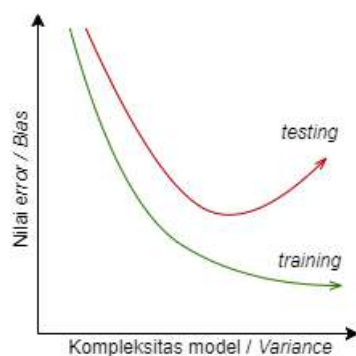
sebesar 7%, dengan arti bahwa terdapat 602.208 dari seluruh mahasiswa yang terdaftar merupakan mahasiswa *drop out*. Angka tersebut lebih kecil dari persentase pada tahun 2018 dengan angka 8%. Pada 2020 persentase mahasiswa *drop out* provinsi Jawa Timur sebesar 7% dan menempati urutan ke 14 pada skala nasional. Banyak penelitian pada studi kasus pengelompokan mahasiswa *drop out*. (Wahyuni et al., 2018) mengklasifikasikan mahasiswa *drop out* di Universitas Pembangunan Panca Budi dengan mendapatkan nilai akurasi 59,58% dan *classification error* sebesar 40,42%. (Alban and Mauricio, 2019) menggunakan *multilayer perceptron* dan *radial basis function* dengan nilai akurasi 96.3% dan 96,8%. (Limsathitwong et al., 2018) menggunakan *Decision Tree* dengan nilai presisi, *recall* dan F1-Measure masing-masing sebesar 0,80, 0,92 dan 0,85. (Jasmir et al., 2018) menggunakan k-Nearest Neighbor dengan nilai akurasi 69%. (Irmayanti et al., 2021) menggunakan metode Iterative Dichotomiser (ID3) dengan nilai akurasi, presisi dan *recall* sebesar 83,1%, 86,2% dan 73,2%. (Nagy and Molontay, 2018) menggunakan Decision Tree, Naive Bayes, k-NN, Linear Models dan Deep Learning dengan akurasi terbaik sebesar 81.1%. Serta (Andri and Paulus, 2021) menggunakan ADTree dan NNge dengan akurasi masing-masing 97.25% dan 96.2%.

Banyak faktor yang menyebabkan terjadinya *drop out* pada mahasiswa dan pada masing-masing perguruan tinggi memiliki peraturan yang berbeda-beda mengenai ketentuan mahasiswa *drop out* (Casanova et al., 2018; Aldowah et al., 2020). Penelitian ini mencoba mengklasifikasi mahasiswa *drop out* di Universitas Islam Negeri Sunan Ampel (UINSA) Surabaya. Dalam Buku Pedoman Akademik 2018 UINSA terdapat beberapa ketentuan mahasiswa *strata 1* yang terancam *drop out* diantaranya jika pada semester 14 belum menyelesaikan tanggungan SKS,

apabila tidak dapat menyelesaikan minimal 40 SKS pada akhir semester 4 serta memperoleh IPK dibawah 2.00. Banyak metode *data mining* yang dapat digunakan untuk klasifikasi (Dalipi et al., 2018; Kuzilek et al., 2018). Banyak hal yang menyebabkan mahasiswa mengalami *drop out* baik dari faktor internal dan maupun faktor eksternal. Contoh faktor internal diantaranya adalah kurangnya nilai IPK, besar biaya UKT dan lainnya. Sedangkan faktor eksternal diantaranya adalah kesibukan di luar perkuliahan, ketidakcocokan dengan jurusan yang diambil ataupun dengan universitas, melakukan tindakan kriminalitas, rendahnya motivasi belajar dan kondisi keuangan keluarga.

Banyak metode *machine learning* yang mendapatkan hasil berbeda-beda walaupun dengan data yang sama. Permasalahan yang dihadapi metode *machine learning* ini dapat diatasi menggunakan *Ensemble Learning*. *Ensemble Learning* mengumpulkan hasil dari beberapa model dan dijadikan dalam satu tempat dengan harapan mendapat hasil yang lebih baik (Maros and Juniar, 2016). *Ensemble Learning* dari beberapa *machine learning* diharapkan memiliki kinerja umum yang lebih baik daripada *machine learning* tunggal, terutama pada kondisi yang bervariasi atau dalam jangka panjang (Kumar and Jain, 2020). *Ensemble Learning* menggabungkan beberapa metode *machine learning* dan digunakan secara bersamaan. Beberapa metode tersebut akan dikombinasikan dengan harapan dapat mendapatkan hasil akurasi yang lebih baik. Ada banyak metode yang digunakan di dalam *Ensemble Learning* diantaranya adalah *Bagging*, *Voting*, *Adaboost* dan *Stacking* (Satmoko et al., 2018). Jika dilihat pada kompensasi nilai bias dan varians pada metode tunggal *machine learning* nilai kompleksitas model akan semakin tinggi namun nilai *error* prediksi akan semakin kecil pada data *training*. Pada saat *testing* nilai *error* yang diharapkan semakin kecil justru semakin

meningkat (Dong and Yu, 2019). Berikut adalah gambaran kompensasi nilai bias (*error*) dan *variance* metode tunggal pada saat *training* dan *testing* data:



Gambar 1.1 Hubungan kompleksitas model dan nilai error (Dong and Yu, 2019)

Saat *training* nilai kompleksitas model tunggal akan semakin tinggi namun nilai *error* prediksi akan semakin kecil. Pada saat *testing* nilai *error* yang diharapkan semakin kecil justru semakin meningkat (Dong and Yu, 2019). Itulah mengapa banyak metode klasifikasi yang mendapatkan hasil berbeda walaupun dengan data yang sama yaitu dikarenakan metode-metode tersebut memiliki pola pikir yang berbeda serta memiliki kekurangan dan kelebihan masing-masing. *Ensemble Learning* diharapkan mampu memperbaiki permasalahan yang dialami oleh metode tunggal dengan menggabungkan model tersebut.

Banyak penelitian menggunakan *Ensemble Learning* pada kasus klasifikasi. (Pawlovsky, 2018) menerapkan *Ensembles Learning* pada metode *K-Nearest Neighbors* (KNN) untuk diagnosis penyakit jantung dan mendapat akurasi mendekati 85%. (Rad and Behjat, 2019) menggunakan *Ensemble Learning* pada klasifikasi dokumen dengan menerapkan metode *Backpropagation*, SVM dan KNN dengan akurasi 98.21%. (Pristyanto, 2019) menerapkan *Ensemble Boosting* untuk meningkatkan kinerja metode *Decision Tree*, *Naive Bayes* dan SVM dengan akurasi terbaik pada metode *Adaptive Boosting Decision Tree* sebesar 94.37%, sensitivitas sebesar 94.37% dan spesifisitas sebesar 97.73%. (Satmoko et al., 2018)

menggunakan *Multiclassifier Ensemble Learning* dan *Chi-Square* untuk meningkatkan akurasi pendeteksian serangan *Distributed Denial of Service* dengan akurasi 99,68%. (Sudiyarno et al., 2021) menggunakan *Ensemble Learning* pada peningkatan performa pendeteksian Anomali dengan nilai akurasi, presisi, *recall* dan F-1 dari *Naive Bayes* sebesar 77,4%, 98,8%, 61,1% dan 75,5% sedangkan *Ensemble Naive Bayes* mendapatkan nilai akurasi, presisi, *recall* dan F-1 sebesar 96,8%, 98%, 96,1% dan 97,2%.

Penelitian ini mencoba menerapkan *Ensemble Learning* pada studi kasus klasifikasi mahasiswa *drop out*. (Patacsil, 2020) menerapkan *Ensemble Learning* pada metode *Decision Tree* dan *Forest Tree* untuk prediksi siswa putus sekolah menggunakan data lulusan sekolah, jenis kelamin, nilai mata pelajaran bahasa Inggris, IPA, Matematika dan nilai rata-rata. Pada penelitian tersebut didapatkan hasil bahwa *Ensemble Learning* memiliki peningkatan marginal dalam kinerja klasifikasi dengan nilai akurasi terbaik adalah 69,22% dengan presisi lulus 56,8 %, presisi *drop out* 77,44 %, *recall* lulus 62,49% dan *recall drop out* 73,04%. (Senthil Kumaran and Malar, 2021) juga menerapkan *Ensemble Learning* pada metode *Naive Bayes*, *SVM* dan *Regresi Logistik* untuk pengelompokan siswa putus sekolah menggunakan data tingkat pendidikan, kinerja, tujuan pembelajaran, referensi belajar, metode belajar, faktor kepercayaan diri dalam pra-tes, waktu belajar, partisipasi forum, interaksi dengan guru, modul belajar dan kinerja modul sebelumnya. Hasilnya menunjukkan bahwa *Ensemble* dapat meningkatkan kinerja prediksi secara signifikan dengan nilai akurasi sebesar 94,24%, presisi 95,62%, *recall* 93,12%, f-measure 94,1% dan AUC 0,91%.

Penelitian ini menerapkan *Ensemble Learning* pada metode *Decision Tree C4.5*, *KNN* dan *Backpropagation*. *Decision Tree C4.5* mampu mengatasi *missing*

data, mampu mengolah data dengan tipe diskrit maupun numerik, dan memiliki *rule* yang mudah diinterpretasikan (Benkercha and Moulahoum, 2018). *Decision Tree C4.5* digunakan oleh (Haidar et al., 2020) pada prediksi mahasiswa *drop out* menggunakan data IPK, jarak rumah, lama kuliah, status pekerjaan, status keluarga dan status keuangan kuliah dengan memperoleh nilai akurasi 96,45%, presisi 96,90% dan *recall* 95,38%. KNN memberikan hasil yang lebih baik dalam menggeneralisasi bahkan dari dataset *training* yang cukup kecil (Agrawal, 2019). KNN terbukti lebih akurat dari pada *Decision Tree C4.5* setelah dibandingkan oleh (Atma and Setyanto, 2018) dalam identifikasi mahasiswa *drop out* menggunakan data jurusan, jenis kelamin, waktu kuliah, status pekerjaan, IPK dan jumlah SKS dengan akurasi KNN sebesar 98.34%, presisi 100% dan *recall* 93.37%. *Backpropagation* cukup handal dalam menyelesaikan masalah (Purba et al., 2019; Setti et al., 2019). *Backpropagation* tidak membutuhkan fitur lain dari fungsi yang akan dipelajari dan hanya memiliki jumlah input yang akan digunakan sebagai parameter tuning. *Backpropagation* digunakan oleh (Sari et al., 2021) untuk melakukan prediksi mahasiswa *drop out* menggunakan data nilai IPS, jumlah mata kuliah mengulang dan jumlah absensi dan mendapat nilai akurasi dari *confusion matrix* sebesar 98,2%.

Penelitian ini mencoba menggabungkan *Decision Tree C4.5*, KNN dan *Backpropagation* menggunakan *Ensemble Voting*. *Ensemble Voting* mampu mempertimbangkan hasil yang didapatkan dari masing-masing metode tunggal sehingga mendapatkan hasil akurasi yang lebih baik. (Satmoko et al., 2018) menggunakan *Ensemble Hard Voting* pada *Multiclassifier Ensemble Learning* dengan metode KNN, *Naive Bayes*, dan *Random Forest* dan mendapat nilai akurasi sebesar 99,68%. (Kajornrit and Chaipornkaew, 2017) menggunakan *Ensemble*

Voting pada metode regresi linier, *Backpropagation*, SVM dan KNN dan mendapatkan fakta bahwa *Ensemble Voting* dapat meningkatkan nilai akurasi dengan rata-rata MAE 0.115% dan RMSE 0.175%. (Assiri et al., 2020) menggabungkan *Logistic Regression*, SVM dengan *Backpropagation* menggunakan *Ensemble Hard Voting* dan menunjukkan akurasi 99,42%, presisi 99,40%, *recall* 99,40%, F1 99,40%, F2 99,40% dan F3 99,40%. Penelitian ini mencoba menerapkan metode *Decision Tree C4.5*, KNN dan *Backpropagation* pada *Ensemble Voting* dalam studi kasus pengelompokan mahasiswa yang berpotensi *drop out*. Penelitian ini menggunakan data semester, jumlah SKS, nilai IPK, nilai IPS, jalur masuk dan nominal UKT dengan evaluasi menggunakan nilai akurasi, presisi dan *recall*. Akurasi menunjukkan persentase ketepatan prediksi mahasiswa *drop out* dan tidak *drop out* dari seluruh mahasiswa. Presisi menunjukkan persentase mahasiswa *drop out* dari seluruh mahasiswa yang diprediksi *drop out*. *Recall* menunjukkan persentase mahasiswa yang diprediksi *drop out* dibandingkan seluruh mahasiswa yang sebenarnya *drop out*.

1.2. Rumusan Masalah

1. Bagaimana perbandingan hasil klasifikasi metode *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting* dari ketiga metode pada pengelompokan mahasiswa yang berpotensi *drop out*?
2. Bagaimana evaluasi klasifikasi mahasiswa *drop out* pada metode *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting*?

1.3. Tujuan Penelitian

1. Mengetahui perbandingan hasil klasifikasi metode *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting* dari ketiga metode pada pengelompokan mahasiswa yang berpotensi *drop out*.
2. Mengetahui evaluasi klasifikasi mahasiswa *drop out* pada metode *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting*.

1.4. Manfaat Penelitian

1. Manfaat bagi peneliti

Penelitian ini dapat menambah wawasan peneliti dalam studi kasus mahasiswa *drop out*. Peneliti dapat memahami metode *Decision Tree C4.5*, KNN dan *Backpropagation* secara mendalam serta penerapannya dalam metode *Ensemble Voting*.

2. Manfaat bagi pembaca

Klasifikasi mahasiswa *drop out* ini dapat memberikan pembelajaran mengenai metode penelitian yang digunakan. Pembaca dapat mempelajari rumus serta perhitungan manual dari *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting*. Pembaca juga akan memahami konsep *drop out* pada mahasiswa.

3. Manfaat bagi institusi

Klasifikasi mahasiswa *drop out* ini dapat membantu instansi perguruan tinggi dalam melakukan persiapan serta antisipasi dalam mengambil langkah untuk mengurangi mahasiswa *drop out*. Informasi mahasiswa *drop out* juga dapat digunakan oleh program studi sebagai persiapan untuk meningkatkan

akreditasi program studi tersebut.

1.5. Batasan Masalah

1. Klasifikasi mahasiswa *drop out* ini menggunakan data mahasiswa aktif dan mahasiswa *drop out* di UINSA dari angkatan 2015 hingga angkatan 2021.
2. Peneliti menggunakan *Ensemble Hard Voting* dan *Ensemble Soft Voting*
3. Evaluasi klasifikasi yang digunakan adalah nilai akurasi, presisi dan *recall*.

1.6. Sistematika Penulisan

1. BAB I PENDAHULUAN

Di dalam bab pendahuluan berisi mengenai latar belakang serta alasan pemilihan topik dan metode penelitian, rumusan masalah yang akan diselesaikan dalam penelitian, tujuan penelitian yang berisi penyelesaian dari rumusan masalah, batasan masalah dalam penelitian, manfaat dari penelitian serta sistematika penulisan pada penelitian ini.

2. BAB II TINJAUAN PUSTAKA

Pada bagian tinjauan pustaka berisi penjelasan dasar teori yang berkaitan dengan topik penelitian. Selain itu juga terdapat penjelasan teori pada masing-masing variabel yang akan dalam klasifikasi mahasiswa *drop out*. Pada bagian ini dijelaskan mengenai pengertian *data mining*, klasifikasi, mahasiswa, *drop out*, semester, satuan kredit semester (SKS), Indeks Prestasi Semester (IPS), Indeks Prestasi Kumulatif (IPK), jalur masuk, Uang Kuliah Tunggal (UKT), normalisasi data, *Decision Tree C4.5*, KNN, *Backpropagation Ensemble Voting*, *Confusion Matrix* dan integrasi keislaman.

3. BAB III METODE PENELITIAN

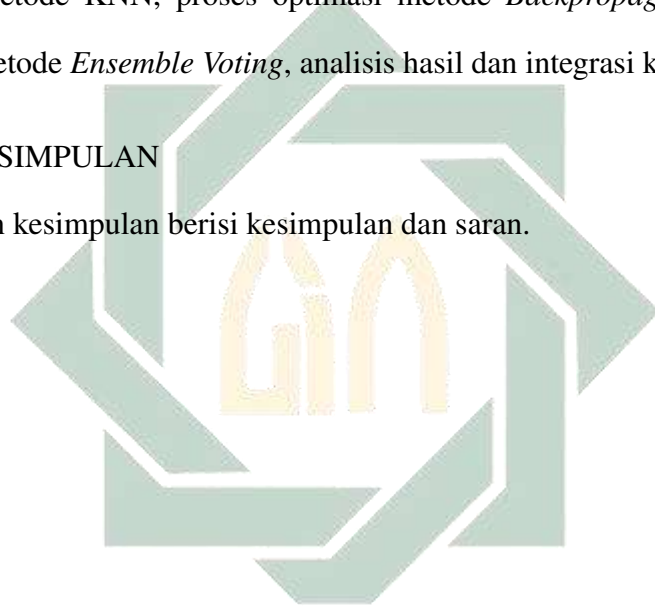
Pada bagian metode penelitian berisi jenis penelitian, sumber data penelitian, variabel data penelitian, tahapan penelitian serta tahapan metode penelitian.

4. BAB IV HASIL DAN PEMBAHASAN

Pada bagian hasil dan pembahasan berisi tentang data penelitian, proses normalisasi data, proses optimasi metode *Decision Tree C4.5*, proses optimasi metode KNN, proses optimasi metode *Backpropagation*, proses optimasi metode *Ensemble Voting*, analisis hasil dan integrasi keislaman.

5. BAB V KESIMPULAN

Pada bagian kesimpulan berisi kesimpulan dan saran.



UIN SUNAN AMPEL
S U R A B A Y A

BAB II

TINJAUAN PUSTAKA

2.1. *Data Mining*

Data mining disebut juga sebagai proses menambang data. *Data mining* dilakukan untuk mencari sebuah pola dari sebuah *database* (Sutoyo and Almaarif, 2020; Mughal et al., 2018) dengan menggunakan metode statistik, *artificial intelligence*, matematika dan *machine learning* (Dogan and Birant, 2021; Russell and Klassen, 2019). *Data mining* juga disebut sebagai *knowledge discovery* yaitu salah satu bidang pekerjaan yang berkembang pesat dikarenakan kebutuhan yang semakin besar (Derisma, 2020). *Data mining* berkaitan dengan data dengan volume yang besar, dalam *Gigabytes* atau *Terabytes* bahkan sebesar *Zetabytes* (Sumathi and Esakkirajan, 2007). Di dalam *data mining* terdapat beberapa pembagian diantaranya adalah prediksi, estimasi, asosiasi, klasterisasi dan klasifikasi (Saragih, 2018).

2.2. Klasifikasi

Klasifikasi mencoba mengelompokkan data baru dengan menggunakan variabel yang diteliti untuk membagi sesuatu menjadi kelompok atau jenis menurut karakteristik tertentu dengan mengenali, membedakan, dan memahami objek (Haut et al., 2018; Lizarti and Ulfah, 2019). Klasifikasi mencoba menemukan model untuk membedakan kelas dari suatu data supaya dapat digunakan dalam memperkirakan kelas dari objek data lain yang belum diketahui kelas atau labelnya (Bagli and Visani, 2020). Klasifikasi populer di bidang kesehatan dan

bioinformatika (Sannino and Pietro, 2018). Ada banyak jenis klasifikasi diantaranya adalah *text classification*, *document classification*, *image classification*, *text classification overview*, *document classification overview* (Mironczuk and Protasiewicz, 2018).

2.3. Mahasiswa

Mahasiswa adalah komponen yang cukup penting dalam dunia pendidikan. Mahasiswa diharapkan dapat mencapai potensi mereka di bidang akademik dan non-akademik (Sutoyo and Almaarif, 2020). Secara umum mahasiswa merupakan siswa yang sudah lulus dari SMA atau sederajat serta sedang menempuh pendidikan dan mencari ilmu di sebuah perguruan tinggi. Perguruan tinggi diantaranya adalah Sekolah tinggi, Akademi, Institut, Politeknik dan Universitas. Secara administratif mahasiswa dapat diartikan sebagai siswa yang terdaftar di sebuah perguruan tinggi.

2.4. Drop Out

Drop out adalah kondisi dimana mahasiswa dengan batas masa studi melebihi batas yang telah ditentukan tapi belum menyelesaikan beberapa kegiatan dalam akademik. (Muhamad et al., 2019; Ramayasa, 2018; Limsathitwong et al., 2018; Laufer and Gorup, 2019; Isphording and Qendrai, 2019). *Drop out* pada mahasiswa sudah diatur dalam UU Republik Indonesia No. 12 tahun 2012 tentang perguruan tinggi (Abdullah and Muhid, 2021). Dampak dari *drop out* tidak hanya dirasakan oleh mahasiswa yang mengalaminya, melainkan beberapa pihak yang terkait juga merasakan dampaknya salah satunya adalah perguruan tinggi (Abdullah and Muhid, 2021). Banyak faktor yang menyebabkan mahasiswa *drop out* dan dapat dipastikan setiap perguruan tinggi memiliki aturan yang berbeda (Casanova et al., 2018; Santos et al., 2020; Nagy and Molontay, 2018).

2.5. Semester

Semester menyatakan waktu mahasiswa berproses dalam kegiatan belajar-mengajar pada jenjang perkuliahan. Satuan waktu dalam satu semester adalah 6 bulan atau setengah tahun. Mahasiswa S1 biasanya menempuh perkuliahan selama waktu 4 tahun atau 8 semester (Irmayanti et al., 2021; Jasmir et al., 2018). Jika dalam waktu 8 semester belum menyelesaikan pendidikan maka kampus memberikan kesempatan hingga 14 semester. Jika dalam kurun waktu 7 tahun belum berhasil menyelesaikan pendidikan, maka akan diberlakukan *drop out*. Namun, beberapa kampus mungkin memiliki kebijakan tersendiri. Di UINSA sudah menerapkan batas maksimal mahasiswa menempuh studi yaitu dengan batasan 14 semester atau 7 tahun. Penelitian ini menggunakan data status semester pada mahasiswa sebagai variabel independen untuk klasifikasi mahasiswa *drop out*.

2.6. Satuan Kredit Semester (SKS)

SKS merupakan salah satu sistem pendidikan yang digunakan sebagai beban studi. Jumlah SKS yang ditempuh juga berbeda-beda di setiap program studi pada perguruan tinggi. Jumlah SKS yang ditentukan oleh institusi harus diselesaikan oleh setiap mahasiswa (Khoerulloh et al., 2019). Mahasiswa dinyatakan lulus apabila sudah menyelesaikan jumlah SKS yang sudah ditetapkan. Pada program sarjana 1 (S1) mewajibkan mahasiswanya untuk menempuh 140 - 160 SKS. Rentan waktu dalam satu SKS pada setiap perguruan tinggi berbeda-beda. Di UINSA rentan waktu dalam satu SKS adalah selama 50 menit. Penelitian ini menggunakan nilai SKS pada mahasiswa sebagai variabel independen untuk klasifikasi mahasiswa *drop out*.

2.7. Indeks Prestasi Semester (IPS) dan Indeks Prestasi Kumulatif (IPK)

Indeks prestasi mahasiswa dibedakan menjadi IPS dan IPK. IPS adalah rata-rata nilai yang didapatkan dari mata kuliah yang sudah dilalui dalam satu semester dan setiap semester akan mendapat nilai yang berbeda-beda (Sijabat et al., 2021). Sedangkan IPK merupakan suatu penilaian indeks prestasi dengan merata-rata nilai mata kuliah yang sudah ditempuh dari awal perkuliahan (Ario, 2020; Yudhistira, 2019). Nilai IPK menunjukkan peningkatan hasil belajar dari awal perkuliahan hingga akhir perkuliahan secara kumulatif (Purnawinadi, 2021). IPK didapatkan dari nilai dari semester awal sampai semester terakhir yang sudah ditempuh. Skala nilai IPS dan IPK di Indonesia adalah mulai dari 1 sampai 4. Penelitian ini menggunakan data nilai IPS dan nilai IPK pada mahasiswa sebagai variabel independen untuk klasifikasi mahasiswa *drop out*.

2.8. Jalur Masuk

Pada program Sarjana di UINSA ada beberapa jalur yang harus dilalui seorang mahasiswa baru agar menjadi mahasiswa di UIN Sunan Ampel Surabaya. Di UIN Sunan Ampel sendiri terdapat dua jenis program studi yaitu program studi umum dan program studi keislaman. Selain jalur mandiri dari kedua jenis program studi memiliki jalur masuk yang berbeda-beda untuk jalur rapor dan tes. Pada program studi keislaman terdapat jalur Seleksi Prestasi Akademik Nasional Perguruan Tinggi Keagamaan Islam Negeri (SPAN-PTKIN) yaitu jalur rapor (Suryani, 2021; Nalim et al., 2021) dan Ujian Masuk Perguruan Tinggi Keagamaan Islam Negeri (UM-PTKIN) sebagai jalur tes tulis (Jusriana et al., 2020; Evendi et al., 2019). Sedangkan pada program studi umum terdapat jalur Seleksi Nasional Masuk Perguruan Tinggi Negeri (SNMPTN) sebagai jalur rapor (Kristian Nazarius

et al., 2021) dan Seleksi Bersama Masuk Perguruan Tinggi Negeri (SBMPTN) sebagai jalur tes tulis (Widodo and Dewi, 2018; Idris, 2021). Pada jalur mandiri juga terdapat dua jenis yaitu mandiri reguler dan mandiri prestasi. Dari beberapa jalur masuk tersebut menentukan kualitas dan kuantitas dari mahasiswa yang mempengaruhi belajar mahasiswa. Dari beberapa jalur tersebut juga menentukan besar kecilnya nilai UKT mahasiswa tersebut. Nilai UKT sangat mempengaruhi nilai *drop out* mahasiswa.

2.9. Uang Kuliah Tunggal (UKT)

UKT adalah sebagian Biaya Kuliah Tunggal (BKT) yang merupakan seluruh biaya operasional mahasiswa pada setiap semester dan ditetapkan berdasarkan BKT dikurangi dengan biaya yang ditanggung oleh pemerintah (Novianti, 2019). UKT mahasiswa mempunyai kelompok yang berbeda-beda, dimana setiap kelompok mempunyai persentase masing-masing yang sudah ditetapkan oleh pemerintah berdasarkan ekonomi mahasiswa (Wawan Firgiawan et al., 2019). Biaya UKT diasumsikan merupakan salah satu faktor yang mempengaruhi mahasiswa *drop out*. Di UIN Sunan Ampel Sendiri terdapat 7 golongan atau kelompok UKT dari UKT terkecil Rp 150.000 dan hingga yang paling besar mendekati Rp 10.000.000.

2.10. Normalisasi

Pada klasifikasi mahasiswa *drop out* ini peneliti menggunakan data semester, jumlah SKS, nilai IPK, nilai IPS, jalur masuk dan nominal UKT. Data-data tersebut memiliki *range* yang berbeda antar variabel (Raharjo et al., 2020). Peneliti menggunakan rumus normalisasi *min-max* untuk mengubah skala data. Data dengan *range* yang berbeda tersebut akan diubah ke dalam *range* 1

sampai 0 dengan menggunakan rumus dari normalisasi *min-max* sebagai berikut :

$$\text{Norm}(x) = \frac{x - \min}{\max - \min} \quad (2.1)$$

Dengan:

x = data yang akan dinormalisasi

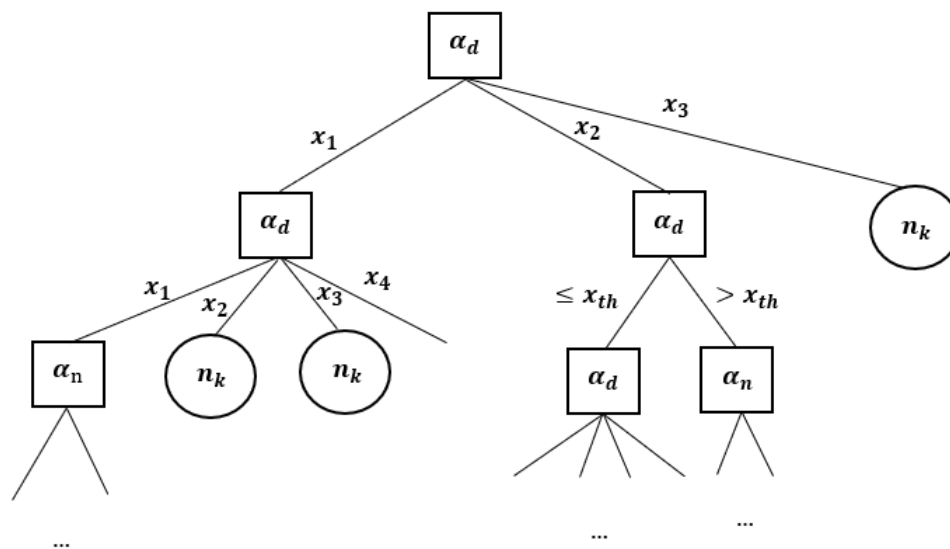
min = nilai minimum dari tiap atribut

max = nilai maksimum dari tiap atribut

2.11. *Decision Tree C4.5*

Selain metode C4.5 ada beberapa metode lain dari *Decision Tree* diantaranya adalah *Iterative Dichotomizer 3* (ID3), *Classification And Regression Tree* (CART), *Conditional Inference Trees* (CIT) dan *CHi-squared Automatic Interaction Detector* (CHAID) (Cherfi et al., 2018). Metode C4.5 merupakan salah satu metode dengan teknik *Decision Tree* (Damanik et al., 2019). Metode ini menerapkan struktur pohon yaitu setiap *node* yang ada mewakili atribut dan setiap cabang juga mewakili nilai atribut lagi. Sedangkan daun digunakan untuk mewakili kelas atau label. *Node* teratas dari pohon keputusan adalah root (Sulistiani and Aldino, 2020). *Input* dari *Decision Tree C4.5* ini adalah himpunan data yang didalamnya terdapat kumpulan sampel data serta kandidat dari atribut yang terdiri dari minimal atribut independen dan dependen. Atribut independen bisa bertipe numerik maupun diskrit, sedangkan atribut dependen harus bertipe diskrit. Jika atribut pada independen adalah dalam bentuk diskrit (a_d), maka cabang dari simpul akan dibuat untuk setiap nilai pada atribut diskrit ini (x_1, x_2, \dots, x_m). Namun jika atribut independen bertipe numerik a_n , maka cabang simpul akan dibuat untuk dua buah nilai yaitu $a_n \leq x_{th}$ dan $a_n > x_{th}$. Dimana x_{th}

merupakan nilai yang didapat dari nilai-nilai a_n . Nilai pada atribut label direpresentasikan oleh daun. Arsitektur *Decision Tree C4.5* dapat digambarkan seperti berikut:



Gambar 2.1 Model *Decision Tree C4.5* (Damanik et al., 2019)

Decision Tree C4.5 adalah model klasifikasi yang menggunakan struktur pohon. *Output* dari model ini adalah pohon keputusan. Dimana *rule* atau aturan yang digunakan didapatkan dari pohon keputusan tersebut. Aturan yang didapatkan akan digunakan untuk melakukan prediksi pada data baru. Beberapa tahapan dalam *Decision Tree C4.5* adalah:

1. Memilih akar *node* dari atribut

Dalam proses memilih *node* digunakanlah *gain ratio*. *Gain ratio* akan dicari pada masing-masing atribut dan dipilih yg paling tinggi. Berikut adalah rumus dalam menghitung *gain ratio*:

$$\text{Gain Ratio}(S, A) = \frac{\text{Gain}(S, A)}{\text{Split Info}(S, A)} \quad (2.2)$$

Dengan:

S = himpunan data *training*

A = atribut data *training*

Setelah mendapatkan *gain ratio*, nilai dari *split info* dan *gain* dari semua atribut akan dihitung menggunakan rumus yang ditentukan. Rumus untuk mendapatkan nilai *split info* adalah sebagai berikut:

$$\text{Split Info}(S, A) = - \sum_{i=1}^n \left| \frac{S_i}{S} \right| * \log_2 \left(\left| \frac{S_i}{S} \right| \right) \quad (2.3)$$

Dengan:

i = index pada S ($i = 1, 2, 3, \dots, n$)

n = banyaknya partisi S

S_i = menyatakan sub himpunan ke i pada sampel S .

$|S_i|$ = jumlah kasus pada partisi ke- i

$|S|$ = jumlah kasus dalam S

Setelah mendapatkan nilai dari *split info*, nilai *gain* akan dicari menggunakan rumus yang ditentukan. *Gain* merupakan ukuran keefektifan dari sebuah atribut. Dalam menghitung nilai *gain* digunakan rumus:

$$\text{Gain}(S, A) = \text{Entropy}(S) - \sum_{x \in \text{Values}(A)} \left| \frac{S_x}{S} \right| * \text{Entropy}(S_x) \quad (2.4)$$

Dengan

$\text{Values}(A)$ = Himpunan yang mungkin untuk atribut A

x = Suatu nilai yang mungkin untuk atribut A

$|S_x|$ = Jumlah sampel untuk nilai x

$\text{Entropy}(S_x)$ = Entropy untuk sampel-sampel yang memiliki nilai x

Entropy digunakan untuk memperkirakan rata-rata jumlah bit per simbol.

Nilai tersebut akan digunakan untuk mentransmisikan data. Rumus menghitung *entropy* adalah:

$$\text{Entropy}(S) = - \sum_{i=1}^n P_i * \log_2 P_i \quad (2.5)$$

Dengan:

P_i = adalah rasio dari kelas C_i di dalam S

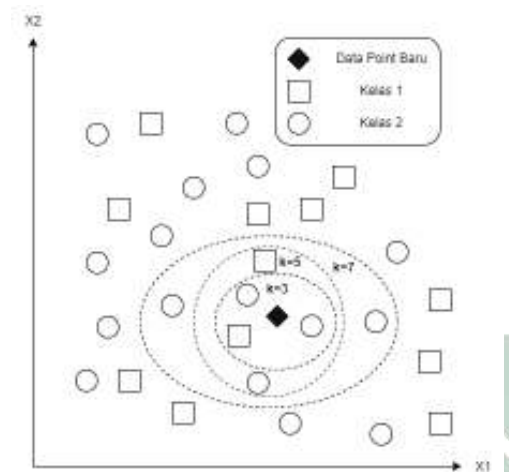
$\text{Entropy}(S)$ adalah *entropy* dari data *training* sedangkan $\text{Entropy}(S_i)$ merupakan *entropy* pada sampel-sampel yang memiliki nilai i .

2. Setelah terpilih atribut yang akan digunakan sebagai akar, langkah selanjutnya adalah membuat cabang pada setiap nilai dari atribut
3. Langkah selanjutnya adalah membagi kasus di dalam cabang dengan tujuan menemukan *node* selanjutnya
4. Proses pada masing-masing cabang tersebut akan diulang terus-menerus hingga semua kasus pada cabang mendapatkan nilai label yg sama. Pada akhirnya semua cabang akan menghasilkan kelas atau label

2.12. *K-Nearest Neighbors* (KNN)

KNN merupakan metode klasifikasi sederhana dengan menggunakan prinsip dasar perhitungan jarak terdekat (Zhang et al., 2018; Triguero et al., 2019). KNN termasuk *supervised learning*, dimana data yang akan digunakan sudah mempunyai kelas pada variabel dependen. Pada tahap *training* KNN hanya menyimpan klasifikasi data *training* dan vektor dari fitur pada data (Mustaqim et al., 2020). Untuk mengukur ketetanggaan dalam klasifikasi KNN menggunakan *distance function or similarity matrix*. KNN mengklasifikasi kelas mayoritas

ketetanggaan terdekat. KNN memberikan bobot yang sama untuk setiap atribut (Mohanapriya and Lekha, 2018). Keakuratan KNN sangat dipengaruhi oleh pemilihan jumlah k . Berikut adalah gambaran cara kerja dari metode KNN:



Gambar 2.2 Klasifikasi K - *Nearest Neighbor* (Zhang et al., 2018)

Penentuan nilai k dalam KNN bisa didapatkan dengan nilai sampel k terdekat k_1, k_2, \dots, k_s . Jika data yang ada semakin banyak maka jumlah k yang dipilih akan semakin kecil, sedangkan jika dimensi data yang ada lebih besar maka jumlah k yang dipilih harus lebih tinggi. Dalam menentukan nilai k yang akan di uji coba lebih baik menggunakan angka ganjil seperti $k = 1, 3, 5, \dots, n$ (Dinata et al., 2020). Jika k terlalu kecil, sensitif terhadap titik noise dan jika terlalu besar, dapat menyebabkan bias model. Beberapa metode digunakan untuk mengatasi masalah seperti itu, tetapi metode tersebut kurang efisien. Nilai k yang digunakan tidak boleh lebih besar atau sama dengan N , dalam menentukan label digunakan rumus jarak. Misal terdapat 2 titik yaitu P_1 dan P_2 sebagai berikut:



Gambar 2.3 Jarak Dua Titik

Maka ada banyak cara untuk menghitung jarak kedua titik diantaranya adalah *Euclidean*, *Manhattan*, *Cosine*, *Correlation*, *Hamming* dan *Minkowski*. Dalam KNN akan dicari jarak antara dua tetangga k terdekat dengan menggunakan rumus *Euclidean*. Adapun rumus dari penghitungan jarak *Euclidean* adalah sebagai berikut:

$$\text{Dist}(x_i, y_i) = \sqrt{\sum_{i=1}^n (x_i - y_i)^2} \quad (2.6)$$

Keterangan:

$\text{Dist}(x_i, y_i)$ = jarak antar objek (*Euclidean Distancing*) ke- i

i = index pada data ($i = 1, 2, 3, \dots, n$)

n = banyaknya data

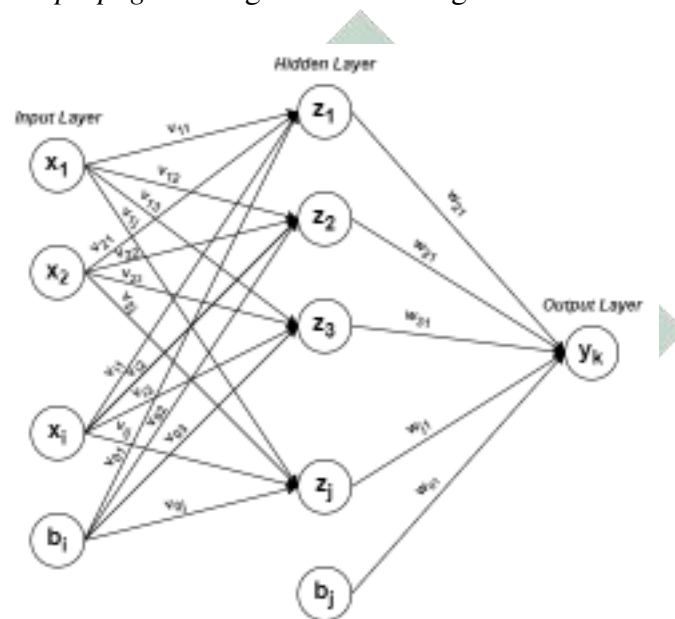
x_i = sampel data

y_i = data uji

2.13. Backpropagation

Backpropagation banyak diaplikasikan pada banyak bidang seperti pengenalan pola, peramalan, pemilihan lokasi, dan evaluasi pekerjaan (Purba et al., 2019). *Backpropagation* memiliki konsep seperti jaringan saraf manusia (Vlachas et al., 2020; Febriadi et al., 2018). Di dalam jaringan saraf manusia meliputi dendrit, akson, badan sel dan sinapsis. Jaringan dendrit digunakan sebagai penerima dan penyalur rangsangan ke jaringan badan sel, dalam *Backpropagation* variabel *input* dianggap sebagai jaringan dendrit dikarenakan memiliki konsep

yang sama. Akson digunakan untuk mengirimkan impuls yang berasal dari badan sel ke jaringan lainnya, dalam *Backpropagation* variabel *output* dianggap sebagai jaringan akson karena memiliki konsep yang sama. Badan sel digunakan sebagai penerima dan penyalur rangsangan dari jaringan dendrit menuju ke jaringan akson, dalam *Backpropagation* fungsi aktivasi $f(x)$ memiliki konsep yang sama dengan badan sel. Sedangkan jaringan yang menghubungkan dua sel saraf adalah sinapsis. Arsitektur dari *Backpropagation* digambarkan sebagai berikut:



Gambar 2.4 Arsitektur *Backpropagation* (Pujianto et al., 2018)

Terdapat 3 layer yang ada dalam *Backpropagation* yaitu *input layer*, *hidden layer*, dan *output layer*. Arsitektur dari *Backpropagation* bisa diganti dengan memberi tambahan *hidden layer* serta jumlah *node* yang ada pada setiap *layer*. Terdapat 3 tahap perhitungan yang ada di *Backpropagation* (Sacramento et al., 2018). Tahap pertama adalah *feed forward*, pada tahap *feed forward* dilakukan perhitungan data *input* dengan menambahkan nilai bobot yang ditentukan dan mendapatkan hasil akhir berupa *output*. Setelah tahap *feed forward* tahap berikutnya adalah tahap *backward*, pada tahap *backward* dilakukan perhitungan pada selisih antara *output* dan target atau biasa disebut sebagai perhitungan *error*.

Setelah melewati tahap *backward* maka akan memasuki tahap selanjutnya yaitu melakukan *update* bobot yang didasarkan pada hasil *error* dari tahap *backward*. Fungsi aktivasi sigmoid digunakan untuk mengaktifkan neuron-neuron pada proses *feed forward*. Adapun rumus fungsi aktivasi *sigmoid* adalah sebagai berikut:

$$f(x) = \frac{1}{1 + e^{-x}} \quad (2.7)$$

Perhitungan *Backpropagation* dimulai dengan melakukan inisialisasi bobot. Inisialisasi ini mengambil nilai acak sebagai bobot awal. Nilai yang diambil adalah nilai yang relatif kecil. Setelah menginisialisasi nilai bobot, langkah-langkah berikutnya adalah:

1. Tahap *feed forward* dilakukan pada setiap pasangan elemen *training*. Berikut adalah tahapan *feed forward*:
 - (a) Sinyal yang berasal dari x_i akan diterima oleh tiap-tiap *input layer* ($x_i, i = 1, 2, 3, \dots, n$). Sinyal tersebut akan diteruskan ke *hidden layer*
 - (b) Sinyal *input* terbobot akan dijumlahkan oleh setiap *hidden layer* ($z_j, j = 1, 2, 3, \dots, n$) menggunakan rumus berikut:

$$z_{inj} = v_{0j} + \sum_{i=1}^n x_i v_{ij} \quad (2.8)$$

sinyal *output* akan dihitung menggunakan fungsi aktivasi berikut:

$$z_j = f(z_{inj}) \quad (2.9)$$

sinyal yang didapatkan akan dikirim ke semua *output layer*.

- (c) Setiap sinyal input terbobot akan dihitung oleh setiap *output layer* ($y_k, k = 1, 2, 3, \dots, m$) menggunakan :

$$y_{ink} = w_{0k} + \sum_{i=1}^p z_i w_{jk} \quad (2.10)$$

sinyal *output* akan dihitung menggunakan fungsi aktivasi berikut:

$$y_k = f(y_{ink}) \quad (2.11)$$

sinyal yang didapatkan akan dikirimkan ke semua *output layer*.

2. Setelah melalui tahap *feed forward*, selanjutnya akan dilakukan tahap *backward*. Berikut adalah tahapan dari *backward*:

- (a) Target pada pola yang mempunyai hubungan dengan pola *input* akan diterima oleh setiap *output layer* ($y_k, k = 1, 2, 3, \dots, m$). Tahap selanjutnya adalah hitung nilai *error* yang ada menggunakan:

$$\delta_k = (t_k - y_k) f'(y_k) \quad (2.12)$$

Dimana:

t_k = merupakan target pada node ke-k

y_k = hasil diagnosis pada output layer node ke-k

$f'(y_k)$ = turunan dari fungsi aktivasi pada *output layer*

koreksi bobot akan dihitung menggunakan:

$$\Delta w_{jk} = \alpha \delta_k z_j \quad (2.13)$$

selanjutnya adalah memperbaiki nilai w_{jk} menggunakan koreksi bobot

yang didapatkan. Hitung koreksi bias menggunakan:

$$\Delta w_{0k} = \alpha \delta_k \quad (2.14)$$

Dengan:

α = *learning rate*

δ_k = nilai kesalahan pada *output layer node* ke-k

z_j = *output* pada *hidden layer node* ke-j

nilai w_{0k} akan diperbaiki menggunakan nilai dari koreksi bias yang sudah didapatkan. Langkah selanjutnya adalah mengirimkan δ_k ke *layer* yang ada di lapisan bawahnya.

- (b) Delta input yang didapatkan akan dijumlahkan oleh setiap *hidden layer* ($Z_j, j = 1, 2, 3, \dots, p$) menggunakan rumus:

$$\delta_{inj} = \sum_{j=1}^p \delta_k w_{jk} \quad (2.15)$$

Dengan:

w_{jk} = bobot *node hidden layer* ke-j *node output* ke-k

jumlah tersebut akan dikalikan dengan turunan dari fungsi aktivitasnya untuk menghitung informasi *error*:

$$\delta_j = \delta_{inj} f'(z_j) \quad (2.16)$$

Dengan:

δ_{inj} = nilai kesalahan pada *hidden layer node* ke-j

$f'(z_j)$ = turunan dari fungsi aktivasi pada *hidden layer*

koreksi bobot akan dihitung menggunakan rumus:

$$\Delta v_{ij} = \alpha \delta_j x_i \quad (2.17)$$

nilai v_{ij} akan diperbaiki menggunakan nilai koreksi bobot yang sudah didapatkan. Langkah selanjutnya adalah perhitungan koreksi bias menggunakan rumus berikut:

$$\Delta v_{0j} = \alpha \delta_j \quad (2.18)$$

nilai v_{0j} akan diperbaiki menggunakan nilai dari koreksi bias yang sudah didapatkan.

- (c) Bias dan bobot dari *output layer* ($j = 0, 1, 2, \dots, p$) akan diperbaiki oleh setiap *output layer* ($y_k, k = 1, 2, 3, \dots, m$) menggunakan rumus:

$$w_{jk}(\text{baru}) = w_{jk}(\text{lama}) + \Delta w_{jk} \quad (2.19)$$

bias dan bobot *hidden layer* ($i = 0, 1, 2, \dots, n$) akan diperbaiki oleh setiap *hidden layer* ($Z_j, j = 1, 2, 3, \dots, n$) menggunakan rumus:

$$v_{ij}(\text{baru}) = v_{ij}(\text{lama}) + \Delta v_{ij} \quad (2.20)$$

- Memasuki tahap yang ketiga yaitu melakukan *update* bobot berdasarkan *error*. Pada tahap ini akan dilakukan tes kondisi berhenti. Tes kondisi berhenti dilakukan jika terdapat 2 kondisi yaitu jika kondisi *error* di bawah target *error* dan iterasi melebihi maksimum epoch (iterasi).

2.14. Ensemble Voting

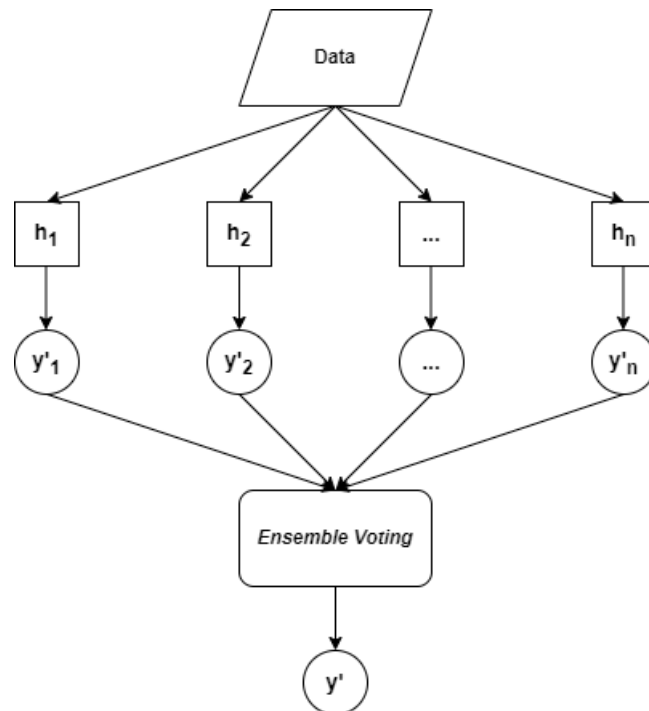
Ensemble Voting adalah salah satu jenis dari *Ensemble Learning*. *Ensemble Learning* adalah metode yang cara kerjanya adalah menggabungkan beberapa *machine learning* dengan harapan mendapatkan hasil yang lebih akurat (Rad and Behjat, 2019). Beberapa model *machine learning* digabungkan dalam *ensemble Learning* dan mendapatkan *output* satu model yang lebih baik (Pawlovsky, 2018). Hubungan linier dan non-linier dapat ditangkap oleh *Ensemble Learning* (Patacsil, 2020). *Ensemble Learning* adalah teknik agregasi keluaran model yang telah berkembang selama satu setengah dekade terakhir di bidang statistik dan *machine learning* (Tattar, 2018). *Ensemble Learning* digunakan untuk menggabungkan atau mengkombinasikan beberapa metode *machine learning* tersebut untuk mendapat hasil yang maksimal. Ada banyak jenis *Ensemble* diantaranya adalah *Bagging Ensemble*, *Random Forest Ensemble*, *AdaBoost Ensemble*, *Gradient Boosting Ensemble*, *Voting Ensemble* dan *Stacking Ensemble* (Brownlee, 2021). (Kumar and Jain, 2020) berpendapat bahwa ada beberapa pembagian dalam *Ensemble Learning*, yaitu *Ensemble* pada data yang didalamnya terdapat metode *Ensemble Bagging (Bootstrap Aggregating)*, *Ensemble* pada metode yang didalamnya terdapat *Ensemble Voting*, *Ensemble* kombinasi data *training* yang didalamnya mencakup *Ensemble Boosting*, *Ensemble AdaBoost*, *Ensemble Gradient Boosting*, *Ensemble XGBoost* dan *Ensemble Stacking*. Secara umum ada 3 pembagian pada *Ensemble learning*, (Kumar and Jain, 2020) yaitu:

1. *Ensemble* pada data *training*, yang didalamnya membagi data *training* menjadi beberapa bagian dan melatih metode klasifikasi pada setiap subset data pelatihan yang dibagi sebelumnya. Dengan menggabungkan *output* dari yang beberapa model diharapkan *Ensemble* mendapatkan nilai akurasi yang

lebih baik dibandingkan dengan melatih satu model pada seluruh data pelatihan. Pembagian data pelatihan semacam ini disebut sebagai metode *Ensemble bagging*.

2. *Ensemble* kombinasi campuran, *Ensemble* ini menumpuk satu lapisan *machine learning* di atas *machine learning* lainnya. Kedua pendekatan ini meningkatkan dan melibatkan pencampuran metode *machine learning* dalam berbagai kombinasi. *Ensemble* kombinasi ini mencakup *Ensemble Boosting*, *Ensemble AdaBoost*, *Ensemble Gradient Boosting*, *Ensemble XGBoost* dan *Ensemble Stacking*.
3. *Ensemble* pencampuran model, *Ensemble* ini menggabungkan dan melatih beberapa model dengan menggunakan beberapa pelatihan berjalan dengan pengaturan *hyperparameter* yang berbeda. Jenis metode *ensemble* yang menggabungkan berbagai model dibahas dalam *Ensemble* pada model metode yang didalamnya terdapat *Ensemble Voting*. Dalam *Ensemble Voting* sendiri dibagi menjadi beberapa bagian diantaranya adalah *Ensemble Soft Voting* dan *Ensemble Hard Voting*.

Penelitian ini hanya akan menggunakan *Ensemble Voting* dari banyak jenis *Ensemble Learning* yang dapat digunakan. Berikut adalah gambaran secara umum bagaimana cara kerja dari *Ensemble Voting*:



Gambar 2.5 Ensemble Voting (Assiri et al., 2020)

Dari gambar di atas dapat dilihat bahwa proses *Ensemble Voting* dapat dijalankan setelah mendapat beberapa hasil dari model tunggal. Hasil dari beberapa model tunggal tersebut akan diolah dalam *Ensemble Voting*. Pada penelitian ini, peneliti mencoba menggunakan *Ensemble Hard Voting* dan *Ensemble Soft Voting* (GitHub, 2022):

1. *Ensemble Hard Voting*

Label pada variabel respon didapatkan dari nilai modus pada setiap pengklasifikasi h_j . Berikut adalah rumus yang digunakan pada *Ensemble Hard Voting*:

$$\hat{y} = \text{mode}\{h_1(x), h_2(x), \dots, h_n(x)\} \quad (2.21)$$

Dengan asumsi jika terdapat 3 model dengan hasil masing-masing adalah 0, 0 dan 1. Berikut adalah contoh perhitungan menggunakan persamaan 2.21

pada 3 model tersebut:

$$\hat{y} = \text{mode}\{0, 0, 1\} = 0$$

Hasil tersebut didapatkan dari modus dan akan mengklasifikasikan sampel sebagai "0".

2. Ensemble Soft Voting

Label pada variabel respon didapatkan berdasarkan probabilitas yang diprediksi p untuk pengklasifikasi menggunakan rumus sebagai berikut:

$$\hat{y} = \arg \max_i \sum_{j=1}^m w_j p_{ij} \quad (2.22)$$

di mana w_j adalah bobot yang dapat diberikan ke pengklasifikasi ke- j .

Dengan asumsi contoh di bagian sebelumnya adalah tugas klasifikasi biner dengan label kelas $i \in \{0, 1\}$, dan diasumsikan pada 3 model tunggal sebelumnya memiliki nilai probabilitas sebagai berikut:

Model 1 $\Rightarrow [0.9, 0.1]$

Model 2 $\Rightarrow [0.8, 0.2]$

Model 3 $\Rightarrow [0.4, 0.6]$

berikut adalah contoh perhitungan menggunakan *Ensemble Soft Voting*:

(a) Dengan menggunakan bobot (1,1,1):

$$\text{probabilitas kelas (0)} = \frac{0.9 + 0.8 + 0.4}{3} = 0.7$$

$$\text{probabilitas kelas (1)} = \frac{0.1 + 0.2 + 0.6}{3} = 0.3$$

$$\hat{y} = \arg \max[\text{probabilitas kelas (0), probabilitas kelas (1)}]$$

$$\hat{y} = \arg \max[0.7, 0.3] = 0$$

(b) Prediksi akan menghasilkan $\hat{y} = 1$ dengan menggunakan nilai bobot $\{0.1, 0.1, 0.8\}$ dengan perhitungan sebagai berikut:

$$\text{probabilitas kelas (0)} = 0.1 \times 0.9 + 0.1 \times 0.8 + 0.8 \times 0.4 = 0.49$$

$$\text{probabilitas kelas (1)} = 0.1 \times 0.1 + 0.1 \times 0.2 + 0.8 \times 0.6 = 0.51$$

$$\hat{y} = \arg \max[\text{probabilitas kelas (0)}, \text{probabilitas kelas (1)}] = 1$$

Terlihat bahwa nilai bobot sangat mempengaruhi hasil dari perhitungan *Ensemble Soft Voting*.

2.15. Confusion Matrix

Confusion matrix adalah proses memprediksi nilai benar dan salah suatu model pada label data uji (Yuliana and Pratomo, 2017). Nilai sebenarnya dan nilai prediksi pada klasifikasi akan dihasilkan dalam proses *confusion matrix* (Irmayanti et al., 2021). Terdapat 4 bagian yang ada pada tabel *confusion matrix* yang berisi nilai *true positive* (TP), *false positive* (FP), *true negative* (TN), dan *false negative* (FN):

1. *True Positive* (TP) adalah banyaknya sampel dengan nilai positif dan diprediksi sebagai positif.
2. *False Positive* (FP) adalah banyaknya sampel dengan nilai negatif dan diprediksi sebagai positif.
3. *False Negative* (FN) adalah banyaknya sampel dengan nilai positif dan diprediksi sebagai negatif.

4. *True Negative* (TN) adalah banyaknya sampel dengan nilai negatif dan diprediksi sebagai negatif.

Berikut adalah tabel dari *confusion matrix*:

Tabel 2.1 Confusion Matrix

	Prediksi Benar	Prediksi Salah
Aktual Benar	TP	FN
Aktual Salah	FP	TN

Nilai yang didapatkan dari proses *confusion matrix* adalah nilai akurasi, presisi dan *recall*. Terdapat fungsi yang berbeda-beda dari nilai akurasi, presisi dan *recall*. Adapun rumus untuk menghitungnya adalah sebagai berikut (Sudiyarno et al., 2021):

$$\text{Akurasi} = \frac{TP + TN}{TP + TN + FP + FN} \times 100\% \quad (2.23)$$

$$\text{Presisi} = \frac{TP}{TP + FP} \times 100\% \quad (2.24)$$

$$\text{Recall} = \frac{TP}{TP + FN} \times 100\% \quad (2.25)$$

Tingkat akurasi *confusion matrix* dibagi menjadi:

Tabel 2.2 Arti Nilai Confusion Matrix (Dinata et al., 2020)

Nilai <i>confusion matrix</i>	Arti nilai
90% - 100 %	Klasifikasi sangat baik
80% - 90%	Klasifikasi baik
70% - 80%	Klasifikasi cukup
60% - 70%	Klasifikasi buruk
0% - 60%	Klasifikasi salah

2.16. Integrasi Keislaman

2.16.1. Bermanfaat dalam Perspektif Islam

Data mining berusaha mencari pola yang berasal dari dalam *database*. Pola tersebut yang nantinya akan dianalisis dan diolah menjadi sebuah informasi yang bermanfaat. *Data mining* dapat membantu orang dengan memberikan informasi yang dibutuhkan. Anjuran berbuat baik dan membantu orang lain terdapat dalam al-Qur'an surat Al-isra' ayat 7 sebagai berikut:

إِنْ أَحْسَنْتُمْ أَحْسَنْتُمْ لِأَنْفُسِكُمْ وَإِنْ أَسَأْتُمْ فَلَهَا فَإِذَا جَاءَ وَعْدُ الْآخِرَةِ لِيَسْئُوا وُجُوهَكُمْ
وَلِيَدْخُلُوا الْمَسْجِدَ كَمَا دَخَلُوهُ أَوَّلَ مَرَّةٍ وَلِيُتَبِّرُوا مَا عَلَوْا تَتْبِيرٌ

Artinya: "Jika berbuat baik, (berarti) kamu telah berbuat baik untuk dirimu sendiri. Jika kamu berbuat jahat, (kerugian dari kejahatan) itu kembali kepada dirimu sendiri. Apabila datang saat (kerusakan) yang kedua, (Kami bangkitkan musuhmu) untuk menyuramkan wajahmu, untuk memasuki masjid (Baitul Maqdis) sebagaimana memasukinya ketika pertama kali, dan untuk membinasakan apa saja yang mereka kuasai."

Dari beberapa dalil mengenai pentingnya menjadi orang yang bermanfaat mendorong peneliti untuk mencari informasi yang bermanfaat dari *database*. Peneliti mencoba mencari informasi klasifikasi mahasiswa yang berpotensi drop out dari *database* yang ada di UINSA. Dalam hadits riwayat Ibnu Majah nomor 2673, Rasulullah Shallallahu Alaihi Wassalam bersabda:

حَدَّثَنَا مُحَمَّدُ بْنُ عَبْدِ الْأَعْلَى الصَّنَعَائِيُّ حَدَّثَنَا الْمُعْتَمِرُ بْنُ سُلَيْمَانَ عَنْ أَبِيهِ عَنْ حَدِيثٍ
عَنْ عِكْرِمَةَ عَنْ ابْنِ عَبَّاسٍ عَنِ النَّبِيِّ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ قَالَ الْمُسْلِمُونَ تَتَكَافَأُونَ
دِمَاؤَهُمْ وَهُمْ يَدُّ عَلَى مَنْ سِوَاهُمْ يَسْعَى بِدِمَتِهِمْ أَذْنَاهُمْ وَيَرُدُّ عَلَى أَفْصَاهُمْ

Artinya: *"Telah menceritakan kepada kami Muhammad bin Abdul 'A'la Ashan'ani, telah menceritakan kepada kami Al Mu'tamir bin Sulaiman dari Ayahnya dari Hanasy dari Ikrimah dari Ibnu Abbas, dari Nabi shallallahu 'alaihi wasallam, beliau bersabda: "Nyawa umat Islam sepadan. seorang muslim harus saling membantu sesama mereka, yang lemah dan yang jauh mempunyai hak yang sama dalam perihal ghanimah.""*

Hadits di atas menjelaskan bahwa tidak ada yang diunggulkan dari seluruh umat muslim dan Allah menganggap semuanya sepadan. Sesama muslim harus saling membantu urusan saudaranya dikarenakan memiliki hak yang sama.

2.16.2. Menuntut Ilmu dalam Perspektif Islam

Menjadi mahasiswa adalah salah satu bukti bahwa seseorang sedang mencari ilmu. Menuntut ilmu adalah usaha mempelajari ilmu untuk kepentingan di dunia dan ilmu untuk akhirat. Perintah untuk menuntut ilmu juga terdapat dalam al-Qur'an surat Al-Ankabut Ayat 43 sebagai berikut:

وَتِلْكَ الْأَمْثَالُ نَضْرِبُهَا لِلنَّاسِ وَمَا يَعْقِلُهَا إِلَّا الْعَالِمُونَ

Artinya: *"Perumpamaan-perumpamaan itu Kami buat untuk manusia. Namun, tidak ada yang memahaminya, kecuali orang-orang yang berilmu."*

Ayat di atas menjelaskan bahwa hanya orang-orang yang berilmu yang bisa memahami beberapa hal yang dilakukan Allah SWT kepada siapa saja yang menyembah selain kepada Allah SWT. Dengan adanya kewajiban menuntut ilmu,

umat muslim harus mencari ilmu sebanyak-banyaknya. Ilmu yang didapatkan harus bermanfaat bagi orang lain supaya ilmu yang didapatkan menjadi amalan yang tidak akan terputus. Dalam hadits riwayat Tirmidzi nomor 2571 juga dijelaskan mengenai anjuran menuntut ilmu bagi seorang muslim:

حَدَّثَنَا نَصْرُ بْنُ عَلِيٍّ قَالَ حَدَّثَنَا خَالِدُ بْنُ يَزِيدَ الْعَتَكِيُّ عَنْ أَبِي جَعْفَرٍ الرَّازِيِّ عَنْ الرَّبِيعِ بْنِ أَنَسٍ عَنْ أَنَسِ بْنِ مَالِكٍ قَالَ قَالَ رَسُولُ اللَّهِ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ مَنْ خَرَجَ فِي طَلَبِ الْعِلْمِ كَانَ فِي سَبِيلِ اللَّهِ حَتَّى يَرْجِعَ قَالَ أَبُو عِيسَى هَذَا حَدِيثٌ حَسَنٌ غَرِيبٌ وَرَوَاهُ بَعْضُهُمْ فَلَمْ يَرْفَعْهُ

Artinya: "Telah bercerita kepada kami Nahsr bin Ali dia berkata, telah bercerita kepada kami Khalid bin Yazid Al Ataki dari Abu Ja'far Ar Razi dari Ar Rabi' bin Anas dari Anas bin Malik dia berkata; Rasulullah Shallallahu Alaihi Wassalam bersabda: "Barangsiapa keluar dalam rangka menuntut ilmu maka dia berada di jalan Allah sampai dia kembali." Abu Isa berkata; 'Hadits ini hasan gharib, sebagian perawi telah meriwayatkannya namun tidak merafa'kannya."

2.16.3. Mengulang Perkuliahan dalam Perspektif Islam

Dalam proses menuntut ilmu mahasiswa akan mengalami banyak kendala dalam menerima materi. Mahasiswa yang kesulitan pada beberapa bidang atau mata kuliah tertentu akan mendapat nilai yang rendah. Jika mahasiswa mendapat nilai mata kuliah yang terlalu dibawah rata-rata, maka harus mengambil kembali mata kuliah yang memiliki nilai rendah pada semester depan dengan menambah SKS. Saat menjalani mata kuliah mengulang mahasiswa akan mengalami kerugian waktu karena harus mengulang materi yang sudah diikuti pada semester sebelumnya. Islam mengajarkan untuk memaksimalkan waktu sebaik mungkin

salah satunya dijelaskan dalam al-Qur'an surat Al-ashr ayat 1-3 berikut:

وَالْعَصْرِ ۝
إِنَّ الْإِنْسَانَ لِفِي خُسْرٍ ۝
إِلَّا الَّذِينَ آمَنُوا وَعَمِلُوا الصَّالِحَاتِ وَتَوَّصُوا بِالحَقِّ ۝
وَتَوَّصُوا بِالصَّبْرِ ۝

Ayat di atas adalah landasan yang ideal mengenai pentingnya menghargai waktu (Ritonga, 2018). Terdapat beberapa hadits mengenai anjuran memanfaatkan waktu sebaik mungkin diantaranya adalah pada hadits riwayat Bukhari nomor 5937, Rasulullah *Shallallahu Alaihi Wassalam* bersabda:

فَقَالَ كُنْ فِي الدُّنْيَا كَأَنَّكَ غَرِيبٌ أَوْ عَابِرُ سَبِيلٍ وَكَانَ ابْنُ عُمَرَ يَقُولُ إِذَا أَمْسَيْتَ فَلَا تَنْتَظِرَ الصَّبَاحَ وَإِذَا أَصْبَحْتَ فَلَا تَنْتَظِرَ الْمَسَاءَ وَخُذْ مِنْ صِحَّتِكَ لِمَرْضِكَ وَمِنْ حَيَاتِكَ لِمَوْتِكَ

Yang artinya: "Jadilah kamu di dunia ini seakan-akan orang asing atau seorang pengembara." Ibnu Umar juga berkata; 'Bila kamu berada di sore hari, maka janganlah kamu menunggu datangnya waktu pagi, dan bila kamu berada di pagi hari, maka janganlah menunggu waktu sore, pergunakanlah waktu sehatmu sebelum sakitmu, dan hidupmu sebelum matimu.'"

Waktu merupakan sesuatu yang berharga. Seorang muslim harus memanfaatkan waktu sebaik mungkin. Mahasiswa harus bisa menyelesaikan studi di perguruan tinggi secara tepat waktu. Ada banyak kerugian yang akan dialami mahasiswa jika mereka menyia-nyiakan waktu. Dapat disimpulkan bahwa tidak ada yang salah dari mahasiswa yang mengalami *drop out* jika ia mengalami hal tersebut dengan niat yang lebih baik. Akan tetapi mahasiswa yang mengalami *drop out* juga sangat merugikan diri sendiri maupun perguruan tinggi. Maka dari itu pentingnya menata niat sebelum melakukan sesuatu apapun tak terkecuali menata

niat untuk apa menjadi mahasiswa dan memperkirakan persiapan baik persiapan mental, pikiran, uang, waktu dan lain-lain. Berikut adalah kaidah fiqh (as-Suyuthi:6) mengenai pentingnya sebuah niat:

الْأُمُورُ بِمَقَاصِدِهَا

Yang artinya: *"Segala perkara tergantung dengan niatnya."*

Banyak makna yang dapat diambil dari kaidah fiqh di atas yang erat kaitanya dengan mahasiswa *drop out*. Dengan adanya niat maka akan terbedalah mahasiswa yang mengalami *drop out* dengan melakukan hal-hal yang lebih penting seperti merawat orang tua atau mencari nafkah untuk kebutuhan keluarga dengan mahasiswa yang mengalami *drop out* dikarenakan memang dia malas belajar dan mendapat nilai dibawah rata-rata.

UIN SUNAN AMPEL
S U R A B A Y A

BAB III

METODE PENELITIAN

3.1. Jenis Penelitian

Klasifikasi mahasiswa *drop out* ini termasuk dalam jenis penelitian kuantitatif dikarenakan dalam prosesnya diperlukan adanya unsur perhitungan secara matematis serta menggunakan data yang berbentuk angka (Hasibuan et al., 2022). Penelitian ini menggunakan data semester, jumlah SKS, nilai IPK, nilai IPS, jalur masuk dan nominal UKT. Perhitungan pada penelitian ini menggunakan *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting*

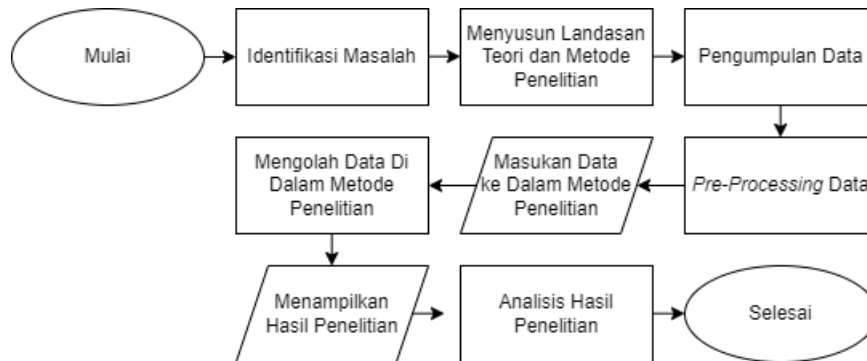
3.2. Sumber Data

Dalam klasifikasi mahasiswa *drop out* ini digunakan data sekunder yang diambil dari data siacad mahasiswa UINSA (<http://my.uinsa.ac.id/index>). Data yang diambil adalah data mahasiswa *drop out* dan mahasiswa aktif *strata 1* angkatan 2015 sampai 2021. Peneliti mengambil sampel mahasiswa dari beberapa sampel program studi di UINSA.

3.3. Variabel Penelitian

Peneliti menggunakan data status mahasiswa sebagai variabel dependen (Y) dan semester, jumlah SKS, nilai IPK, nilai IPS, jalur masuk dan nominal UKT sebagai variabel independen (X).

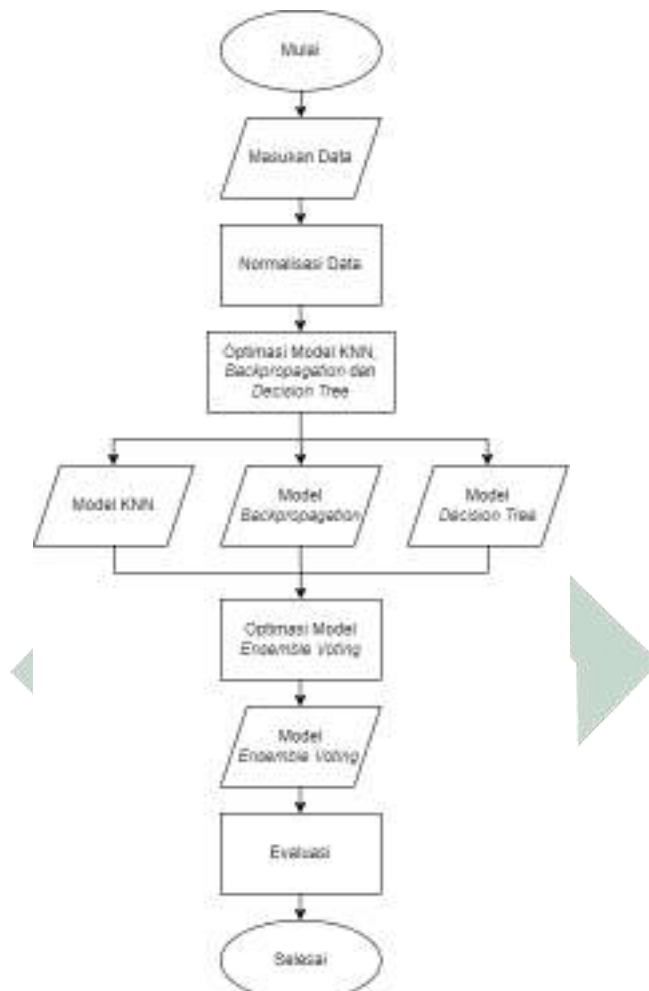
3.4. Tahapan Penelitian



Gambar 3.1 Tahapan Penelitian

Peneliti melakukan identifikasi dengan menganalisa beberapa penelitian terdahulu pada studi kasus pengelompokan mahasiswa *drop out*. Peneliti juga melakukan identifikasi pada metode yang akan digunakan yaitu *Ensemble Voting*, *Decision Tree C4.5*, KNN dan *Backpropagation*. Peneliti mengambil data dari *database* UINSA dengan variabel semester, jumlah SKS, nilai IPK, nilai IPS, jalur masuk dan nominal UKT. Data yang didapatkan akan diolah menggunakan *software python*. Peneliti melakukan estimasi pada masing-masing metode klasifikasi untuk mendapatkan model terbaik. Pada tahap akhir peneliti mengevaluasi dan membandingkan hasil akurasi pada masing-masing metode tunggal dan hasil akurasi pada metode *Ensemble Voting*.

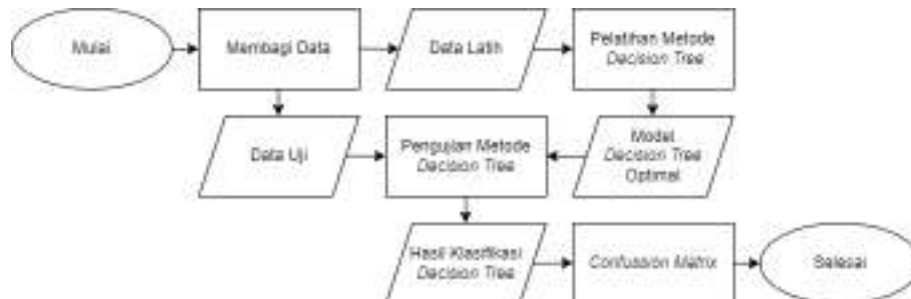
3.5. Tahapan Metode Penelitian



Gambar 3.2 Tahapan Metode Penelitian

Pada tahap awal metode penelitian ini, peneliti mengekspor data ke dalam aplikasi *python*. Data yang sudah terbaca akan dinormalisasi terlebih dahulu. Penelitian ini menggabungkan hasil klasifikasi dari 3 metode tunggal. Sebelum menerapkan *Ensemble Voting* peneliti mencari model optimal dari 3 metode tunggal. Hasil klasifikasi metode tunggal akan dimasukan ke metode *Ensemble Voting*. Setelah didapatkan hasil klasifikasi dan nilai akurasi dari 3 metode tunggal dan metode *Ensemble Voting* peneliti melakukan evaluasi metode penelitian.

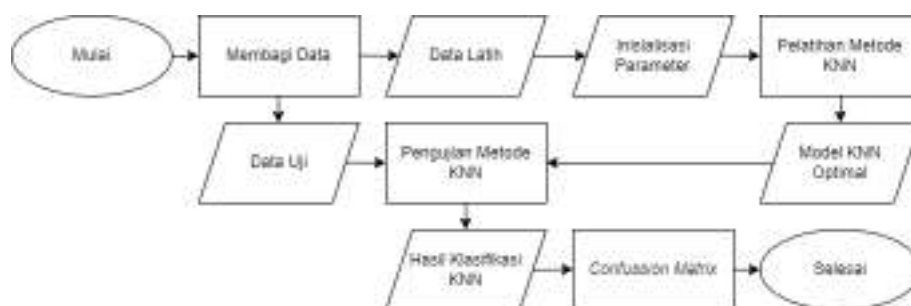
3.5.1. Tahapan Metode *Decision Tree C4.5*



Gambar 3.3 Tahapan Metode *Decision Tree C4.5*

Data yang didapatkan akan dibagi ke data *training* dan data *testing*. Data *training* akan diolah untuk mendapatkan model *Decision Tree C4.5* terbaik. Pada proses *training* digunakan persamaan (2.2, 2.3, 2.4, 2.5) untuk memilih atribut yang digunakan sebagai *node*. Setelah mendapatkan model terbaik peneliti memasukan data *testing* ke dalam pohon keputusan untuk melakukan *training* metode *Decision Tree C4.5*. Hasil dari metode *Decision Tree C4.5* akan dievaluasi menggunakan *confusion matrix*.

3.5.2. Tahapan Metode KNN



Gambar 3.4 Tahapan Metode KNN

Data yang didapatkan akan dibagi ke data *training* dan data *testing*. Data *training* akan diolah untuk mendapatkan model KNN terbaik dengan melakukan inisialisasi nilai parameter k . Peneliti melakukan uji coba parameter k pada nilai 3, 5, 7 dan 9. Dalam menentukan jarak dari 2 titik peneliti menggunakan rumus

Euclidean pada persamaan [2.6](#). Setelah mendapatkan model terbaik dari uji coba nilai k , peneliti memasukan data uji untuk melakukan *training* metode KNN. Hasil dari metode KNN akan dievaluasi menggunakan *confusion matrix*.

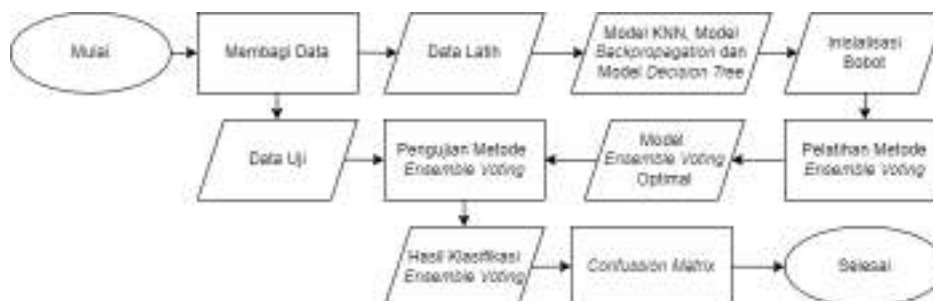
3.5.3. Tahapan Metode *Backpropagation*



Gambar 3.5 Tahapan Metode *Backpropagation*

Data yang didapatkan akan dibagi ke data *training* dan data *testing*. Data *training* akan diolah untuk mendapatkan model *Backpropagation* terbaik dengan melakukan inisialisasi nilai bobot. Peneliti melakukan uji coba nilai *hidden layer* yaitu *hidden layer* (10, 10), (10, 50), (10, 100), ..., (500, 300), (500, 400) dan (500, 500). Dalam *Backpropagation* terdapat 3 tahap yaitu *feed forward*, *backward*, *update* bobot. Saat proses *forward* digunakan [\(2.8, 2.9, 2.10, 2.11\)](#) untuk menghitung data *input* dan bobot untuk mencari *output*. Pada tahap *backward* dihitung nilai *error* dengan menghitung selisih antara *output* dan target menggunakan persamaan [\(2.12, 2.13, 2.14, 2.15, 2.16, 2.17, 2.18, 2.19, 2.20\)](#). Pada tahap ketiga yaitu *update* bobot berdasarkan *error* yang telah didapatkan pada tahap *backward*. Setelah mendapatkan model terbaik peneliti memasukan data *testing* untuk melakukan pengujian metode *Backpropagation*. Hasil perhitungan *Backpropagation* akan dievaluasi menggunakan *confusion matrix*.

3.5.4. Tahapan Metode *Ensemble Voting*



Gambar 3.6 Tahapan Metode *Ensemble Voting*

Data yang didapatkan akan dibagi ke data *training* dan data *testing*. Data *training* akan diolah untuk mendapatkan model *Ensemble Voting* terbaik dengan melakukan inisialisasi nilai bobot dan memasukan hasil klasifikasi dari metode KNN, *Backpropagation* dan *Decision Tree C4.5* pada persamaan 2.21 pada *Hard Voting* dan persamaan 2.22 pada *Soft Voting*. Peneliti menggunakan uji coba nilai bobot dengan nilai uji coba $(1, 1, 1)$, $(1, 1, 2)$, $(1, 2, 1)$ dan $(2, 1, 1)$. Peneliti melakukan uji coba pada kedua jenis *Ensemble Voting* tersebut. Setelah mendapatkan model terbaik pada masing-masing jenis *Ensemble Voting*, peneliti memasukan data *testing* untuk melakukan pengujian metode *Ensemble Voting*. Hasil dari metode *Ensemble Voting* akan dievaluasi menggunakan *confusion matrix*.

BAB IV

HASIL DAN PEMBAHASAN

4.1. Deskripsi Data

Peneliti menggunakan 1092 data mahasiswa dari seluruh fakultas yang ada di UIN Sunan Ampel Surabaya (UINSA, 2023). Peneliti mengambil data mahasiswa dari beberapa program studi yang ada di fakultas tersebut. Berikut adalah data mahasiswa yang digunakan:

	NIM	Program Studi	Masuk	Status	Semester	SKS	IPK	IPS	UKT	Jalur Masuk
0	3010121001	Bahasa dan Sastra Arab	20211	Aktif	3	40	3.62	0.00	2405000	RAPORT
1	3020121031	Bahasa dan Sastra Arab	20211	Aktif	3	47	3.77	3.77	2405000	TES
2	3010121002	Bahasa dan Sastra Arab	20211	Aktif	3	40	3.35	3.35	3605000	RAPORT
3	3010121003	Bahasa dan Sastra Arab	20211	Aktif	3	40	3.57	0.00	3605000	RAPORT
4	3010120001	Bahasa dan Sastra Arab	20201	Aktif	4	91	3.56	3.63	3605000	RAPORT
...
1087	J91218098	Psikologi	20181	Drop Out	4	21	1.65	0.00	1325000	TES
1088	J91218098	Psikologi	20181	Drop Out	4	22	1.65	0.00	2935000	TES
1089	J71215122	Psikologi	20151	Drop Out	2	0	0.15	0.00	2215000	RAPORT
1090	J71215044	Psikologi	20151	Drop Out	2	0	0.00	0.00	2215000	RAPORT
1091	J71215069	Psikologi	20151	Drop Out	2	0	0.00	0.00	4030000	RAPORT

Gambar 4.1 Data Mahasiswa

Gambar di atas menunjukkan data sebanyak 1092 data mahasiswa dengan 557 data mahasiswa aktif dan 535 data mahasiswa *drop out*. Dari data di atas terdapat banyak atribut yang dapat digunakan namun peneliti hanya akan menggunakan beberapa atribut saja yang akan dijadikan sebagai variabel independen (X) dan variabel dependen (Y). Pada atribut jalur didalamnya terdapat banyak nilai diantaranya raport, tes, mandiri reguler, mandiri prestasi dan program

khusus. Data mahasiswa dengan jalur tes yang dimaksud adalah mahasiswa yang masuk melalui jalur SBMPTN dan UMPTKIN. Sedangkan data mahasiswa dengan jalur raport yang dimaksud adalah mahasiswa yang masuk melalui jalur SNMPTN dan SPANPTKIN:

Status	Semester	SKS	IPK	IPS	UKT	Jalur Masuk
Aktif	3	41	3.67	3.67	3230000	RAPORT
Drop Out	4	72	3.11	1.93	1960000	Program Khusus
Aktif	3	41	2.96	3.06	1831750	RAPORT
Drop Out	4	2	0.38	0.00	1180000	TES
Drop Out	4	22	1.86	0.00	5300000	Mandiri Reguler
Aktif	3	41	3.24	3.24	2155000	RAPORT
Drop Out	4	21	1.57	0.00	150000	Mandiri Prestasi
Aktif	3	42	3.82	3.82	1470000	RAPORT

Gambar 4.2 Sampel Data Penelitian

Dari gambar 4.2 di atas ditunjukkan bahwa terdapat beberapa jalur masuk yang ada. Jalur mandiri prestasi akan direpresentasikan dengan angka 0, jalur mandiri reguler akan direpresentasikan dengan angka 1, jalur program khusus akan direpresentasikan dengan angka 2, jalur raport akan direpresentasikan dengan angka 3 dan jalur tes akan direpresentasikan dengan angka 4. Pada atribut status terdapat 2 data yaitu data mahasiswa aktif dan mahasiswa *drop out*. Pada data mahasiswa aktif akan dikonversi menjadi angka 0 sedangkan pada data mahasiswa *drop out* dikonversi menjadi angka 1. Berikut adalah hasil dari representasi data ke dalam bentuk angka pada atribut status dan jalur masuk:

Status	Semester	SKS	IPK	IPS	UKT	Jalur Masuk
0	3	41	3.67	3.67	3230000	3
1	4	72	3.11	1.93	1960000	2
0	3	41	2.96	3.06	1831750	3
1	4	2	0.38	0.00	1180000	4
1	4	22	1.86	0.00	5300000	1
0	3	41	3.24	3.24	2155000	3
1	4	21	1.57	0.00	150000	0
0	3	42	3.82	3.82	1470000	3

Gambar 4.3 Sampel Data Penelitian

Setelah merubah data pada atribut status dan jalur masuk berikut adalah visualisasi distribusi data pada masing-masing atribut dan keterkaitan terhadap atribut lainnya:



Gambar 4.4 Sebaran Data

Dalam visualisasi di atas terlihat distribusi atribut status mahasiswa terhadap atribut pada variabel dependen (X). Data mahasiswa aktif disimbolkan dengan *scatter* warna biru. Sedangkan data mahasiswa *drop out* disimbolkan dengan *scatter* berwarna orange. Dapat diasumsikan bahwa status mahasiswa drop out memiliki perbedaan karakter cukup signifikan jika dibandingkan dengan mahasiswa aktif. Dari salah satu tampilan distribusi di atas yaitu hubungan antara

atribut semester dengan atribut UKT menunjukkan bahwa semua mahasiswa dengan data semester 2 mengalami drop out. Pada data mahasiswa semester 3 hampir semua mahasiswa berstatus mahasiswa aktif dan hanya ada sedikit saja mahasiswa yang mengalami *drop out*. Sedangkan pada data mahasiswa semester 4 terlihat merata antara sebaran mahasiswa aktif dan mahasiswa *drop out*. Namun data mahasiswa semester 4 dengan nilai UKT tinggi belum tentu mengalami *drop out*. Justru mahasiswa yang mengalami *drop out* dari data semester 4 didominasi oleh mahasiswa dengan nilai UKT kecil atau dibawah rata-rata.

4.2. Pre-processing Data

Dalam melakukan klasifikasi mahasiswa *drop out* ini peneliti hanya akan menggunakan data semester, jumlah SKS, nilai IPK, nilai IPS, jalur masuk serta nominal UKT sebagai variabel independen (X). Sedangkan sebagai variabel respon (Y) peneliti menggunakan data status mahasiswa yang terdiri dari 2 nilai yaitu data mahasiswa aktif dan mahasiswa *drop out*. Dari gambar 4.1 di atas didapatkan distribusi dari masing-masing atribut yang menjadi variabel independen (X):

Tabel 4.1 Distribusi Data

Data	Status	Semester	SKS	IPK	IPS	UKT	Jalur Masuk
Paling Rendah	0	2	0	0	0	150000	0
Rata-rata	0.489	3.753	52.141	2.762	1.713	3004805	2.362
Paling Tinggi	1	4	96	3.9	3.9	9260000	4

Dari gambar di atas ditunjukkan nilai rata-rata, nilai paling rendah dan nilai paling tinggi dari masing-masing atribut. Nilai paling tinggi dan nilai paling rendah dari masing-masing atribut ini akan dijadikan acuan dalam melakukan normalisasi data ke dalam bentuk 0 hingga 1. Pada tahap awal penelitian peneliti melakukan

normalisasi pada data. Berikut adalah contoh perhitungan untuk normalisasi pada data pada variabel SKS menggunakan persamaan [2.1](#):

$$X' = \frac{40 - 0}{96 - 0} = 0.416666666667 \quad (4.1)$$

Perhitungan di atas adalah contoh perhitungan manual untuk menormalisasi data pada atribut SKS. Dimana pada atribut SKS ini memiliki nilai paling kecil sebesar 0 dan nilai paling besar sebesar 96, sedangkan nilai pada baris pertama adalah 40. Dengan menerapkan perhitungan yang sama seperti perhitungan di atas, didapatkan data dengan skala 0 sampai 1 pada seluruh variabel dan ditampilkan pada gambar dibawah ini:

	Status	Semester	SKS	IPK	IPS	UKT	Jalur Masuk
0	0.0	0.5	0.416667	0.928205	0.000000	0.247530	0.75
1	0.0	0.5	0.489583	0.966667	0.966667	0.247530	1.00
2	0.0	0.5	0.416667	0.858974	0.858974	0.379254	0.75
3	0.0	0.5	0.416667	0.915385	0.000000	0.379254	0.75
4	0.0	1.0	0.947917	0.912821	0.930769	0.379254	0.75
...
1087	1.0	1.0	0.218750	0.423077	0.000000	0.128979	1.00
1088	1.0	1.0	0.229167	0.423077	0.000000	0.305708	1.00
1089	1.0	0.0	0.000000	0.038462	0.000000	0.226674	0.75
1090	1.0	0.0	0.000000	0.000000	0.000000	0.226674	0.75
1091	1.0	0.0	0.000000	0.000000	0.000000	0.425906	0.75

Gambar 4.5 Normalisasi Data Penelitian

Data yang sudah dinormalisasi ke dalam range 0 hingga 1 akan digunakan pada proses perhitungan masing-masing metode. Pada atribut Status nilai 1 merepresentasikan status mahasiswa *drop out*, sedangkan nilai 0 merepresentasikan mahasiswa aktif. Sebelum melakukan pelatihan pada

masing-masing metode, peneliti menganalisis data yang akan digunakan. Dataset yang sudah dinormalisasi akan dibagi menjadi data latih dan data uji. Setelah melakukan uji coba perbandingan data dengan uji coba 50:50, 60:40, 70:30, 80:20 dan 90:10 peneliti menggunakan perbandingan data sebesar 90:10 dikarenakan mendapatkan akurasi yang paling tinggi. Data latih akan digunakan untuk membangun masing-masing model yang digunakan. Sedangkan data uji digunakan untuk melakukan evaluasi dari model.

4.3. Perhitungan Metode *Decision Tree C4.5*

4.3.1. Membangun Model *Decision Tree C4.5*

Dalam proses memilih *node* digunakanlah *gain ratio*. *Gain ratio* akan dicari pada masing-masing atribut dan dipilih yg paling tinggi. Dalam mencari nilai *gain ratio* langkah pertama yang dilakukan adalah mencari nilai *entropy* dari masing-masing atribut. Dari dataset yang digunakan diketahui jumlah data yang ada sebanyak 1096 kasus, dengan kelas mahasiswa aktif sebanyak 557 data dan kelas mahasiswa *drop out* sebanyak 535 data. Maka didapatkan nilai *entropy* total menggunakan persamaan 2.5 adalah:

$$entropy \text{ total } [557, 535] = -\frac{557}{1096} \log_2\left(\frac{557}{1096}\right) - \frac{535}{1096} \log_2\left(\frac{535}{1096}\right) = 1.001314322$$

Setelah mendapatkan nilai *entropy* total, dilakukan perhitungan nilai *entropy* dan nilai *gain* dari nilai pada setiap atribut. Berikut adalah contoh perhitungan nilai *entropy* pada atribut semester:

1. Semester 2

jumlah kasus : 29

mahasiswa aktif : 0

mahasiswa *drop out* : 29

$$entropy = -\frac{0}{29} \log_2\left(\frac{0}{29}\right) - \frac{29}{29} \log_2\left(\frac{29}{29}\right) = 0$$

2. Semester 3

jumlah kasus : 211

mahasiswa aktif : 210

mahasiswa *drop out* : 1

$$entropy = -\frac{1}{91} \log_2\left(\frac{1}{91}\right) - \frac{90}{91} \log_2\left(\frac{90}{91}\right) = 0.043414076$$

3. Semester 4

jumlah kasus : 825

mahasiswa aktif : 347

mahasiswa *drop out* : 505

$$entropy = -\frac{347}{825} \log_2\left(\frac{347}{825}\right) - \frac{505}{825} \log_2\left(\frac{505}{825}\right) = 0.975048517$$

Setelah didapatkan *entropy* pada masing-masing atribut, akan dilakukan perhitungan mencari nilai gain menggunakan nilai *entropy* yang didapatkan. Setelah mendapatkan nilai gain dari semua atribut maka akan dapat ditentukan atribut apa yang memiliki nilai tertinggi dan menjadi akar pertama *decision tree* C4.5. Dengan menggunakan nilai gain akan digunakan untuk membangun *rule* atau pohon dari *decision tree* C4.5. Dari pohon keputusan *decision tree* C4.5 yang dibentuk (LAMPIRAN A) menunjukkan bahwa atribut IPS menjadi akar pertama *decision tree* C4.5 dan dilanjutkan atribut yang lain sebagai cabang dari *decision tree* C4.5. Pohon keputusan tersebut digunakan sebagai aturan dalam melakukan klasifikasi mahasiswa *drop out*.

4.3.2. Hasil dan Evaluasi *Decision Tree* C4.5

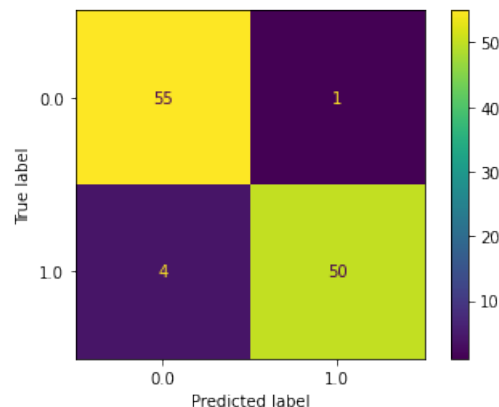
Setelah mendapatkan model dari data latih, peneliti dapat melakukan klasifikasi menggunakan data uji. Berikut adalah tabel sampel hasil klasifikasi

model *Decision Tree C4.5* pada data uji dengan variabel independen menggunakan data yang sudah dinormalisasi dan variabel dependen adalah hasil klasifikasi yang berisi mahasiswa aktif atau mahasiswa *drop out*:

Tabel 4.2 Sampel Hasil Klasifikasi *Decision Tree C4.5*

Semester	SKS	IPK	IPS	UKT	Jalur Masuk	Prediksi
1	0.8125	0.92051282	0.92051282	0.5740944	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.5	0.20833333	0.20833333	0	0.22008782	0.75	0
1	0.97916667	0.89230769	0.89230769	0.11306257	0.25	1
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0.25	0.46410256	0.0	0.1701427	1	1

Dari tabel di atas dapat dilihat sampel hasil klasifikasi menggunakan data uji yang terdiri dari 110 data. Dari hasil klasifikasi di atas dapat dicari berapa jumlah data yang terprediksi benar dan berapa jumlah data yang terprediksi salah. Untuk mengetahui hal jumlah tersebut digunakan data uji pada atribut status mahasiswa yang akan dibandingkan dengan hasil klasifikasi. Hasil dari perbandingan hasil tersebut dapat divisualisasikan menggunakan *Confusion Matrix*. Berikut adalah hasil *Confusion Matrix* dari metode *Decision Tree C4.5*:



Gambar 4.6 Confusion Matrix Decision Tree C4.5

Dari hasil *Confusion Matrix* metode *Decision Tree C4.5* pada gambar 4.6 terdapat 55 data mahasiswa dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa aktif. Terdapat 1 data dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa *drop out*. Terdapat 4 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa aktif. Terdapat 50 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa *drop out*. Dengan menggunakan nilai-nilai pada *Confusion Matrix* dapat dicari nilai akurasi, presisi dan *recall*. Berikut adalah tabel nilai akurasi, presisi dan *recall* dari metode *Decision Tree C4.5* menggunakan persamaan 2.23, 2.24 dan 2.25:

Tabel 4.3 Evaluasi Decision Tree C4.5

Akurasi	Presisi	Recall
95.45%	98.03%	92.59%

Dari tabel di atas dapat diartikan bahwa terdapat 95.45% data terprediksi tepat sebagai mahasiswa aktif dan mahasiswa *drop out* dari seluruh mahasiswa. Terdapat 98.03% data mahasiswa aktif dari seluruh mahasiswa yang diprediksi sebagai mahasiswa aktif. Terdapat 92.59% data mahasiswa yang diprediksi sebagai mahasiswa aktif dibandingkan seluruh mahasiswa yang sebenarnya aktif.

Pada tahap selanjutnya akan dibuktikan apakah metode KNN dan *Backpropagation* mampu memperbaiki nilai akurasi dari metode *Decision Tree C4.5* atau justru mendapatkan nilai akurasi yang lebih kecil.

4.4. Perhitungan Metode KNN

4.4.1. Membangun Model KNN

Dalam membangun model KNN digunakan data dari variabel independen (X) dan dependen (Y). Dengan menggunakan persamaan 2.6 akan dilakukan perhitungan jarak euclidean. Berdasarkan persamaan yang akan digunakan maka diperlukan adanya data uji baru yang akan dilakukan klasifikasi berdasarkan jarak euclidean. Dalam menghitung jarak menggunakan rumus euclidean diperlukan sampel data latih dan data uji untuk mengukur jarak antar data. Diberikan data latih dan data uji sebagai berikut:

Tabel 4.4 Sampel Data Latih Dan Data Uji

Data	Semester	SKS	IPK	IPS	UKT	Jalur Masuk
Data Latih	1	0.92708333	0.93846154	0.97179487	0.97179487	0.75
	0.5	0.4375	0.94871795	0	0.30570801	0.75
	0.5	0.44791667	0.95384615	0	0.30996158	0.75
	⋮	⋮	⋮	⋮	⋮	⋮
Data Uji	1	0.8125	0.92051282	0.92051282	0.5740944	0.25
	⋮	⋮	⋮	⋮	⋮	⋮

Dari sampel data latih dan data uji di atas didapatkan sampel nilai jarak berdasarkan *Euclidean Distance* menggunakan persamaan 2.6 adalah sebagai

berikut:

$$\text{Dist}(x_1, y) = \left((1-1)^2 + (0.92708333 - 0.8125)^2 + (0.93846154 - 0.92051282)^2 + (0.97179487 - 0.92051282)^2 + (0.97179487 - 0.5740944)^2 + (0.75 - 0.25)^2 \right)^{\frac{1}{2}} = 0.290206834$$

$$\text{Dist}(x_2, y) = \left((0.5 - 1)^2 + (0.4375 - 0.8125)^2 + (0.94871795 - 0.92051282)^2 + (0 - 0.92051282)^2 + (0.30570801 - 0.5740944)^2 + (0.75 - 0.25)^2 \right)^{\frac{1}{2}} = 1.560795635$$

$$\text{Dist}(x_3, y) = \left((0.5 - 1)^2 + (0.44791667 - 0.8125)^2 + (0.95384615 - 0.92051282)^2 + (0 - 0.92051282)^2 + (0.30996158 - 0.5740944)^2 + (0.75 - 0.25)^2 \right)^{\frac{1}{2}} = 1.551142114$$

Setelah menghitung jarak euclidean pada semua data latih maka akan didapatkan urutan data yang memiliki jarak terkecil hingga terbesar. Dalam penulisan ini peneliti tidak menampilkan tabel tersebut dikarenakan data yang digunakan cukup banyak dan terlalu kompleks jika dilakukan perhitungan. Tahap selanjutnya adalah menentukan jarak terkecil yang masuk peringkat sejumlah nilai k. Dalam hal ini peneliti melakukan uji coba nilai k. Pada permasalahan ini, nilai parameter k yang akan digunakan yaitu k = 3, k = 5, k = 7 dan k = 9. Berikut adalah tabel hasil uji coba nilai k yang sudah dilakukan:

Tabel 4.5 Uji Coba Parameter K

Nilai k	Akurasi	Presisi	<i>Recall</i>
3	95.45%	98.03%	92.59%
5	95.45%	96.22%	94.44%
7	96.36%	100%	92.59%
9	96.36%	100%	92.59%

Dari tabel di atas dapat diketahui bahwa akurasi terbaik dihasilkan oleh nilai parameter $k = 7$ dengan nilai akurasi yang dihasilkan yaitu sebesar 96.36%. Nilai k terbaik tersebut yang akan digunakan sebagai model terbaik KNN untuk klasifikasi mahasiswa *drop out*.

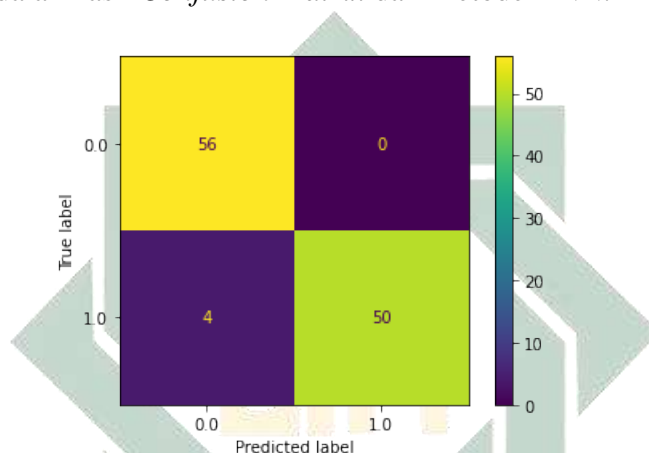
4.4.2. Hasil dan Evaluasi KNN

Berikut adalah tabel sampel hasil klasifikasi model KNN pada data uji dengan variabel independen menggunakan data yang sudah dinormalisasi dan variabel dependen adalah hasil klasifikasi yang berisi mahasiswa aktif atau mahasiswa *drop out*:

Tabel 4.6 Sampel Hasil Klasifikasi KNN

Semester	SKS	IPK	IPS	UKT	Jalur Masuk	Prediksi
1	0.8125	0.92051282	0.92051282	0.5740944	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.5	0.20833333	0.20833333	0	0.22008782	0.75	0
1	0.97916667	0.89230769	0.89230769	0.11306257	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0.25	0.46410256	0.0	0.1701427	1	1

Dari tabel di atas dapat dilihat sampel hasil klasifikasi menggunakan data uji yang terdiri dari 110 data. Dari hasil klasifikasi di atas dapat dicari berapa jumlah data yang terprediksi benar dan berapa jumlah data yang terprediksi salah. Untuk mengetahui hal jumlah tersebut digunakan data uji pada atribut status mahasiswa yang akan dibandingkan dengan hasil klasifikasi. Hasil dari perbandingan hasil tersebut dapat divisualisasikan menggunakan *Confusion Matrix*. Berikut adalah hasil *Confusion Matrix* dari metode KNN:



Gambar 4.7 Confusion Matrix KNN

Dari hasil *Confusion Matrix* metode KNN pada gambar 4.7 terdapat 56 data mahasiswa dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa aktif. Terdapat 0 data dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa *drop out*. Terdapat 4 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa aktif. Terdapat 50 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa *drop out*. Dengan menggunakan nilai-nilai pada *Confusion Matrix* dapat dicari nilai akurasi, presisi dan *recall*. Berdasarkan tabel 4.5 didapatkan nilai akurasi presisi dan *recall* terbaik pada metode KNN adalah 96.36%, 100%, 92.59%. Nilai akurasi, presisi dan *recall* dihitung menggunakan persamaan 2.23, 2.24 dan 2.25. Dari nilai yang didapatkan dapat diartikan bahwa terdapat 96.36% data terprediksi tepat sebagai mahasiswa

aktif dan mahasiswa *drop out* dari seluruh mahasiswa. Terdapat 100% data mahasiswa aktif dari seluruh mahasiswa yang diprediksi sebagai mahasiswa aktif. Terdapat 92.59% data mahasiswa yang diprediksi sebagai mahasiswa aktif dibandingkan seluruh mahasiswa yang sebenarnya aktif. Hasil akurasi yang didapatkan dari metode KNN ini lebih akurat jika dibandingkan dengan hasil akurasi dari metode *Decision Tree C4.5* dengan nilai 95.45%. Pada tahap selanjutnya akan dibuktikan apakah metode *Backpropagation* mampu memperbaiki nilai akurasi dari metode *Decision Tree C4.5* dan KNN atau justru mendapatkan nilai akurasi yang lebih kecil.

4.5. Perhitungan Metode *Backpropagation*

4.5.1. Membangun Model *Backpropagation*

Dalam mencari model *Backpropagation* yang terbaik peneliti akan melakukan uji coba pada beberapa ukuran *hidden layer*. Parameter *hidden layer* sangat mempengaruhi hasil klasifikasi. Banyaknya data dan atribut juga mempengaruhi kesulitan dalam menentukan ukuran *hidden layer* yang optimal. Sebelum membangun model *Backpropagation* peneliti menentukan parameter pelatihan data. Berikut adalah beberapa parameter untuk pelatihan data yang telah ditetapkan:

1. Fungsi aktivasi : sigmoid

Dengan menggunakan fungsi aktivasi ini akan sesuai dengan rumus yang akan digunakan yaitu pada persamaan 2.7 pada *hidden layer*.

2. *Learning rate*: 0,001

Parameter *training* untuk menghitung nilai koreksi bobot saat *training*.

3. *Max iter* : 200

Jumlah maksimum iterasi.

Setelah melakukan inisialisasi pada beberapa parameter peneliti melakukan uji coba ukuran *hidden layer*. Berikut adalah hasil uji coba *hidden layer*:

Tabel 4.7 Uji Coba Parameter *Hidden Layer*

<i>Hidden Layer</i>	Akurasi	Presisi	<i>Recal</i>
10,10	50.90%	0%	0%
⋮	⋮	⋮	⋮
200,10	90.90%	95.83%	95.18%
⋮	⋮	⋮	⋮
500,500	50.90%	0%	0%

Setelah melakukan beberapa uji coba dan didapatkan hasil akurasi terbaik adalah model *Backpropagation* pada ukuran *hidden layer* (200,10) dengan akurasi sebesar 90.90%. Peneliti hanya melakukan uji coba pada 2 *hidden layer* dikarenakan setelah melakukan uji coba pada 1 *hidden layer* hasil akurasi yang didapatkan selalu dibawah uji coba 2 *hidden layer*.

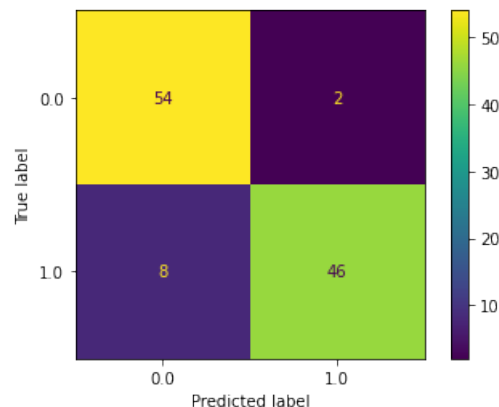
4.5.2. Hasil dan Evaluasi *Backpropagation*

Setelah mendapatkan model terbaik peneliti melakukan klasifikasi menggunakan data uji. Berikut adalah tabel sampel hasil klasifikasi model *Backpropagation* dengan variabel independen menggunakan data yang sudah dinormalisasi dan variabel dependen adalah hasil klasifikasi yang berisi mahasiswa aktif atau mahasiswa *drop out*:

Tabel 4.8 Sampel Hasil Klasifikasi *Backpropagation*

Semester	SKS	IPK	IPS	UKT	Jalur Masuk	Prediksi
1	0.8125	0.92051282	0.92051282	0.5740944	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.5	0.20833333	0.20833333	0	0.22008782	0.75	1
1	0.97916667	0.89230769	0.89230769	0.11306257	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0.25	0.46410256	0.0	0.1701427	1	1

Dari tabel di atas dapat dilihat sampel hasil klasifikasi menggunakan data uji yang terdiri dari 110 data. Pada salah satu sampel di atas menunjukkan hasil klasifikasi mahasiswa adalah mahasiswa *drop out*, sedangkan dengan data yang sama pada model *Decision Tree C4.5* dan KNN mendapatkan hasil klasifikasi mahasiswa aktif. Hal ini yang menjadi tantangan dan tugas yang akan coba diselesaikan oleh *ensemble voting*. Dari hasil klasifikasi di atas dapat dicari berapa jumlah data yang terprediksi benar dan berapa jumlah data yang terprediksi salah. Untuk mengetahui jumlah tersebut digunakan data uji pada atribut status mahasiswa yang akan dibandingkan dengan hasil klasifikasi. Hasil dari perbandingan hasil tersebut dapat divisualisasikan menggunakan *Confusion Matrix*. Berikut adalah hasil *Confusion Matrix* dari metode *Backpropagation*:



Gambar 4.8 Confusion Matrix Backpropagation

Dari hasil *Confusion Matrix* metode *Backpropagation* pada gambar 4.8 terdapat 54 data mahasiswa dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa aktif. Terdapat 2 data dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa *drop out*. Terdapat 8 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa aktif. Terdapat 46 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa *drop out*. Dengan menggunakan nilai-nilai pada *Confusion Matrix* dapat dicari nilai akurasi, presisi dan *recall*. Berdasarkan tabel 4.7 didapatkan nilai akurasi presisi dan *recall* terbaik pada metode *Backpropagation* adalah 90.90%, 95.83%, 85.18%. Nilai akurasi, presisi dan *recall* dihitung menggunakan persamaan 2.23, 2.24 dan 2.25. Dari nilai yang didapatkan dapat diartikan bahwa terdapat 90.90% data terprediksi tepat sebagai mahasiswa aktif dan mahasiswa *drop out* dari seluruh mahasiswa. Terdapat 95.83% data mahasiswa aktif dari seluruh mahasiswa yang diprediksi sebagai mahasiswa aktif. Terdapat 85.18% data mahasiswa yang diprediksi sebagai mahasiswa aktif dibandingkan seluruh mahasiswa yang sebenarnya aktif. Hasil akurasi yang didapatkan dari metode *Backpropagation* terbukti mendapatkan nilai akurasi dibawah dari metode *Decision Tree C4.5* dengan akurasi 95.45% dan metode KNN dengan akurasi 96.36%. Setelah mendapatkan nilai akurasi, presisi

dan *recall* dari masing-masing metode tunggal didapatkan pernyataan bahwa metode tunggal yang mendapatkan nilai akurasi terbaik adalah metode KNN, diikuti metode *Decision Tree C4.5* dan *Backpropagation*. Pada tahap selanjutnya akan dibuktikan apakah metode *Ensemble Voting* mampu memperbaiki nilai akurasi dari metode tunggal *Decision Tree C4.5*, KNN dan *Backpropagation*.

4.6. Perhitungan Metode *Ensemble Voting*

4.6.1. Membangun Model *Ensemble Voting*

Untuk mendapatkan model terbaik dari *ensemble voting* peneliti melakukan uji coba pada *rule* dan bobot. *Rule* yang digunakan dalam *ensemble voting* ada 2 macam yaitu *ensemble hard voting* dan *ensemble soft voting*. Peneliti juga akan melakukan uji coba besarnya bobot pada *rule ensemble soft voting*. *Ensemble voting* dipengaruhi oleh hasil klasifikasi dari model tunggal. Berikut adalah sampel hasil perhitungan model *Decision Tree C4.5*, KNN dan *Backpropagation* pada data uji ke 87 dan 88 yang mendapatkan hasil klasifikasi berbeda pada salah satu model:

Tabel 4.9 Sampel Hasil Klasifikasi Model Tunggal

<i>Decision Tree C4.5</i>	KNN	<i>Backpropagation</i>	<i>Ensemble Voting</i>
⋮	⋮	⋮	⋮
Aktif	Aktif	<i>Drop Out</i>	?
<i>Drop Out</i>	Aktif	Aktif	?
⋮	⋮	⋮	⋮

Pada tabel di atas terlihat bahwa pada data uji ke 87 model *Backpropagation* mendapatkan hasil dari klasifikasi mahasiswa *drop out* sedangkan pada model *Decision Tree C4.5* dan KNN mendapatkan hasil klasifikasi

mahasiswa aktif. Sedangkan pada data uji ke 88 model *Decision Tree C4.5* mendapatkan hasil dari klasifikasi mahasiswa *drop out* sedangkan pada KNN dan *Backpropagation* mendapatkan hasil klasifikasi mahasiswa aktif. Berikut adalah contoh perhitungan manual untuk mendapatkan hasil klasifikasi *ensemble voting* menggunakan hasil output dari 3 model tunggal pada data uji ke 87 mengacu pada hasil dari tabel 4.9:

1. Perhitungan *Ensemble Hard Voting*

Berdasarkan sampel dari tabel 4.9 diketahui terdapat 3 metode dengan hasil klasifikasi sebagai berikut:

Decision Tree C4.5 : mahasiswa aktif

KNN : mahasiswa aktif

Backpropagation : mahasiswa *drop out*

Misal hasil prediksi mahasiswa aktif dinyatakan dalam angka 0 sedangkan hasil prediksi mahasiswa *drop out* dinyatakan dalam angka 1, maka 3 model klasifikasi tersebut dimasukan ke dalam persamaan *Ensemble Hard Voting* pada persamaan 2.21 sebagai berikut:

$$\hat{y} = \text{mode}\{0, 0, 1\} = 0 \quad (4.2)$$

Hasil 0 atau prediksi mahasiswa aktif dapat secara langsung ditentukan dikarenakan hasil dari pemungutan suara / voting dari 3 model tanpa menggunakan nilai bobot. Dengan hasil tersebut *Ensemble Hard Voting* memprediksi bahwa data mahasiswa pada data uji ke 87 adalah mahasiswa *drop out*.

2. *Ensemble Soft Voting*

Dalam *Ensemble Soft Voting* digunakan nilai bobot. Peneliti mencoba

menggunakan bobot (Bobot) (1,1,1), (2,1,1), (1,2,1) dan (1,1,2). Dalam *Ensemble Soft Voting* nilai probabilitas pada masing-masing metode tunggal juga mempengaruhi hasil dari pemungutan suara yang dilakukan. Dalam masing-masing model terdapat nilai probabilitas yang berjumlah 1. Nilai probabilitas pada model tunggal didapatkan dari jumlah nilai yang diharapkan dibagi dengan jumlah seluruh nilai pada keluaran metode tunggal. Pada kasus ini peneliti menggunakan 3 model dan masing-masing model memiliki nilai probabilitas nilai positif (mahasiswa aktif) dan nilai negatif (mahasiswa *drop out*). Jumlah prediksi masing-masing kelas dapat dilihat pada Gambar *Confusion Matrix* 4.6, 4.7 dan 4.8. Pada 3 model tunggal dilakukan klasifikasi menggunakan data uji dengan data sebanyak 110 data. Berikut adalah nilai probabilitas pada masing-masing model tunggal:

(a) *Decision Tree C4.5*

Prediksi aktif : 59

Prediksi *drop out* : 51

UIN SUNAN AMPEL
SURABAYA

$$\frac{59}{110} = 0.54$$

$$\frac{51}{110} = 0.46$$

(b) KNN

Prediksi aktif : 60

Prediksi *drop out* : 50

$$\frac{60}{110} = 0.55$$

$$\frac{50}{110} = 0.45$$

(c) *Backpropagation*

Prediksi aktif : 62

Prediksi *drop out* : 48

$$\frac{62}{110} = 0.56$$

$$\frac{48}{110} = 0.44$$

Dari perhitungan nilai probabilitas di atas, didapatkan nilai probabilitas pada masing-masing model tunggal sebagai berikut:

$$\text{Decision Tree C4.5} = [0.54, 0.46]$$

$$\text{KNN} = [0.55, 0.45]$$

$$\text{Backpropagation} = [0.56, 0.44]$$

Pada perhitungan *Ensemble Soft Voting* ini peneliti menggunakan uji coba nilai bobot (1,1,1), (2,1,1), (1,2,1) dan (1,1,2). Berikut adalah perhitungan pada bobot (1,1,1) dengan menggunakan nilai probabilitas dari masing-masing model tunggal:

$$\text{probabilitas kelas (0)} = \frac{0.54 + 0.55 + 0.56}{3} = 0.55$$

$$\text{probabilitas kelas (1)} = \frac{0.46 + 0.45 + 0.44}{3} = 0.45$$

$$\hat{y} = \arg \max[\text{probabilitas kelas (0)}, \text{probabilitas kelas (1)}]$$

$$\hat{y} = \arg \max[0.55, 0.45] = 0$$

Setelah melakukan perhitungan menggunakan bobot (1,1,1), peneliti

melakukan perhitungan menggunakan bobot (2,1,1) dengan perhitungan sebagai berikut:

$$\text{probabilitas kelas (0)} = 2 \times 0.54 + 1 \times 0.55 + 1 \times 0.56 = 2.19$$

$$\text{probabilitas kelas (1)} = 2 \times 0.46 + 1 \times 0.45 + 1 \times 0.44 = 1.81$$

$$\hat{y} = \arg \max[\text{probabilitas kelas (0), probabilitas kelas (1)}]$$

$$\hat{y} = \arg \max[2.19, 1.81] = 0$$

Setelah melakukan perhitungan menggunakan bobot (2,1,1) didapatkan hasil bahwa mahasiswa pada data uji ke 87 diprediksi sebagai mahasiswa aktif. Selanjutnya dilakukan perhitungan menggunakan bobot (1,2,1) akan menghasilkan prediksi sebagai berikut:

$$\text{probabilitas kelas (0)} = 1 \times 0.54 + 2 \times 0.55 + 1 \times 0.56 = 2.2$$

$$\text{probabilitas kelas (1)} = 1 \times 0.46 + 2 \times 0.45 + 1 \times 0.44 = 1.8$$

$$\hat{y} = \arg \max[\text{probabilitas kelas (0), probabilitas kelas (1)}]$$

$$\hat{y} = \arg \max[2.2, 1.8] = 0$$

Setelah melakukan perhitungan menggunakan bobot (1,2,1) didapatkan hasil bahwa mahasiswa pada data uji ke 87 diprediksi sebagai mahasiswa aktif. Selanjutnya dilakukan perhitungan menggunakan bobot (1,1,2) akan menghasilkan prediksi sebagai berikut:

$$\text{probabilitas kelas (0)} = 1 \times 0.54 + 1 \times 0.55 + 2 \times 0.56 = 2.21$$

$$\text{probabilitas kelas (1)} = 1 \times 0.46 + 1 \times 0.45 + 2 \times 0.44 = 1.79$$

$$\hat{y} = \arg \max[\text{probabilitas kelas (0), probabilitas kelas (1)}]$$

$$\hat{y} = \arg \max[2.21, 1.79] = 0$$

Dari contoh perhitungan menggunakan *rule ensemble hard voting* dan *ensemble soft voting* diatas didapatkan hasil 0 atau prediksi mahasiswa aktif. Berikut adalah contoh perhitungan *ensemble voting* yang diharapkan dapat memperbaiki model *Decision Tree C4.5*, KNN dan *Backpropagation* pada data uji ke 87:

Tabel 4.10 Sampel Hasil Klasifikasi Model Tunggal Dan Ensemble Voting

<i>Decision Tree C4.5</i>	KNN	<i>Backpropagation</i>	<i>Ensemble Voting</i>
Aktif	Aktif	<i>Drop Out</i>	Aktif

Dari tabel di atas didapatkan hasil bahwa *ensemble voting* mendapatkan hasil klasifikasi mahasiswa aktif. Hasil ini adalah sesuai harapan bahwa ensemble voting akan menghasilkan klasifikasi mahasiswa aktif dikarenakan terdapat 2 model yang menghasilkan klasifikasi mahasiswa aktif dan hanya ada 1 model yang menghasilkan klasifikasi mahasiswa *drop out*. Pada model ensemble voting ini peneliti melakukan uji coba pada jenis *rule ensemble voting* yang terdiri dari *ensemble hard voting* dan *ensemble soft voting*. Peneliti juga melakukan uji coba pada nilai bobot yang digunakan. Nilai bobot yang akan di uji coba adalah (1,1,1), (2,1,1), (1,2,1) dan (1,1,2). Berikut adalah tabel hasil uji coba *rule* dan bobot pada *ensemble voting*:

Tabel 4.11 Uji Coba Rule Dan Bobot Ensemble Voting

Ensemble Rule	Bobot	Akurasi	Presisi	Recall
Hard Voting	Tidak Ada	96.36%	100%	92.59%
Soft Voting	(1,1,1)	97.27%	98.11%	96.29%
	(2,1,1)	98.18%	100%	96.29%
	(1,2,1)	97.27%	98.11%	96.29%
	(1,1,2)	95.45%	98.03%	92.59%

Dari hasil uji coba *rule* dan bobot pada tabel di atas didapatkan nilai akurasi yang paling baik adalah *rule ensemble soft voting* dengan bobot (2,1,1). Hasil akurasi yang didapatkan dari *rule ensemble soft voting* dengan bobot (2,1,1) adalah 98.18%.

4.6.2. Hasil dan Evaluasi Ensemble Voting

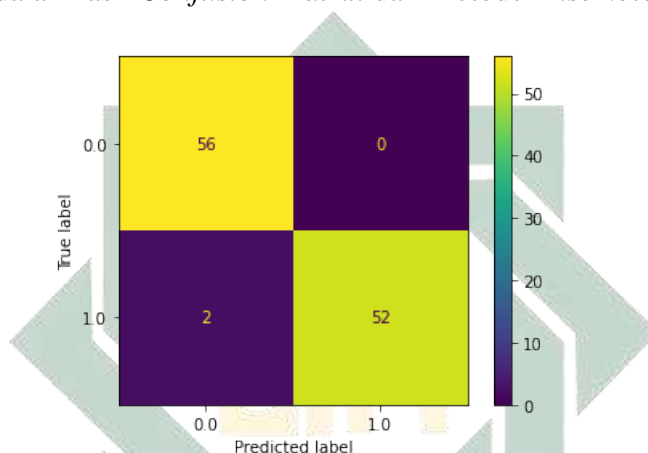
Setelah mendapatkan *rule* dan bobot terbaik peneliti melakukan klasifikasi pada data uji. Berikut adalah hasil klasifikasi dari metode *Ensemble Voting*:

Tabel 4.12 Sampel Hasil Klasifikasi Ensemble Voting

Semester	SKS	IPK	IPS	UKT	Jalur Masuk	Prediksi
1	0.8125	0.92051282	0.92051282	0.5740944	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
0.5	0.20833333	0.20833333	0	0.22008782	0.75	0
1	0.97916667	0.89230769	0.89230769	0.11306257	0.25	0
⋮	⋮	⋮	⋮	⋮	⋮	⋮
1	0.25	0.46410256	0.0	0.1701427	1	1

Dari tabel di atas dapat dilihat sampel hasil klasifikasi menggunakan data

uji yang terdiri dari 110 data. Peneliti mencoba menampilkan output prediksi dari masing-masing metode tunggal. Dari hasil klasifikasi di atas dapat dicari berapa jumlah data yang terprediksi benar dan berapa jumlah data yang terprediksi salah. Untuk mengetahui hal jumlah tersebut digunakan data uji pada atribut status mahasiswa yang akan dibandingkan dengan hasil klasifikasi. Hasil dari perbandingan hasil tersebut dapat divisualisasikan menggunakan *Confusion Matrix*. Berikut adalah hasil *Confusion Matrix* dari metode *Ensemble Voting*:



Gambar 4.9 *Confusion Matrix Ensemble Voting*

Dari hasil *Confusion Matrix* metode *Ensemble Voting* pada gambar 4.9 terdapat 56 data mahasiswa dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa aktif. Terdapat 0 data dengan status asli mahasiswa aktif dan diprediksi sebagai mahasiswa *drop out*. Terdapat 2 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa aktif. Terdapat 52 data dengan status asli mahasiswa *drop out* dan diprediksi sebagai mahasiswa *drop out*. Dengan menggunakan nilai-nilai pada *Confusion Matrix* dapat dicari nilai akurasi, presisi dan *recall*. Berdasarkan tabel 4.11 didapatkan nilai akurasi presisi dan *recall* terbaik pada metode *Ensemble Voting* adalah 98.18%, 100%, 96.29%. Nilai akurasi, presisi dan *recall* dihitung menggunakan persamaan 2.23, 2.24 dan 2.25. Dari nilai yang didapatkan dapat diartikan bahwa terdapat 98.18% data terprediksi

tepat sebagai mahasiswa aktif dan mahasiswa *drop out* dari seluruh mahasiswa. Terdapat 100% data mahasiswa aktif dari seluruh mahasiswa yang diprediksi sebagai mahasiswa aktif. Terdapat 96.29% data mahasiswa yang diprediksi sebagai mahasiswa aktif dibandingkan seluruh mahasiswa yang sebenarnya aktif. Hasil akurasi yang didapatkan dari metode *Ensemble Voting* terbukti mendapatkan nilai akurasi lebih baik dari pada metode tunggal *Decision Tree C4.5* dengan akurasi 95.45%, metode KNN dengan akurasi 96.36% dan *Backpropagation* dengan akurasi 90.90%.

4.7. Evaluasi Metode *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting*

Berdasarkan pada tabel 4.2, 4.6 dan 4.8 terlihat bahwa terdapat perbedaan hasil klasifikasi pada data uji ke 87 dan 88. Berikut adalah gambar tampilan sampel *output* program pada data uji ke 87 dan 88 di masing-masing model tunggal dan *Ensemble Voting* serta nilai pada data asli:

Out[39]:

	KNN	backpo	dec tree	Voting	Y True
0	0.0	1.0	0.0	0.0	0.0
1	0.0	0.0	1.0	0.0	0.0

Gambar 4.10 Sampel Perbandingan Hasil Metode Klasifikasi

Gambar di atas menunjukkan sampel klasifikasi masing-masing metode tunggal, hasil klasifikasi *Ensemble Voting* dan data asli. Dapat dilihat pada sampel data ke 87 menunjukkan perbedaan hasil klasifikasi dari metode tunggal. Metode *Backpropagation* memprediksi 1 atau mahasiswa *drop out* sedangkan 2 metode tunggal lain memprediksi 0 atau mahasiswa aktif. *Ensemble Voting* dengan menggunakan *rule Ensemble Soft Voting* dengan parameter (1,1,1) dapat memprediksi data uji ke 87 sebagai mahasiswa aktif dan dapat dilihat pada kolom

data asli sebenarnya pada data tersebut adalah mahasiswa dengan kode 0 atau dengan status mahasiswa aktif. Sedangkan pada sampel data ke 88 menunjukan metode *Decision Tree C4.5* memprediksi 1 atau mahasiswa *drop out* sedangkan 2 metode tunggal lain memprediksi 0 atau mahasiswa aktif. *Ensemble Voting* dengan menggunakan *rule Ensemble Soft Voting* dengan parameter (1,1,1) dapat memprediksi data uji ke 88 sebagai mahasiswa aktif dan dapat dilihat pada kolom data asli sebenarnya pada data tersebut adalah mahasiswa dengan kode 0 atau dengan status mahasiswa aktif. Berikut adalah tabel hasil akurasi, presisi dan *recall* dari masing-masing metode:

Tabel 4.13 Perbandingan Evaluasi Metode Klasifikasi

Metode / Evaluasi	Akurasi	Presisi	<i>Recall</i>
<i>Decision Tree C4.5</i>	95.45%	98.03%	92.59%
KNN	96.36%	100%	92.59%
<i>Backpropagation</i>	90.90%	95.83%	85.18%
<i>Ensemble Soft Voting</i>	98.18%	100%	96.29%

Dari tabel di atas dapat dibuktikan bahwa *ensemble voting* mampu memperbaiki bukan hanya nilai akurasi masing-masing metode tunggal namun juga memperbaiki nilai presisi dan nilai *recall*. Sebelum melakukan klasifikasi pada 1092 data peneliti melakukan uji coba perbandingan data latih dan data uji. Peneliti melakukan uji coba perbandingan data dengan uji coba 50:50, 60:40, 70:30, 80:20 dan 90:10 dengan hasil seperti pada tabel berikut:

Tabel 4.14 Uji Coba Perbandingan Data Latih dan Data Uji

Metode	50:50	60:40	70:30	80:20	90:10
<i>Decision Tree C4.5</i>	95.42%	96.56%	96.95%	97.26%	95.45%
KNN	95.05%	94.96%	94.81%	94.52%	96.36%
<i>Backpropagation</i>	86.81%	89.70%	86.89%	88.12%	90.90%
<i>Ensemble Hard Voting</i>	94.87%	95.65%	96.03%	95.43%	96.36%
<i>Ensemble Soft Voting</i>	96.15%	96.79%	97.25%	97.26%	98.18%

Dari tabel di atas peneliti menggunakan perbandingan data 90:10 dikarenakan memiliki akurasi paling tinggi yaitu sebesar 98.18%. Berikut adalah perbandingan nilai akurasi, presisi dan *recall* pada masing-masing metode dengan menggunakan perbandingan data 90:10:

4.8. Integrasi Keilmuan

4.8.1. Kajian Keislaman Terkait Hasil Klasifikasi

Penelitian ini menggunakan metode *Ensemble Learning*, dimana cara kerja dari metode ini adalah memaksimalkan hasil dengan cara melihat hasil dari beberapa model. Peneliti menggunakan metode *Ensemble Voting* dan erat kaitanya dengan diskusi. Islam mengajarkan untuk melakukan diskusi untuk menyelesaikan suatu permasalahan. Dalam melakukan klasifikasi mahasiswa *drop out* ini peneliti menggunakan 3 metode tunggal dan menggabungkan mereka ke dalam 1 metode yaitu *Ensemble Voting*. *Ensemble Voting* terbukti dapat memperbaiki nilai akurasi yang didapatkan dari 3 metode tunggal yaitu *Decision Tree C4.5*, KNN dan *Backpropagation*. *Ensemble Voting* mendapatkan akurasi sebesar 98.50% sedangkan 3 metode tunggal hanya mendapatkan akurasi sebesar 96.58%, 97.64%

dan 89.74%. Sesuai dengan namanya yaitu *Ensemble Voting*, metode ini melakukan pemungutan suara terbanyak dari masing-masing metode tunggal. Pemungutan suara ini disebut juga melakukan diskusi untuk mendapatkan keputusan bersama atau biasa disebut dengan musyawarah. Perintah untuk melakukan musyawarah dijelaskan dalam al-Qur'an surat Asy-Syura Ayat 38 dan sebagai berikut:

وَالَّذِينَ اسْتَجَابُوا لِرَبِّهِمْ وَأَقَامُوا الصَّلَاةَ وَأَمْرُهُمْ شُورَىٰ بَيْنَهُمْ وَمِمَّا رَزَقْنَاهُمْ يُنفِقُونَ

Artinya: *”(juga lebih baik dan lebih kekal bagi) orang-orang yang menerima (mematuhi) seruan Tuhan dan melaksanakan salat, sedangkan urusan mereka (diputuskan) dengan musyawarah di antara mereka. Mereka menginfakkan sebagian dari rezeki yang Kami anugerahkan kepada mereka.”*

Ayat di atas menunjukkan kedudukan musyawarah yang sangat penting dalam kehidupan umat Islam, karena perihal musyawarah disebutkan setelah perihal kepatuhan/ketaatan dan perihal shalat. Ayat di atas juga menjelaskan bahwa jika mereka menghadapi masalah maka harus diselesaikan dengan cara musyawarah, selalu bermusyawarah untuk menentukan sikap di dalam menghadapi hal-hal yang bersifat penting serta tidak otoriter atau memaksakan pendapat. Selain berpedoman pada ayat al-quran berikut adalah hadits yang menjelaskan mengenai anjuran untuk bermusyawarah pada hadits riwayat Ahmad nomor 78 dan sebagai berikut:

حَدَّثَنَا عَبْدُ الرَّحْمَنِ بْنُ مَهْدِيٍّ عَنْ سُفْيَانَ عَنْ أَبِي إِسْحَاقَ عَنْ حَارِثَةَ قَالَ جَاءَ نَاسٌ مِنْ أَهْلِ الشَّامِ إِلَى عُمَرَ رَضِيَ اللَّهُ عَنْهُ فَقَالُوا إِنَّا قَدْ أَصَبْنَا أَمْوَالًا وَخَيْلًا وَرَقِيقًا نُحِبُّ أَنْ يَكُونَ لَنَا فِيهَا زَكَاةٌ وَظَهُورٌ قَالَ مَا فَعَلَهُ صَاحِبَايَ قَبْلِي فَأَفْعَلَهُ وَاسْتَشَارَ أَصْحَابَ مُحَمَّدٍ صَلَّى اللَّهُ عَلَيْهِ وَسَلَّمَ وَفِيهِمْ عَلِيٌّ رَضِيَ اللَّهُ عَنْهُ فَقَالَ عَلِيٌّ هُوَ حَسَنٌ إِنْ لَمْ يَكُنْ جَزِيَّةً رَاتِبَةً يُؤْخَذُونَ بِهَا مِنْ بَعْدِكَ

Artinya: "Telah menceritakan kepada kami Abdurrahman Bin Mahdi dari Sufyan dari Abu Ishaq dari Haritsah dia berkata; sekelompok orang dari syam datang kepada Umar, kemudian mereka berkata; "Sesungguhnya kami mempunyai harta, kuda dan budak, dan kami ingin mengeluarkan zakat dan pensucinya." Umar menjawab; "Sebagaimana yang dilakukan oleh kedua sahabatku sebelumku maka aku akan melakukan seperti yang dilakukan mereka berdua." Kemudian dia bermusyawarah dengan para sahabat Nabi Muhammad shallallahu 'alaihi wasallam dan diantara mereka ada Ali, maka Ali berkata; "Itu adalah baik jika itu bukan pajak rutin yang diambil dari mereka setelah kamu.""

Salah satu contoh musyawarah yang diadakan Nabi Shallallahu Alaihi Wassalam terjadi sebelum dan sesudah Perang Badar. Hubab bin Mundzir yang mengenal betul medan pertempuran dan ia bertanya kepada Nabi Muhammad Shallallahu Alaihi Wassalam apakah lokasi ini memang dipilih berdasarkan wahyu yang tak bisa dirubah atautkah ini pendapatmu sebagai strategi perang. Nabi Muhammad Shallallahu Alaihi Wassalam menjawab bahwa ini hanya pendapat beliau dan strategi perang. Kemudian Hubab bin Mundzir mengatakan bahwa lokasi tersebut bukanlah tempat yang tepat. Rasulullah Shallallahu Alaihi Wassalam menerima pendapat sahabatnya itu dan meneruskan berjalan hingga mata air yang terdekat dengan musuh. Usai Perang Badar Nabi Muhammad

Shallallahu Alaihi Wassalam pun menyelenggarakan musyawarah dengan sahabat-sahabatnya. Nabi *Shallallahu Alaihi Wassalam* meminta pendapat beberapa sahabatnya tentang nasib para tawanan. Abu Bakar berpendapat untuk membebaskan tawanan dengan membayar tebusan. Sedangkan Umar bin Khattab berpendapat untuk hukum mati saja para tawanan itu. Dan Abdullah bin Rawahah berpendapat untuk melempar para tawanan ke dalam lembah yang banyak kayu dan membakarnya. Pada akhirnya Nabi *Shallallahu Alaihi Wassalam* lantas lebih condong pada pendapat Abu Bakar. Maka, para tawanan itu bisa dibebaskan dengan jaminan.

Beberapa ayat al-quran dan hadits di atas menjelaskan bahwa pentingnya bermusyawarah dalam menghadapi suatu permasalahan. Dalam penelitian ini metode yang digunakan menerapkan sikap bermusyawarah atau diskusi dengan melakukan pemungutan suara dari beberapa model tunggal. Seperti yang terlihat pada tabel 4.2, 4.6 dan 4.8 terlihat bahwa salah satu metode memprediksi menggunakan data tersebut sebagai mahasiswa aktif, sedangkan pada 2 metode lainnya memprediksi status mahasiswa tersebut adalah *drop out*. Setelah melakukan perhitungan menggunakan *Ensemble Voting* yaitu dengan melakukan pemungutan suara dari masing-masing metode tunggal didapatkan hasil bahwa metode *Ensemble Voting* memprediksi data yang sama sebagai mahasiswa *drop out*.

4.8.2. Kajian Keislaman Terkait Analisi Studi Kasus

Studi kasus dalam penelitian ini adalah klasifikasi mahasiswa *drop out* di UINSA Surabaya. Peneliti menggunakan data mahasiswa dari angkatan 2015 hingga 2021. Namun data yang digunakan hanya data mahasiswa ketika ia memasuki semester 4 dikarenakan salah satu ketentuan mahasiswa *drop out* yang tercantum dalam Buku Pedoman Akademik 2018 UINSA adalah apabila pada

semester 4 mahasiswa tidak dapat menyelesaikan minimal 40 SKS serta memperoleh IPK dibawah 2.00. Selain menggunakan data mahasiswa pada semester 4 peneliti juga menggunakan data mahasiswa aktif semester 2 dan 3 dengan harapan sistem dapat melakukan prediksi sejak dini apakah mahasiswa tersebut berpotensi mengalami *drop out* atau tidak. Dengan adanya penelitian ini diharapkan perguruan tinggi dapat melakukan introspeksi jika mengetahui ada mahasiswa yang berpotensi *drop out*. Ada banyak faktor yang menyebabkan mahasiswa mengalami *drop out*. Rata-rata mahasiswa yang mengalami *drop out* dikarenakan ia tidak hanya fokus ke perkuliahan saja namun memikirkan hal lain seperti keuangan keluarga, masalah dengan orang-orang terdekat dan lain-lain. Selain berkewajiban menuntut ilmu seseorang akan gugur kewajiban menuntut ilmunya ketika menghadapi beberapa permasalahan yang menjadi fokus utama. Seperti yang dijelaskan dalam al-Qur'an surat Al-Baqarah ayat 215 berikut:

يَسْأَلُونَكَ مَاذَا يُنْفِقُونَ قُلْ مَا أَنْفَقْتُ مِنْ خَيْرٍ فَلِلْوَالِدَيْنِ وَالْأَقْرَبِينَ وَالْيَتَامَى وَالْمَسْكِينِ
وَابْنِ السَّبِيلِ وَمَا تَفْعَلُوا مِنْ خَيْرٍ فَإِنَّ اللَّهَ بِهِ عَلِيمٌ

Artinya: *"Mereka bertanya kepadamu (Nabi Muhammad) tentang apa yang harus mereka infakkan. Katakanlah, "Harta apa saja yang kamu infakkan, hendaknya diperuntukkan bagi kedua orang tua, kerabat, anak-anak yatim, orang-orang miskin, dan orang yang dalam perjalanan (dan membutuhkan pertolongan)." Kebaikan apa saja yang kamu kerjakan, sesungguhnya Allah Maha Mengetahuinya."*

Ayat di atas menjelaskan bahwa seseorang anak akan menanggung nafkah kedua orang tuanya ketika mereka sudah berusia lanjut. Jika seorang anak akan memilih fokus bekerja dengan niat ingin membantu perekonomian keluarga maka

ia akan menyingkirkan terlebih dahulu kewajibannya untuk menuntut ilmu dengan mengambil cuti dalam perkuliahan dan lebih memprioritaskan perekonomian keluarga. Adapun permasalahan lain yang memungkinkan mahasiswa mengambil cuti selain membantu perekonomian keluarga. Hal lain yang menjadi salah satu faktor mahasiswa mengambil cuti adalah ketika salah satu anggota keluarga atau orang tuanya mengalami sakit atau sudah berusia lanjut sehingga perlu untuk ditemani setiap hari. Dalam al-Qur'an surat Al Isra ayat 23 menjelaskan bahwa seorang anak akan bertanggung jawab ketika kedua orang tuanya sudah berumur sebagai berikut:

﴿وَقَضَىٰ رَبُّكَ أَلَّا تَعْبُدُوا إِلَّا إِيَّاهُ وَبِالْوَالِدَيْنِ إِحْسَانًا ۖ إِنَّمَا يُبَلِّغُنَّ عَنْكَ الْكِبَرَ أَحَدُهُمَا أَوْ كِلَاهُمَا فَلَا تَقُلْ لَهُمَا أِفْ وَلَا تَنْهَرْهُمَا وَقُلْ لَهُمَا قَوْلًا كَرِيمًا

Artinya: "Tuhanmu telah memerintahkan agar kamu jangan menyembah selain Dia dan hendaklah berbuat baik kepada ibu bapak. Jika salah seorang di antara keduanya atau kedua-duanya sampai berusia lanjut dalam pemeliharaanmu, maka sekali-kali janganlah engkau mengatakan kepada keduanya perkataan "ah" dan janganlah engkau membentak keduanya, serta ucapkanlah kepada keduanya perkataan yang baik. Sekadar mengucapkan kata ah (atau kata-kata kasar lainnya) kepada orang tua tidak dibolehkan oleh agama, apalagi memperlakukan mereka dengan lebih kasar."

Dalam penelitian ini juga menggunakan data mahasiswa semester 2 dan 3 dengan harapan sistem klasifikasi ini dapat memprediksi apakah mahasiswa tersebut berpotensi *drop out* atau tidak. Selain merugikan perguruan tinggi mahasiswa yang mengalami *drop out* juga merugikan diri sendiri. Mahasiswa *drop out* akan mendapatkan banyak kerugian diantaranya adalah rugi biaya dan rugi

waktu. Mahasiswa yang mengalami *drop out* bisa dikatakan rugi dalam hal finansial dikarenakan jika ia mengalami *drop out* maka biaya yang ia gunakan dalam pembayaran kuliah akan sia-sia dikarenakan ia tidak berhasil mendapatkan gelar sebagai sarjana. Mahasiswa *drop out* juga mengalami kerugian dikarenakan biaya yang ia keluarkan untuk pelaksanaan perkuliahan juga cukup banyak, dalam hal ini bisa dikatakan mahasiswa yang mengalami *drop out* sudah melakukan pemborosan. Islam mengajarkan untuk berhemat dan menjauhi sifat boros karena sangat disukai oleh setan. Hal ini dijelaskan dalam al-Qur'an surat Al-Isra' ayat 27 sebagai berikut:

إِنَّ الْمُبَذِّرِينَ كَانُوا إِخْوَانَ الشَّيْطَانِ ۖ وَكَانَ الشَّيْطَانُ لِرَبِّهِ كَفُورًا

Artinya: "Sesungguhnya para pemboros itu adalah saudara-saudara setan dan setan itu sangat ingkar kepada Tuhannya."

Mahasiswa *drop out* tidak mendapat jaminan potongan biaya perkuliahan dan biaya untuk perkuliahan akan sia-sia. Darti potongan hadits riwayat Muslim nomor 3236, Rasulullah Shallallahu Alaihi Wassalam bersabda:

إِنَّ اللَّهَ يَرْضَى لَكُمْ ثَلَاثًا وَيَكْرَهُ لَكُمْ ثَلَاثًا فَيَرْضَى لَكُمْ أَنْ تَعْبُدُوهُ وَلَا تُشْرِكُوا بِهِ شَيْئًا وَأَنْ تَعْتَصِمُوا بِحَبْلِ اللَّهِ جَمِيعًا وَلَا تَفَرَّقُوا وَيَكْرَهُ لَكُمْ قِيلَ وَقَالَ وَكَثْرَةُ السُّؤَالِ

Artinya: "Sesungguhnya Allah menyukai bagimu tiga perkara dan membenci tiga perkara; Dia menyukai kalian supaya beribadah kepada-Nya dan tidak menyekutukan-Nya dengan sesuatu apapun, kalian berpegang teguh dengan agama-Nya dan tidak berpecah belah. Dan Allah membenci kalian dari mengatakan sesuatu yang tidak jelas sumbernya, banyak bertanya dan menyia-nyiakan harta."

Dalam sabda di atas Rasulullah *Shallallahu Alaihi Wassalam* ingin menyampaikan bahwa ada 3 hal yang disukai Allah SWT dan 3 hal yang dibenci Allah SWT. Tiga hal yang disukai Allah SWT adalah orang-orang yang menyembah Allah SWT, orang-orang yang tidak mempersekutukan Allah SWT, dan orang-orang yang berpegang teguh pada agama. Sedangkan tiga hal yang dibenci Allah SWT dari hadits di atas adalah orang-orang yang berkata namun tidak jelas dari mana sumber perkataannya, orang-orang yang banyak bertanya dan orang-orang yang menyia-nyiakan harta.

Dari hadits di atas didapatkan pernyataan bahwa orang-orang yang menyia-nyiakan harta atau pemboros adalah termasuk orang-orang yang dibenci oleh Allah SWT. Ibnu Mas'ud dan Ibnu 'Abbas berpendapat "Pemborosan atau *tabdzir* merupakan kegiatan menginfakkan sesuatu di jalan yang keliru". Pendapat ini didukung oleh Mujahid yang berpendapat bahwa, "Jika seseorang menggunakan semua kekayaan dengan bijaksana, hal tersebut bukan merupakan pemborosan. Tapi ketika seseorang menggunakan kekayaannya walaupun hanya satu mud saja namun pada hal keburukan, maka itulah yang disebut pemborosan". Qotadah juga mengatakan: "Yang disebut sebagai pemborosan adalah mengeluarkan nafkah dalam berbuat maksiat pada Allah, pada jalan yang keliru dan pada jalan untuk berbuat kerusakan". Berdasarkan dari ketiga pendapat ulama tersebut dibuktikan bahwa mereka sepakat mengenai arti dari pemborosan. Pemborosan adalah menyia-nyiakan sesuatu ke jalan yang salah. Selain kehilangan biaya kuliah, mahasiswa yang mengalami *drop out* juga akan kehilangan waktu. Mahasiswa *drop out* akan kehilangan uang untuk biaya kuliah dan waktu yang digunakan pada saat kuliah akan sia-sia. Ibnul Qayyim Al-Jauziyah berkata: "Menyia-nyiakan waktu lebih berbahaya dari kematian, karena menyia-nyiakan

waktu akan memutuskanmu dari Allah dan negeri akhirat, sedangkan kematian hanya memutuskan dirimu dari dunia dan penduduknya”. Anjuran untuk meninggalkan hal yang dianggap sia-sia dijelaskan dalam hadits riwayat At-Tirmidzi dan Ibnu Majah 2239 yang didalamnya Rasulullah *Shallallahu Alaihi Wassalam* bersabda:

مِنْ حُسْنِ إِسْلَامِ الْمَرْءِ تَرْكُهُ مَا لَا يَنْفَعُهُ

Yang artinya: ”Di antara tanda baiknya Islam seseorang adalah meninggalkan sesuatu yang tidak bermanfaat baginya.”

Dari hadits di atas ditunjukkan bahwa seorang muslim yang baik adalah dia yang akan menjauhi sesuatu yang dianggap sia-sia karena masih banyak kegiatan lain yang bisa dilakukan. Ibnu Mas’ud *Radhiallahu ‘Anhu* berkata : “Tiada yang pernah kusesali selain keadaan ketika matahari tenggelam, ajalku berkurang (bertambahnya usia), namun amalanku tidak bertambah”. Seorang muslim harus bijak dalam menghemat keuangan dan memanfaatkan waktu sebaik mungkin, salah satunya sebagai mahasiswa harus mengantisipasi terjadinya *drop out*.

4.8.3. Kajian Keislaman Terkait Keberhasilan Klasifikasi

Dalam melakukan klasifikasi mahasiswa *drop out* ini peneliti menggunakan metode *Ensemble Voting* dengan harapan metode ini dapat memperbaiki tingkat akurasi yang didapatkan dari metode tunggal. Dalam penelitian ini peneliti menggunakan 3 metode tunggal untuk membuktikan asumsi bahwa ensemble voting dapat memperbaiki tingkat akurasi metode tunggal. Metode tunggal yang digunakan adalah *Decision Tree C4.5*, KNN dan *Backpropagation*. Peneliti melakukan uji coba pada nilai parameter dan bobot pada metode tunggal dan juga ensemble voting. peneliti juga melakukan uji coba pada perbandingan data latih

dan data uji yang digunakan. Allah SWT berfirman dalam al-Qur'an surat At-Taubah ayat 105 yang menjelaskan bahwa hasil yang diharapkan akan didapat dengan bekerja keras:

وَقُلِ اعْمَلُوا فَسَيَرَى اللَّهُ عَمَلَكُمْ وَرَسُولُهُ وَالْمُؤْمِنُونَ وَسَتُرَدُّونَ إِلَىٰ عِلْمِ الْغَيْبِ وَالشَّهَادَةِ
فَيُنَبِّئُكُم بِمَا كُنتُمْ تَعْمَلُونَ

Artinya: *"Katakanlah (Nabi Muhammad), "Bekerjalah! Maka, Allah, rasul-Nya, dan orang-orang mukmin akan melihat pekerjaanmu. Kamu akan dikembalikan kepada (Zat) yang mengetahui yang gaib dan yang nyata. Lalu, Dia akan memberitakan kepada kamu apa yang selama ini kamu kerjakan."*

Peneliti mendapatkan hasil bahwa *ensemble voting* dapat memperbaiki tingkat akurasi dari metode tunggal. Jika dilihat pada tabel 4.13 nilai akurasi dari metode *ensemble voting* sebesar 0.9871% sedangkan ketiga metode tunggal hanya menghasilkan akurasi sebesar 0.9700%, 0.9764% dan 0.9487%. Peneliti dapat membuktikan hal tersebut setelah berusaha mendapatkan hasil yang maksimal pada setiap metode yang digunakan. Dalam melakukan klasifikasi mahasiswa *drop out* tidak menutup kemungkinan terjadi kesalahan perhitungan yang dilakukan peneliti. Metode pendekatan numerik akan selalu menghasilkan suatu kesalahan sekecil apapun ukurannya dalam menyelesaikan permasalahan matematika. Terdapat sebuah hadits yang membahas mengenai anjuran untuk memaafkan kesalahan orang lain yaitu dalam hadits riwayat Tirmidzi 2423, Rasulullah *Shallallahu Alaihi Wassalam* bersabda:

قَالَ كُلُّ ابْنِ آدَمَ خَطَّاءٌ وَخَيْرُ الْخَطَّائِينَ التَّوَّابُونَ

Yang artinya: *"Semua anak cucu Adam banyak salah dan sebaik-baik orang*

yang bersalah adalah mereka yang bertaubat.”

Dari beberapa dalil di atas jelas bahwa seorang muslim harus saling memaafkan kesalahan dari saudara dan kerabat baik itu kesalahan yang besar maupun kesalahan yang kecil. Salah satu kesalahan yang harus dimaafkan adalah kesalahan dalam menghitung ataupun dalam usaha untuk melakukan prediksi suatu kejadian tertentu. Karena pada dasarnya segala sesuatu yang sudah terjadi adalah sudah menjadi ketetapan dan izin Allah SWT. Pada penelitian ini peneliti mencoba memaklumi kesalahan dari beberapa metode tunggal dan mencoba menggabungkan beberapa metode dengan harapan mendapat akurasi yang lebih baik.

Peneliti mendapatkan hasil yang cukup bagus dan memuaskan dikarenakan sudah melakukan banyak uji coba. Seperti pada metode KNN peneliti melakukan uji coba nilai k mulai dari $k=3$ hingga $k=9$. Pada metode *Backpropagation* peneliti melakukan uji coba nilai bobot pada *Hidden Layer* dari nilai (10,10) hingga (500,500). Bahkan pada metode *Ensemble Voting* peneliti melakukan uji coba *rule* yang terdiri dari *Ensemble Hard Voting* dan *Ensemble Soft Voting*. Tak hanya itu peneliti bahkan juga melakukan uji coba pada perbandingan data latih dan data uji dengan uji coba 50:50 hingga 90:10. Agama Islam juga mengajarkan tentang bekerja keras untuk mendapatkan hasil yang memuaskan. Seperti dalam salah satu kaidah fiqih (as-Suyuthi, 98) sebagai berikut:

مَا كَانَ أَكْثَرُ فِعْلًا كَانَ أَكْثَرَ فَضْلًا

Artinya: ”Sesuatu yang banyak dikerjakan, lebih banyak keutamaannya.”

Kaidah fiqih di atas bahwa dengan banyak melakukan usaha maka seseorang muslim akan mendapatkan banyak imbalan. Setelah melakukan banyak

uji coba peneliti mendapatkan hasil yang diharapkan yaitu dapat membuktikan bahwa metode *Ensemble Voting* mampu memperbaiki nilai akurasi, presisi dan *recall* dari masing-masing metode tunggal.



UIN SUNAN AMPEL
S U R A B A Y A

BAB V

PENUTUP

5.1. Kesimpulan

Berdasarkan hasil yang telah didapatkan dari penelitian ini dan pembahasan yang telah dipaparkan sebelumnya, maka dapat disimpulkan bahwa:

1. Untuk melakukan perbandingan hasil klasifikasi dari metode *Decision Tree C4.5*, KNN, *Backpropagation* dan *Ensemble Voting* perlu melakukan optimasi pada beberapa metode. Pada metode KNN peneliti melakukan uji coba pada nilai ketetanggaan k . Pada metode *Backpropagation* peneliti melakukan uji coba pada nilai *hidden layer*. Sedangkan pada *Ensemble Voting* peneliti melakukan optimasi dengan menggunakan uji coba *rule* dan nilai bobot pada *rule Ensemble Soft Voting*. Pada metode tunggal terdapat beberapa kali perbedaan hasil klasifikasi mahasiswa. Sering terjadi ketika salah satu metode tunggal mendapatkan hasil klasifikasi mahasiswa aktif atau mahasiswa *drop out* sedangkan pada 2 metode tunggal lainnya mendapatkan hasil klasifikasi yang berbeda. Metode *Ensemble Voting* juga tidak selalu mendapatkan hasil klasifikasi yang diharapkan yaitu mengambil hasil klasifikasi berdasarkan pemungutan suara dari metode tunggal.
2. Setelah melakukan uji coba pada perbandingan data latih dan data uji, peneliti menggunakan perbandingan data latih dan data uji sebesar 90:10 dikarenakan mendapatkan hasil yang paling baik. Dari 557 data mahasiswa aktif dan 535 data mahasiswa *drop out* akan dibagi menjadi 982 data latih

dan 110 data uji. Metode *Decision Tree C4.5* mendapat akurasi 95.45%, presisi 98.03% dan *recall* 92.59%. Dengan parameter k sebesar 7 metode KNN mendapatkan akurasi 96.36%, presisi 100% dan *recall* 92.59%. Dengan menggunakan *hidden layer* (200,10) metode *Backpropagation* mendapat akurasi 90.90%, presisi 95.83% dan *recall* 95.18%. Sedangkan pada *Ensemble Voting rule* yang digunakan adalah *Ensemble Soft Voting* dengan bobot (2,1,1). *Ensemble Voting* mampu memperbaiki akurasi presisi dan *recall* dari masing-masing metode tunggal. *Ensemble Voting* mendapat akurasi 98.18%, presisi 100% dan *recall* 96.29%.

5.2. Saran

Pada beberapa hasil klasifikasi dari *Ensemble Soft Voting* dengan menggunakan bobot (2,1,1) belum bisa mendapatkan hasil klasifikasi yang diharapkan. Hasil klasifikasi *Ensemble Soft Voting* terkadang mendapatkan hasil klasifikasi dari salah satu hasil metode tunggal bahkan mendapatkan hasil yang berbeda dari ketiga metode tunggal. Peneliti mengasumsikan hal ini terjadi dikarenakan peneliti hanya melakukan uji coba menggunakan bobot (1,1,1), (2,1,1), (1,2,1) dan (1,1,2) saja. Saran untuk penelitian selanjutnya untuk melakukan uji coba nilai bobot pada *Ensemble Soft Voting* yang lebih bervariasi seperti (0.1, 0.2, 0.8), (0.7, 0.3, 0.4), (0.9, 0.7, 0.5) dan masih banyak lagi. Pada penelitian berikutnya bisa menambah data sebagai variabel dependen yang lebih banyak lagi dengan mencari faktor lain yang dapat digunakan seperti pekerjaan orang tua, gaji orang tua, beban yang ditanggung keluarga, status apakah mahasiswa mendapatkan beasiswa atau tidak dan faktor lain yang dapat menyebabkan mahasiswa mengalami *drop out*.

DAFTAR PUSTAKA

- Abdullah, A. W. and Muhid, A. (2021). Dukungan Sosial, Academic Satisfaction, Dan Kecenderungan Drop Out Pada Mahasiswa. *Psikoislamika J. Psikol. dan Psikol. Islam*, 18(1):174–187.
- Aesyi, U. S., Lahitani, A. R., Diwangkara, T. W., and Kurniawan, R. T. (2021). Deteksi Dini Mahasiswa Drop Out Menggunakan C5.0. *JISKA (Jurnal Inform. Sunan Kalijaga)*, 6(2):113–119.
- Agrawal, R. (2019). *Smart Intelligent Computing And Applications*, volume 104. Springer Singapore.
- Alban, M. and Mauricio, D. (2019). Neural Networks To Predict Dropout At The Universities. *Int. J. Mach. Learn. Comput.*, 9(2):149–153.
- Aldowah, H., Al-Samarraie, H., Alzahrani, A. I., and Alalwan, N. (2020). Factors Affecting Student Dropout In MOOCs: A Cause And Effect Decision-Making Model. *J. Comput. High. Educ.*, 32(2):429–454.
- Andri, A. and Paulus, P. (2021). Prediksi Mahasiswa Berisiko Drop Out (DO) dengan ADTree dan NNge. *J. SIFO Mikroskil*, 22(1):47–58.
- Ario, T. S. (2020). Problematika Pada Mahasiswa Pekerja Paruh Waktu "Part Time". *Univ. Muhammadiyah Yogyakarta*, page 3.
- Armansyah, A. (2021). Prototipe Jaringan Syaraf Tiruan Multilayer Perceptron Untuk Prediksi Mahasiswa Dropout. *J. Nas. Komputasi dan Teknol. Inf.*, 4(4):265–271.

- Assiri, A. S., Nazir, S., and Velastin, S. A. (2020). Breast Tumor Classification Using An Ensemble Machine Learning Method. *J. Imaging*, 6(6).
- Atma, Y. D. and Setyanto, A. (2018). Perbandingan Algoritma C4.5 Dan KNN Dalam Identifikasi Mahasiswa Berpotensi Drop Out. *Metik J. ISSN 2580-1503*, 2(2):31–37.
- Bagli, E. and Visani, G. (2020). Metrics For Multi-Class Classification: An Overview. pages 1–17.
- Benkercha, R. and Moulahoum, S. (2018). Fault Detection And Diagnosis Based On C4.5 Decision Tree Algorithm For Grid Connected PV System. *Sol. Energy*, 173(July):610–634.
- Brownlee, J. (2021). Ensemble Machine Learning With Python: 7-Day Mini-Course. *Mach. Learn. Mastery*, page 20.
- Casanova, J. R., Cervero, A., Núñez, J. C., Almeida, L. S., and Bernardo, A. (2018). Factors That Determine The Persistence And Dropout Of University Students. *Psicothema*, 30(4):408–414.
- Cherfi, A., Nouira, K., and Ferchichi, A. (2018). Very Fast C4.5 Decision Tree Algorithm. *Appl. Artif. Intell.*, 32(2):119–137.
- Dalipi, F., Imran, A. S., and Kastrati, Z. (2018). MOOC Dropout Prediction Using Machine Learning Techniques: Review And Research Challenges. *IEEE Glob. Eng. Educ. Conf. EDUCON*, 2018-April:1007–1014.
- Damanik, I. S., Windarto, A. P., Wanto, A., Poningsih, Andani, S. R., and Saputra, W. (2019). Decision Tree Optimization in C4.5 Algorithm Using Genetic Algorithm. *J. Phys. Conf. Ser.*, 1255(1).

- Derisma, D. (2020). Perbandingan Kinerja Algoritma Untuk Prediksi Penyakit Jantung Dengan Teknik Data Mining. *J. Appl. Informatics Comput.*, 4(1):84–88.
- Dinata, R. K., Akbar, H., and Hasdyna, N. (2020). Algoritma K-Nearest Neighbor dengan Euclidean Distance dan Manhattan Distance untuk Klasifikasi Transportasi Bus. *Ilk. J. Ilm.*, 12(2):104–111.
- Dogan, A. and Birant, D. (2021). Machine Learning And Data Mining In Manufacturing. *Expert Syst. Appl.*, 166:114060.
- Dong, X. and Yu, Z. (2019). A Survey On Ensemble Learning. pages 1–18.
- Evendi, E., Tadriss, J., Nusa, P., Barat, T., Rekapitulasi, T., Mahasiswa, S., and Tadriss, J. (2019). Jurnal Riset Teknologi dan Inovasi Pendidikan Evaluasi Penerimaan Mahasiswa Baru Jurusan Tadris Matematika PTKIN-UIN Mataram. 2(1):256–271.
- Febriadi, B., Zamzami, Z., Yunefri, Y., and Wanto, A. (2018). Bipolar Function In Backpropagation Algorithm In Predicting Indonesia's Coal Exports By Major Destination Countries. *IOP Conf. Ser. Mater. Sci. Eng.*, 420(1).
- GitHub (2022). EnsembleVoteClassifier: A majority voting classifier. https://rasbt.github.io/mlxtend/user_guide/classifier/EnsembleVoteClassifier/.
- Haidar, L. R., Sedyono, E., and Iriani, A. (2020). Analisa Prediksi Mahasiswa Drop Out Menggunakan Metode Decision Tree Dengan Algoritma Id3 Dan C4.5. *J. Transform.*, 17(2):97.
- Hasibuan, A. T., Sianipar, M. R., Ramdhani, A. D., Putri, F. W., and Ritonga,

- N. Z. (2022). Konsep Dan Karakteristik Penelitian Kualitatif Serta Perbedaannya Dengan Penelitian Kuantitatif. *J. Pendidik. Tambusa*, 6(2):8686–8692.
- Haut, J. M., Member, S., Paoletti, M. E., Member, S., Plaza, J., Member, S., Li, J., Member, S., and Plaza, A. (2018). Active Learning With Convolutional Neural Networks For Hyperspectral Image Classification Using A New Bayesian Approach. *IEEE Trans. Geosci. Remote Sens.*, PP:1–22.
- Idris, M. (2021). Analisis Prestasi Akademik Mahasiswa Teknik Informatika Institut Teknologi Sumatera (ITERA) Berdasarkan Jalur Seleksi Masuk Perguruan Tinggi Negeri. *J. Sci. Appl. Technol.*, 5(1):126.
- Irmayanti, D., Muhyidin, Y., and Nurjaman, D. A. (2021). Prediksi Mahasiswa Berpotensi Drop Out Dengan Metode Iteratif Dichotomiser 3 (ID3). *J. Teknol. Inf.*, 5(2):103–113.
- Isphording, I. E. and Qendrai, P. (2019). Gender Differences In Student Dropout In STEM. *IZA Discuss. Pap.*, 3(82):126–134.
- Jasmir, Zaenal, D., J. P. A., and Rasywir, E. (2018). Prediksi Mahasiswa Drop Out Dengan Menggunakan Algoritma Klasifikasi Data Mining. *Pros. Annu. Res. Semin.*, 4(1):82–87.
- Jusriana, A., Sulasteri, S., Jurusan, N. A., Fisika, P., Tarbiyah, F., Uin, K., and Makassar, A. (2020). Perbandingan Pemahaman Konsep Fisika Dasar Berdasarkan Jalur Masuk Perguruan Tinggi Pada Program Studi Pendidikan Fisika Uin Alauddin Makassar. *J. Pendidik. Fis.*, 10(2):81–87.
- Kajornrit, J. and Chaipornkaew, P. (2017). A Comparative Study Of Ensemble

- Back-Propagation Neural Network For The Regression Problems. *Proceeding 2017 2nd Int. Conf. Inf. Technol. INCIT 2017*, 2018-Janua:1–6.
- Khoerulloh, A. R., Nurjanah, D., and Romadhony, A. (2019). Sistem Rekomendasi Mata Kuliah Pilihan Menggunakan Association Rule Dan Ant Colony Optimization (Studi Kasus Mata Kuliah Di Jurusan Teknik Informatika Universitas Telkom). *eProceedings Eng.*, 6(2):9597–9617.
- Kristian Nazarius, J., Nusa Mandiri, S., Jatiwaringin Raya No, J., Melayu, C., and Timur, J. (2021). Analisis Pemilihan Siswa Untuk Jalur SNMPTN dengan Metode Weighted Product (WP) Dan Weighted Sum Model (WSM). *J. Sains Komput. Inform. (J-SAKTI)*, 5(1):135–142.
- Kumar, A. and Jain, M. (2020). *Ensemble Learning For AI Developers*.
- Kuzilek, J., Vaclavek, J., Fuglik, V., and Zdrahal, Z. (2018). *Student Drop-out Modelling Using Virtual Learning Environment Behaviour Data*, volume 11082 LNCS. Springer International Publishing.
- Laufer, M. and Gorup, M. (2019). The Invisible Others: Stories Of International Doctoral Student Dropout. *High. Educ.*, 78(1):165–181.
- Limsathitwong, K., Tiwatthanont, K., and Yatsungnoen, T. (2018). Dropout Prediction System To Reduce Discontinue Study Rate Of Information Technology Students. *Proc. 2018 5th Int. Conf. Bus. Ind. Res. Smart Technol. Next Gener. Information, Eng. Bus. Soc. Sci. ICBIR 2018*, pages 110–114.
- Lizarti, N. and Ulfah, A. N. (2019). Penerapan Algoritma K-Nearest Neighbor Untuk Penentuan Peminatan Studi (Studi Kasus : Program Studi Teknik Informatika STMIK Amik Riau). *Fountain Informatics J.*, 4(1):1.

- Manrique, R., Nunes, B. P., Marino, O., Casanova, M. A., and Nurmikko-Fuller, T. (2019). An Analysis Of Student Representation, Representative Features And Classification Algorithms To Predict Degree Dropout. *ACM Int. Conf. Proceeding Ser.*, pages 401–410.
- Maros, H. and Juniar, S. (2016). Advanced Decision Sciences Based On Deep Learning And Ensemble Learning Algorithms A Practical Approach Using Python. pages 1–23.
- Mironczuk, M. M. and Protasiewicz, J. (2018). A Recent Overview Of The State-Of-The-Art Elements Of Text Classification.
- Mohanapriya, M. and Lekha, J. (2018). Comparative Study Between Decision Tree And Knn Of Data Mining Classification Technique. *J. Phys. Conf. Ser.*, 1142(1).
- Mughal, J., Jawad, M., and Mughal, H. (2018). Data Mining: Web Data Mining Techniques, Tools And Algorithms: An Overview. *IJACSA) Int. J. Adv. Comput. Sci. Appl.*, 9(6).
- Muhamad, M., Windarto, A. P., and Suhada, S. (2019). Penerapan Algoritma C4.5 Pada Klasifikasi Potensi Siswa Drop Out. *KOMIK (Konferensi Nas. Teknol. Inf. dan Komputer)*, 3(1):1–8.
- Mustaqim, T., Umam, K., and Muslim, M. A. (2020). Twitter Text Mining For Sentiment Analysis On Government's Response To Forest Fires With Vader Lexicon Polarity Detection And K-Nearest Neighbor Algorithm. *J. Phys. Conf. Ser.*, 1567(3).
- Nagy, M. and Molontay, R. (2018). Predicting Dropout In Higher Education Based

- Pn Secondary School Performance. *INES 2018 - IEEE 22nd Int. Conf. Intell. Eng. Syst. Proc.*, pages 000389–000394.
- Nalim, N., Dewi, H. L., and Safii, M. A. (2021). Analisis Faktor-Faktor yang Mempengaruhi Keberhasilan Studi Mahasiswa di PTKIN Provinsi Jawa Tengah. *J. Kependidikan J. Has. Penelit. dan Kaji. Kepustakaan di Bid. Pendidikan, Pengajaran dan Pembelajaran*, 7(4):1003.
- Novianti, H. (2019). Sistem Pendukung Keputusan Penentuan Uang Kuliah Tunggal Dengan Metode Analitical Hierarchy Process (Ahp) Di Universitas Sriwijaya. *JSI J. Sist. Inf.*, 11(1):1711–1721.
- Patacsil, F. F. (2020). Survival Analysis Approach For Early Prediction Of Student Dropout Using Enrollment Student Data And Ensemble Models. *Univers. J. Educ. Res.*, 8(9):4036–4047.
- Pawlovsky, A. P. (2018). An Ensemble Based On Distances For A KNN Method For Heart Disease Diagnosis. *Int. Conf. Electron. Inf. Commun. ICEIC 2018*, 2018-Janua:1–4.
- Pristyanto, Y. (2019). Penerapan Metode Ensemble Untuk Meningkatkan Kinerja Algoritme Klasifikasi Pada Imbalanced Dataset. *J. Teknoinfo*, 13(1):11.
- Pujianto, A., Kusrini, K., and Sunyoto, A. (2018). Perancangan Sistem Pendukung Keputusan Untuk Prediksi Penerima Beasiswa Menggunakan Metode Neural Network Backpropagation. *J. Teknol. Inf. dan Ilmu Komput.*, 5(2):157.
- Purba, I. S., Wanto, A., Riansah, R. M., Ahmad, Y., Siregar, S. P., Winanjaya, R., Julham, M., and Silitonga, H. (2019). Accuracy Level Of Backpropagation

- Algorithm To Predict Livestock Population Of Simalungun Regency In Indonesia. *J. Phys. Conf. Ser.*, 1255(1).
- Purnawinadi, I. G. (2021). Dampak Pandemi Covid-19 Pada Prestasi Belajar Mahasiswa Keperawatan Tahap Akademik. *J. Sk. Keperawatan*, 7(1):63–69.
- Rad, S. E. and Behjat, A. R. (2019). Document Classification Base On Ensemble Classifiers Support Vector Machine , Multilayer Perceptron And K-Nearest Neighbors. *J Biochem Tech*, (2):174–182.
- Raharjo, M., Napiah, M., Putra, J. L., and Mustofa, M. (2020). Prediksi Pengaruh Matakuliah Terhadap Peminatan Outline Tugas Akhir Mahasiswa Dengan Jaringan Syaraf Tiruan. *J. Infortech*, 2(1):78–83.
- Ramayasa, I. P. (2018). Perancangan Sistem Klasifikasi Mahasiswa Drop Out Menggunakan Algoritma K-Nearest Neighbor. *Semin. Nas. Sist. Inf. dan Teknol. Inf. 2018*, 1(1):585–589.
- Ritonga, H. J. (2018). Manajemen Waktu Dalam Islam. *Al-Idarah*, 5(6):51–53.
- Russell, M. and Klassen, M. (2019). *Mining The Social Web: Data Mining Facebook, Twitter, LinkedIn, Instagram, GitHub, And More.*
- Sacramento, J., Bengio, Y., Costa, R. P., and Senn, W. (2018). Dendritic cortical microcircuits approximate the backpropagation algorithm. *Adv. Neural Inf. Process. Syst.*, 2018-Decem(NeurIPS):8721–8732.
- Sannino, G. and Pietro, G. D. (2018). A Deep Learning Approach For ECG-Based Heartbeat Classification For Arrhythmia Detection. *Futur. Gener. Comput. Syst.*

- Santos, G. A., Belloze, K. T., Tarrataca, L., Haddad, D. B., Bordignon, A. L., and Brandao, D. N. (2020). Evolved Tree: Analyzing Student Dropout In Universities. *Int. Conf. Syst. Signals, Image Process.*, 2020-July:173–178.
- Saragih, J. (2018). Analisa Penambangan Data Menggunakan Algoritma Aco (Ant Colony Optimization) : *Ant_Miner3*. 3(1):2528–5114.
- Sari, E. Y., Kusriani, K., and Sunyoto, A. (2021). Analisis Akurasi Jaringan Syaraf Tiruan Dengan Backpropagation Untuk Prediksi Mahasiswa Dropout. *Creat. Inf. Technol. J.*, 6(2):85.
- Satmoko, D. B., Sukarno, P., and Jadied, E. M. (2018). Peningkatan Akurasi Pendeteksian Serangan DDoS Menggunakan Multiclassifier Ensemble Learning dan Chi-Square. *e-Proceeding Eng.*, 5(3):7877–7985.
- Senthil Kumaran, V. and Malar, B. (2021). Distributed Ensemble Based Iterative Classification For Churn Analysis And Prediction Of Dropout Ratio In E-Learning. *Interact. Learn. Environ.*, 0(0):1–16.
- Setti, S., Wanto, A., Syafiq, M., Andriano, A., and Sihotang, B. K. (2019). Analysis Of Backpropagation Algorithms In Predicting World Internet Users. *J. Phys. Conf. Ser.*, 1255(1).
- Sijabat, A., Sianipar, H. F., and Sijabat, O. P. (2021). Pengaruh Pembelajaran Daring Terhadap Indeks Prestasi Semester Mahasiswa Pendidikan Fisika UHKBNP. ... *dan Pengabdian Masyarakat*. 2021, 2019:434–439.
- Sudiyarno, R., Setyanto, A., and Luthfi, E. T. (2021). Peningkatan Performa Pendeteksian Anomali Menggunakan Ensemble Learning Dan Feature Selection. *Creat. Inf. Technol. J.*, 7(1):1.

- Sulistiani, H. and Aldino, A. A. (2020). Decision Tree C4.5 Algorithm For Tuition Aid Grant Program Classification (Case Study: Department of Information System, Universitas Teknokrat Indonesia). *Edutic - Sci. J. Informatics Educ.*, 7(1):40–50.
- Sumathi, S. and Esakkirajan, S. (2007). *Data Mining And Data Warehousing*, volume 47.
- Suryani, S. (2021). Media Sosialisasi Penerimaan Dalam Meningkatkan Recruitment Calon Mahasiswa Baru. *BERDAYA J. Pendidik. dan Pengabd. Kpd. Masy.*, 3(2):53–58.
- Sutoyo, E. and Almaarif, A. (2020). Educational Data Mining Untuk Prediksi Kelulusan Mahasiswa Menggunakan Algoritme Naïve Bayes Classifier. *J. RESTI (Rekayasa Sist. dan Teknol. Informasi)*, 4(1):95–101.
- Tattar, P. N. (2018). *Hands-On Ensemble Learning With R: A Beginner's Guide To Combining The Power Of Machine Learning Algorithms Using Ensemble Techniques*.
- Triguero, I., García-Gil, D., Maillo, J., Luengo, J., García, S., and Herrera, F. (2019). Transforming Big Data Into Smart Data: An Insight On The Use Of The K-Nearest Neighbors Algorithm To Obtain Quality Data. *Wiley Interdiscip. Rev. Data Min. Knowl. Discov.*, 9(2):1–24.
- UINSA (2023). SIM Akademik Universitas Islam Negeri Sunan Ampel Surabaya. https://sinau.uinsby.ac.id/siakad/data_mahasiswa/detail.
- Vlachas, P. R., Pathak, J., Hunt, B. R., Sapsis, T. P., Girvan, M., Ott, E., and Koumoutsakos, P. (2020). Backpropagation Algorithms And Reservoir

- Computing In Recurrent Neural Networks For The Forecasting Of Complex Spatiotemporal Dynamics. *Neural Networks*, 126:191–217.
- Wahyuni, S., Saragih, k. S., and Perangin-angin, M. I. (2018). Implemntasi Metode Decision Tree C4.5 Untuk Menganalisa Mahasiswa Dop Out. *Ethos*, 6(1):42–51.
- Wawan Firgiawan, Sugiarto Cokrowibowo, and Nuralamsah Zulkarnaim (2019). Komparasi Algoritma Saw, Ahp, Dan Topsis Dalam Penentuan Uang Kuliah Tunggal (Ukt). *J. Comput. Inf. Syst. (J-CIS)*, 1(2):1–11.
- Widodo, D. R. and Dewi, D. K. (2018). Perbedaan Sense Of Belongingness Ditinjau Dari Jalur Seleksi Pada Mahasiswa Universitas Negeri Surabaya. (1995):43–51 Rizqia.
- Yudhistira, S. (2019). Dapatkah Nilai Prestasi Belajar Di Sma Dipergunakan Untuk Meramalkan Nilai Prestasi Belajar Di Perguruan Tinggi. I(11):1–8.
- Yuliana, A. and Pratomo, D. B. (2017). Algoritma Decision Tree (C4.5) Untuk Memprediksi Kepuasan Mahasiswa Terhadap Kinerja Dosen Politeknik Tedc Bandung. *Semnasinotek 2017*, pages 377–384.
- Zhang, S., Cheng, D., Deng, Z., Zong, M., and Deng, X. (2018). A Novel KNN Algorithm With Data-Driven K Parameter Computation. *Pattern Recognit. Lett.*, 109:44–54.