

TESIS

**PERBANDINGAN PREDIKSI PELUANG KARIR ALUMNI
BERDASARKAN KINERJA PROGRAM STUDI
MENGUNAKAN
METODE C4.5 DAN NEURAL NETWORK**

***THE COMPARISON OF ALUMNI'S CAREER OPPORTUNITY
PREDICTION BASED ON STUDY PROGRAM
PERFORMANCES UTILIZING
C4.5 AND NEURAL NETWORK METHODS***

Disusun dan Diajukan oleh

ALI ISRA

D032181030



**PROGRAM MAGISTER TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
MAKASSAR
2022**

**PERBANDINGAN PREDIKSI PELUANG KARIR ALUMNI
BERDASARKAN KINERJA PROGRAM STUDI
MENGUNAKAN
METODE C4.5 DAN NEURAL NETWORK**

Tesis

Sebagai Salah Satu Syarat Untuk Mencapai Gelar Magister

Program Studi

Teknik Elektro

Disusun dan diajukan Oleh

ALI ISRA

Kepada

**PROGRAM MAGISTER TEKNIK ELEKTRO
FAKULTAS TEKNIK
UNIVERSITAS HASANUDDIN
MAKASSAR
2022**

LEMBAR PENGESAHAN TESIS

PERBANDINGAN PREDIKSI PELUANG KARIR ALUMNI BERDASARKAN KINERJA PROGRAM STUDI MENGUNAKAN METODE C4.5 DAN NEURAL NETWORK

Disusun dan diajukan oleh

ALI ISRA
D032181030

Telah dipertahankan dihadapan Panitia Ujian yang dibentuk dalam rangka penyelesaian studi Program Magister Program Studi Teknik Elektro Fakultas Teknik Universitas Hasanuddin.
Pada Tanggal 20 Juni 2022
Dan dinyatakan telah memenuhi syarat kelulusan

Pembimbing Utama,

Pembimbing Pendamping,



Dr. Ir. Zahir Zainuddin, M.Sc.
NIP. 196404271989101002


Dr. Eng. Zulkifli Tahir, S.T.,M.Sc.
NIP. 198404032010121004

Dekan Fakultas Teknik
Universitas Hasanuddin

Ketua Program Studi S2 Teknik Elektro


Prof. Dr. Eng. Ir. Muhammad Isran Ramli, S.T.,M.T.
NIP. 197309262000121002


Prof. Dr. Eng. Ir. Syafaruddin, ST.,M.Eng.,IPU
NIP. 197405301999031003

PERNYATAAN KEASLIAN TESIS

Yang bertanda tangan di bawah ini:

Nama : Ali Isra
Nomor Mahasiswa : D032181030
Program Studi : Teknik Elektro

Menyatakan dengan sebenarnya bahwa tesis yang saya tulis ini benar-benar merupakan hasil karya saya sendiri, bukan merupakan pengambilalihan tulisan atau pemikiran orang lain. Apabila dikemudian hari terbukti atau dapat dibuktikan bahwa Sebagian atau keseluruhan tesis ini hasil karya orang lain, saya bersedia menerima sanksi atas perbuatan tersebut.

Makassar, 14 Juli 2022

Yang menyatakan



Ali Isra

PRAKATA

Alhamdulillah, segala puji bagi Allah SWT atas segala rahmat, taufik dan hidayahnya sehingga tesis dengan judul **“PERBANDINGAN PREDIKSI PELUANG KARIR ALUMNI BERDASARKAN KINERJA PROGRAM STUDI MENGGUNAKAN METODE C4.5 DAN NEURAL NETWORK”** ini dapat diselesaikan. Shalawat dan salam disampaikan kepada Nabi Muhammad SAW yang telah menyinari dunia ini dengan keindahan ilmu dan akhlak yang diajarkan kepada seluruh umatnya. Tesis ini disusun untuk memenuhi salah satu persyaratan memperoleh gelar Magister Teknik (M.T.) dalam bidang keahlian Teknik Informatika pada program studi Teknik Elektro Universitas Hasanuddin.

Oleh karena itu pada kesempatan ini penulis menyampaikan rasa hormat dan menghaturkan terima kasih yang sebesar-besarnya kepada:

1. Yang mulia Ibundaku tercinta Ny. Rasiah, Istriku tersayang Muliana,S.E. dan anak-anakku serta segenap keluarga atas doa dan dukungan serta segala pengorbanan selama proses perjuangan penyelesaian tesis ini.
2. Bapak Dr. Ir. Zahir Zainuddin, M.Sc. dan Bapak Dr.Eng. Zulkifli Tahir, S.T., M.Sc. atas bimbingan, arahan serta waktu yang telah diluangkan kepada penulis dalam proses penyusunan tesis ini mulai dari seminar proposal, *conference* hingga ujian tutup.

3. Segenap Pimpinan, Dosen, staf dan civitas akademika Departemen dan Program studi Teknik Elektro atas bimbingan, pelayanan dan kerjasamanya selama proses akademik.
4. Pimpinan dan rekan-rekan Tim Kerja Sekretariat Jenderal Pendidikan Tinggi Kemdikbudristek maupun Lembaga Layanan Pendidikan Tinggi Wilayah IX atas dukungan dan fasilitasi akses data, juga motivasi dan pengorbanannya dalam mem-*backup* tugas dan tanggung jawab tim selama penulis menekuni proses penelitian.
5. Keluarga NewNormal (Dr. Firman,M.M.),dkk, rekan-rekan *Doctor Candidate* (Pak Budi, Pak Syarif),dkk atas dukungan dan bantuan yang banyak memfasilitasi penyelesaian Tesis maupun administrasi akademik.

Dengan keterbatasan pengalaman, ilmu maupun pustaka yang ditinjau, penulis menyadari bahwa tesis ini masih banyak kekurangan dan pengembangan lanjut agar benar-benar bermanfaat. Oleh sebab itu, penulis sangat mengharapkan kritik dan saran agar tesis ini lebih sempurna serta sebagai masukan bagi penulis untuk penelitian dan penulisan karya ilmiah di masa yang akan datang.

Akhir kata, penulis berharap tesis ini memberikan manfaat bagi kita semua terutama untuk pengembangan ilmu pengetahuan.

Makassar, Juni 2022

Ali Isra

ABSTRAK

Ali Isra. Perbandingan Prediksi Peluang Karir Alumni berdasarkan Kinerja Program Studi Menggunakan Metode C4.5 dan Neural Network

(dibimbing oleh Zahir Zainuddin dan Zulkifli Tahir)

Beberapa indikator kinerja program studi adalah mampu menghasilkan alumni yang terserap dalam dunia kerja dengan masa tunggu mulai bekerja yang cepat, sesuai dengan kompetensi bidang ilmunya (keselarasan horizontal) dan sesuai dengan jenjang pendidikannya (keselarasan vertikal). Untuk mengoptimalkan hal tersebut, diperlukan alat untuk mendeteksi lebih dini peluang keberhasilan program studi berdasarkan parameter kinerja yang sedang diterapkan. Penelitian ini bertujuan untuk memprediksi peluang karir meliputi masa tunggu bekerja, keselarasan horizontal dan vertikal alumni suatu program studi di perguruan tinggi yang merupakan salah satu indikator kinerja utama perguruan tinggi. Metode *Decision Tree* (pohon keputusan) menggunakan Algoritma *C4.5* dan *Neural Network* (Jaringan saraf tiruan) digunakan untuk melakukan klasifikasi pada prediksi karir masing-masing alumni. Penelitian ini menggunakan data latih dan data uji masing-masing sebanyak 25.118 dan 10.766 data alumni. Data bersumber dari laporan tracer study, laporan Kerjasama dan pangkalan data Pendidikan tinggi Kemdikbud Ristek tahun 2016 – 2020 dengan melakukan *query* dan pemrosesan data *json*. Proses pengujian menggunakan *Confusion Matrix Accuracy* dengan hasil pengujian prediksi masa tunggu bekerja, hubungan horizontal dan hubungan vertikal masing-masing 77%, 68% dan 87%.

Kata kunci: *data science*, prediksi, karir, jaringan saraf, masa tunggu, hubungan horizontal, hubungan vertikal

ABSTRACT

Ali Isra. The Comparison of Alumni's Career Opportunities Prediction based on Study Program Performances Using C4.5 and Neural Network Method

(supervised by Zahir Zainuddin and Zulkifli Tahir)

Some of the performance indicators of the study program are capable of producing alumni who are quickly employed in the working industries with a short waiting period, based on their competence in the field of science (horizontal alignment) and according to their level of education (vertical alignment). In order to optimize this, a tool is surely needed to detect early the success opportunity of a study program based on the performance parameters being applied. This study aimed to predict the alumni's career opportunities, namely working waiting period, horizontal and vertical alignment of an alumni of a study program at a university, which is considerably one of the achievement indicators mainly used in higher education institutions. The Decision Tree method used the C4.5 Algorithm and the Neural Network used to classify the career predictions of each alumni. This study utilized both training data and test data as many as 25,188 and 10,776 alumni data respectively. The data was sourced from study tracer reports, Cooperation reports and the database of the Ministry of Education and Technology Higher Education in 2016 – 2020 by querying and processing json data. The testing was carried out using the Confusion Matrix Accuracy whose results revealed that the prediction of the waiting work period, the horizontal and the vertical relationship were 77%, 68% and 87% respectively.

Keywords: data science, prediction, career, neural network, waiting period, horizontal relationship, vertical relationship

DAFTAR ISI

PRAKATA	2
ABSTRAK	4
ABSTRACT.....	5
DAFTAR ISI.....	6
DAFTAR TABEL.....	9
DAFTAR GAMBAR.....	10
BAB I PENDAHULUAN	12
A. LATAR BELAKANG	12
B. RUMUSAN MASALAH	17
C. TUJUAN PENELITIAN	18
D. MANFAAT PENELITIAN	18
E. BATASAN MASALAH.....	19
F. SISTEMATIKA PENULISAN	19
<i>BAB I PENDAHULUAN.....</i>	<i>19</i>
BAB II TINJAUAN PUSTAKA	22
A. LANDASAN TEORI	22
1. <i>Data Mining</i>	22
2. <i>Tracer Study</i>	22
3. <i>Pangkalan Data Pendidikan Tinggi (PDDikti)</i>	24

4. <i>Kerjasama Perguruan Tinggi Kampus Merdeka – Merdeka Belajar</i>	
25	
5. <i>Data Wrangling</i>	28
6. <i>Data Scaling (Min Max Normalization)</i>	29
7. <i>Oversampling</i>	30
8. <i>SMOTE-NC</i>	31
9. <i>Cross Validation</i>	33
10. <i>Jaringan Saraf Tiruan (Neural Network)</i>	34
11. <i>Algoritma C45</i>	37
12. <i>Proses Klasifikasi</i>	38
13. <i>Evaluasi Performansi Metode Klasifikasi</i>	41
14. <i>Python</i>	42
15. <i>Google Colaboratory</i>	42
16. <i>Jupyter Notebook</i>	44
17. <i>Navicat</i>	46
B. PENELITIAN TERKAIT.....	47
C. STATE OF THE ART	49
D. KERANGKA PIKIR	55
BAB III METODE PENELITIAN	57
A. TAHAPAN PENELITIAN	57
B. WAKTU DAN LOKASI PENELITIAN	59
C. INSTRUMENTASI PENELITIAN	59

D. DESAIN SISTEM.....	60
1. <i>Akuisisi Data</i>	61
2. <i>Data Wrangling</i>	67
3. <i>Data Split and Scaling</i>	71
4. <i>Data Balancing</i>	72
5. <i>Training</i>	74
6. <i>Pengujian</i>	76
7. <i>Deployment</i>	77
BAB IV HASIL PENELITIAN DAN PENGUJIAN	79
A. DATASET	79
B. PENGOLAHAN DATA	80
1. <i>Proses Split dan Balancing Data</i>	80
2. <i>Hasil Prediksi</i>	85
C. HASIL PENGUJIAN	97
1. <i>Pengujian Confusion Matrix</i>	97
2. <i>Evaluasi Model</i>	102
3. <i>Deployment</i>	110
BAB V KESIMPULAN DAN SARAN.....	111
A. KESIMPULAN.....	111
B. SARAN.....	112
DAFTAR PUSTAKA.....	113
LAMPIRAN.....	117

DAFTAR TABEL

Tabel 2. 1 <i>State of the art</i>	51
Tabel 3. 1 Tabel <i>Confusion Matrix</i>	76
Tabel 4. 1 Tabel Variabel (X) Dataset	79
Tabel 4. 2 Tabel Data Target (<i>Output</i>) Dataset	80

DAFTAR GAMBAR

Gambar 2. 1 Ilustrasi Kelas Klasifikasi Kelas	39
Gambar 2. 2 Ilustrasi Klasifikasi Multi Kelas.....	40
Gambar 2. 3 Kerangka Pikir	55
Gambar 3. 1 Diagram tahapan penelitian.....	57
Gambar 3. 2 Desain Alur Sistem Prediksi Peluang Karir Alumni.....	60
Gambar 3. 3 Korelasi Data	69
Gambar 4. 1 Komposisi Masing-masing Target (y) Dataset.....	81
Gambar 4. 2 Grafik Perubahan Komposisi Dataset (Masa Tunggu)	81
Gambar 4. 3 Grafik Perubahan Sebaran Dataset Masa Tunggu.....	82
Gambar 4. 4 Grafik Perubahan Komposisi Dataset Hubungan Horisontal.....	83
Gambar 4. 5 Grafik Perubahan Sebaran Dataset Hubungan Horisontal ...	83
Gambar 4. 6 Grafik Perubahan Komposisi Dataset Hubungan Vertikal ...	84
Gambar 4. 7 Grafik Perubahan Sebaran Dataset Hubungan Vertikal	84
Gambar 4. 8 Pohon Keputusan Masa Tunggu	86
Gambar 4. 9 Deskripsi Masa Tunggu	87
Gambar 4. 10 Pohon Keputusan Hubungan Horisontal	89
Gambar 4. 11 Deskripsi Hubungan Horizontal	91
Gambar 4. 12 Pohon Keputusan Hubungan Vertikal.....	93
Gambar 4. 13 Deskripsi Hubungan Vertikal	94
Gambar 4. 14 <i>Confusion Matrix</i> Masa Tunggu dengan metode C4.5	98
Gambar 4. 15 <i>Confusion Matrix</i> Hubungan Horizontal dengan metode C4.5	98
Gambar 4. 16 <i>Confusion Matrix</i> Hubungan Vertikal dengan metode C4.5	99
Gambar 4. 17 <i>Confusion Matrix</i> Masa Tunggu dengan metode ANN	100
Gambar 4. 18 <i>Confusion Matrix</i> Hubungan Horizontal dengan metode ANN	101
Gambar 4. 19 <i>Confusion Matrix</i> Vertikal dengan metode ANN	101
Gambar 4. 20 Hasil Evaluasi Model C4.5 untuk Target Masa Tunggu...	102

Gambar 4. 21 Hasil Evaluasi Model C4.5 untuk Target Hubungan Horizontal	104
Gambar 4. 22 Hasil Evaluasi Model C4.5 untuk Target Hubungan Vertikal	106
Gambar 4. 23 Hasil Evaluasi Model ANN untuk Target Masa Tunggu...	108
Gambar 4. 24 Hasil Evaluasi Model ANN untuk Target Hubungan Horizontal	108
Gambar 4. 25 Hasil Evaluasi Model ANN untuk Target Hubungan Vertikal	108
Gambar 4. 26 Aplikasi yang telah di-deploy dalam versi website.....	110

BAB I

PENDAHULUAN

A. LATAR BELAKANG

Pendidikan di Indonesia melalui Undang-undang Republik Indonesia No 20 Tahun 2003 tentang sistem Pendidikan nasional dalam ketentuan umum pasal 1 (satu) menyebutkan bahwa peserta didik merupakan anggota masyarakat yang berusaha mengembangkan potensi diri melalui proses pembelajaran yang tersedia pada jalur, jenjang, dan jenis pendidikan tertentu salah satunya ditempuh pada Perguruan Tinggi. Perguruan tinggi memiliki peran yang sangat vital untuk mengembangkan potensi diri pada masyarakat sehingga tersedianya Sumber Daya manusia yang unggul, siap bersaing dan mampu beradaptasi dengan era zaman yang terus berkembang.

Pada Agustus tahun 2019 badan pusat statistik merilis berita resmi statistik tentang keadaan ketenagakerjaan Indonesia yang menyebutkan bahwa tingkatan pengangguran terbuka (TPT) mengalami penurunan sebesar 0.06% dari jumlah tingkat pengangguran terbuka di tahun 2018 sebesar 5,34 Persen dan berdasarkan dari hasil data survey tentang ketenagakerjaan, tingkat pengangguran terbuka periode agustus tahun 2015 – 2019, tercatat TPT alumni pendidikan diploma rata-rata 6,49 Persen dan TPT alumni universitas sebanyak 5,60 Persen. Tercatat pula dalam berita acara tersebut bahwa 8,13 juta orang yang bekerja kurang dari 35

jam perminggu dan 28,41 juta orang bekerja paruh waktu dibawah jam kerja normal kurang dari 35 jam perminggu (BPS, 2019).

Di sisi lain, berdasarkan data *tracer study online* yang diselenggarakan oleh Direktorat Pembelajaran dan Kemahasiswaan Kementerian Pendidikan dan Kebudayaan, dari 391.433 laporan alumni yang terserap di dunia kerja hanya menunjukkan 11,70% yang tidak memiliki keselarasan horizontal antara kompetensi bidang ilmu terhadap bidang karir dan 23,86% yang tidak memiliki keselarasan vertikal antara jenjang studi dengan tingkatan karir yang digeluti.

Dengan melihat data-data tersebut tentunya perguruan tinggi negeri maupun swasta dituntut untuk lebih meningkatkan kinerja institusinya masing-masing dari segi akademik dan non akademik dengan tujuan agar para alumni lulusan dari perguruan tinggi mudah terserap di dunia usaha dan dunia industri sehingga kedepannya angka TPT lebih baik dari tahun-tahun sebelumnya. Guna meningkatkan kinerja perguruan tinggi agar sesuai dengan tujuan Pendidikan tinggi sebagaimana dituangkan pada pasal 5 undang-undang No 12 tahun 2012 tentang Pendidikan tinggi, perguruan tinggi oleh pemerintah telah menetapkan standar untuk dipatuhi oleh perguruan tinggi melalui peraturan Menteri pendidikan dan kebudayaan No 03 Tahun 2020 tentang Standar Nasional Pendidikan Tinggi dan peraturan Menteri Pendidikan dan kebudayaan No 05 tahun 2020 tentang Akreditasi Program Studi dan Perguruan Tinggi sebagai alat untuk mengukur kinerja suatu perguruan tinggi.

Tidak dapat dipungkiri bahwasanya berkembangannya zaman, motivasi masyarakat Indonesia untuk merasakan belajar di perguruan tinggi baik pada perguruan tinggi negeri dan swasta sangat tinggi. Karena secara *defacto* yang beredar dikalangan masyarakat bahwasanya dengan kuliah mampu menaikkan taraf kualitas hidup seseorang. Tentunya oleh Perguruan Tinggi disambut secara positif dan juga oleh para elit yang memiliki kelebihan secara finansial berlomba -lomba untuk mendirikan perguruan tinggi dengan tujuan dapat ikut ambil bagian adanya fenomena masyarakat tentang menaikkan kualitas hidup masyarakat sehingga secara *de jure* didapati kompetensi yang dilakukan perguruan tinggi terjadi persaingan yang tidak profesional karena hanya berpusat pada *income* yang lebih tanpa memikirkan kualitas dari alumni yang dihasilkan sehingga perguruan tinggi yang memiliki riwayat pengelolaan patuh terhadap ketentuan pengelolaan perguruan tinggi yang baik dengan status akreditasi baik tetap menjadi primadona oleh masyarakat.

Bentuk perguruan tinggi dengan kinerja akademik yang baik tentunya mampu menghasilkan alumni dengan kualitas yang baik pula karena mampu memprediksikan peluang para alumni untuk dapat langsung diserap oleh dunia usaha dan dunia industri sehingga memberi dampak yang positif yang besar buat perguruan tinggi sendiri (Muhson et al., 2012). Kehadiran perguruan tinggi baik negeri dan swasta yang hadir di tengah-tengah masyarakat, mampu meningkatkan kualitas masyarakat yang melanjutkan Pendidikan di perguruan tinggi tersebut dan juga memiliki kontribusi positif yang baik untuk peningkatan daerah hingga ke tingkat

nasional melalui program *link and match* dengan stakeholder, dunia usaha dan dunia industri. Akan tetapi hal tersebut belum dilaksanakan secara maksimal karena didapati bahwasannya standar yang telah ditetapkan oleh pemerintah tidak dijalankan dengan optimal oleh perguruan tinggi baik negeri maupun perguruan tinggi swasta sehingga berdampak pada kualitas alumni yang dihasilkan.

Kemampuan secara otonom yang dimiliki oleh setiap perguruan tinggi dalam pengelolaannya, harusnya menjadi peluang besar agar berkembang menjadi lebih baik dan berkualitas melalui pembaharuan dan peningkatan kualitas kinerja akademik sehingga perguruan tinggi mampu melakukan prediksi alumni perguruan tersebut setelah lulus. Kinerja akademik yang baik umumnya memenuhi aspek peningkatan sumber daya pendidik dan tenaga kependidikan, sarana prasarana dan kerjasama dengan pihak dunia usaha dan dunia industri yang terukur positif antara kedua pihak sehingga mampu melakukan adaptasi dengan *trend issue* pendidikan abad 21 salah yang saat ini digunakan secara nasional yaitu revolusi industri 4.0 dan 5.0 diseluruh negara maju. Akan tetapi kewenangan otonomi yang dimiliki oleh perguruan tinggi sering salah untuk diterapkan sehingga proses pengelolaannya tidak sesuai dengan kaidah-kaidah standar pendidikan tinggi yang baik dan lebih memilih menerapkan proses pengelolaan yang berorientasi pada mutu lulusan yang siap kerja sehingga menjadi masalah tersendiri dihadapi oleh pemerintah.

Mutu lulusan yang siap kerja yang di proses pada perguruan tinggi dapat diprediksi dengan meningkatkan kinerja akademik yang berorientasi dan

atau sesuai dengan kebutuhan dunia usaha dan dunia industri. Proses prediksi ini tujuannya adalah untuk mendapatkan informasi yang sistematis terukur dan melalui bukti data kinerja akademik yang terpola mengikuti ketentuan standar atau pendapat para ahli dan atau melalui kesepakatan forum ilmiah. Salah satu teknik prediksi yang digunakan adalah dengan cara menggali data atau dikenal dengan *data mining* yang bertujuan untuk menggali data-data yang saling berkorelasi. Teknik prediksi menggunakan *data mining* telah banyak dirasakan manfaatnya untuk menyelesaikan kasus - kasus di berbagai bidang ilmu salah satu diantaranya adalah kasus yang datanya menggunakan data-data Pendidikan seperti penelitian yang dilakukan oleh Tair dan kawan-kawan tahun 2012 dengan mengangkat topik tentang study kasus mengetahui kemampuan individu siswa melalui data Pendidikan dan setelah diperoleh data-data tersebut, dilanjutkan proses Analisis menggunakan *data mining* sehingga dapat dilakukan peningkatan kinerja akademik siswa tersebut (Tair and El-halees, 2012).

Penelitian ini tentang prediksi karir alumni pada perguruan tinggi yang dilakukan dalam penelitian ini menggunakan data mining dengan metode yang digunakan adalah C45 dan Neural Network. Penelitian-penelitian tersebut diantaranya adalah pemberian rekomendasi yang menggabungkan situasi aktual perguruan tinggi dengan menggunakan kumpulan data tipikal dan metode pohon keputusan yang melakukan pengurangan atribut berdasarkan teori himpunan kasar untuk mengurangi beban komputasi algoritma (Xinjuan and Jing, 2013a). Model prediksi karir menggunakan data mining dan klasifikasi linear juga merupakan

pendekatan yang realistis untuk konseling karena mempertimbangkan baik kepribadian dan bakat, yang bertanggung jawab untuk keputusan karir (Rangnekar et al., 2018a).

Berdasarkan masalah yang telah diuraikan pada paragraf sebelumnya serta perolehan data berdasarkan sumber yang eligible, penelusuran sumber referensi untuk melihat keterkaitan metode yang digunakan sesuai dengan penelitian yang akan dilakukan. Oleh karena itu dalam penelitian ini akan dilakukan PERBANDINGAN PREDIKSI PELUANG KARIR ALUMNI BERDASARKAN KINERJA PROGRAM STUDI MENGGUNAKAN METODE C4.5 DAN NEURAL NETWORK. Dengan dilakukan prediksi ini, diharapkan dapat meningkatkan efektifitas dan efisiensi prediksi karir lulusan, membantu perguruan tinggi dalam melakukan peningkatan mutu penyelenggaraan akademik yang diselenggarakan program studi dalam rangka penguatan kompetensi mahasiswa, serta dapat mendukung pemerintah dalam menentukan arah kebijakan terkait penyelenggaraan pendidikan tinggi.

B. RUMUSAN MASALAH

Rumusan masalah pada penelitian ini adalah:

1. Bagaimana mengimplementasikan metode Decision tree C4.5 untuk klasifikasi karir alumni program studi tertentu berdasarkan masa tunggu bekerja, hubungan horizontal dan hubungan vertikal?

2. Bagaimana mengimplementasikan Neural Network untuk klasifikasi karir alumni program studi tertentu berdasarkan masa tunggu bekerja, hubungan horizontal dan hubungan vertikal?

C. TUJUAN PENELITIAN

Tujuan penelitian ini adalah sebagai berikut:

1. Membuat sistem prediksi peluang karir alumni meliputi masa tunggu bekerja, hubungan horizontal dan hubungan vertikal menggunakan metode Decision Tree C4.5
2. Membuat sistem prediksi peluang karir alumni meliputi masa tunggu bekerja, hubungan horizontal dan hubungan vertikal menggunakan metode Neural Network

D. MANFAAT PENELITIAN

Manfaat penelitian ini adalah sebagai berikut:

1. Manfaat Teoritis
 - a. Untuk menambah kajian ilmiah tentang algoritma klasifikasi dan algoritma prediksi
 - b. Menambah pengetahuan serta dapat mengaplikasikan algoritma *decision tree* (pohon keputusan) dan *Neural Network* (jaringan saraf tiruan) untuk pengembangan penelitian.
2. Manfaat Praktis

Sebagai referensi program studi dalam upaya peningkatan mutu kinerja akademik sehingga Alumni yang dihasilkan program studi memiliki peluang yang besar dapat terserap pada dunia usaha dan

dunia industri dengan bidang dan tingkatan yang selaras dengan kompetensi yang dibentuk oleh program studi.

E. BATASAN MASALAH

Batasan masalah pada penelitian ini adalah sebagai berikut:

1. Data yang digunakan dalam penelitian ini adalah data laporan *tracer study*, laporan Kerjasama dan laporan akademik program studi untuk jenjang S1, D4 dan D3 pada perguruan tinggi swasta di lingkungan LLDIKTI Wilayah III, V, VII dan IX.
2. Objek penelitian yaitu Menghitung Peluang Karir Alumni dalam masa tunggu mendapatkan pekerjaan, keselarasan horizontal antara bidang ilmu dengan bidang pekerjaan dan keselarasan tingkat jenjang Pendidikan dengan jenjang pekerjaan.
3. Pengambilan data dilakukan pada bulan Januari 2020

F. SISTEMATIKA PENULISAN

Adapun sistematika penulisan pada penelitian ini adalah sebagai berikut:

BAB I PENDAHULUAN

Bab I berisi penjelasan tentang latar belakang penelitian dengan objek data *tracer study* dan pangkalan data Pendidikan tinggi.

BAB II TINJAUAN PUSTAKA

Bab II berisi penjelasan tentang tinjauan pustaka yang merupakan penjelasan tentang hasil-hasil penelitian yang relevan dengan penelitian yang dilakukan berupa sumber acuan terbaru dari pustaka primer seperti

buku, artikel, jurnal, regulasi serta tulisan asli lainnya. Dalam bab ini juga diuraikan tentang kerangka pikir tentang masalah, metode, pengukuran dan hasil dari penggunaan metode yang sesuai dengan objek dalam penelitian yang diusulkan

BAB III METODE PENELITIAN

Bab III ini merupakan penjelasan tentang tahapan penelitian, bagaimana pengembangan dan penerapan machine learning dengan menggunakan algoritma Fuzzy yang diusulkan penulis untuk memprediksi peluang karir alumni berdasarkan data tracer study dan pangkalan data pendidikan ting

BAB IV HASIL DAN PEMBAHASAN

Pada bab IV ini menjelaskan tentang hasil dan pembahasan penelitian serta implikasi dari penelitian yang dilakukan. Hasil merupakan suatu penjelasan tentang data kuantitatif yang dikumpulkan sesuai dengan metodologi yang telah ditetapkan. Pembahasan merupakan suatu penjelasan tentang pengolahan data dan interpretasinya, baik dalam bentuk deskriptif maupun penarikan inferensinya. Implikasi penelitian merupakan suatu penjelasan tentang tindak lanjut penelitian yang terkait dengan aspek sistem, maupun aspek penelitian lanjutan.

BAB V KESIMPULAN DAN SARAN

Pada bab V ini berisi ringkasan temuan, rangkuman kesimpulan dan saran. Kesimpulan merupakan pernyataan secara general atau spesifik yang berisi hal-hal penting dan menjadi temuan penelitian yang bersumber pada

hasil dan pembahasan. Saran merupakan pernyataan atau rekomendasi peneliti yang berisi hal-hal penting sebagaimana yang telah disampaikan.

BAB II

TINJAUAN PUSTAKA

A. LANDASAN TEORI

1. Data Mining

Data Mining adalah suatu istilah yang digunakan untuk menguraikan penemuan pengetahuan di dalam basis data. Data Mining adalah proses yang menggunakan teknik statistik, matematika, kecerdasan buatan, machine learning untuk mengekstraksi dan mengidentifikasi informasi yang bermanfaat dan pengetahuan yang terkait dari berbagai basis data besar (Kusrini and Taufiq Luthfi, 2019).

Data Mining adalah salah satu bidang yang berkembang pesat karena besarnya kebutuhan akan nilai tambah dari database skala besar yang makin banyak terakumulasi sejalan dengan pertumbuhan teknologi informasi.

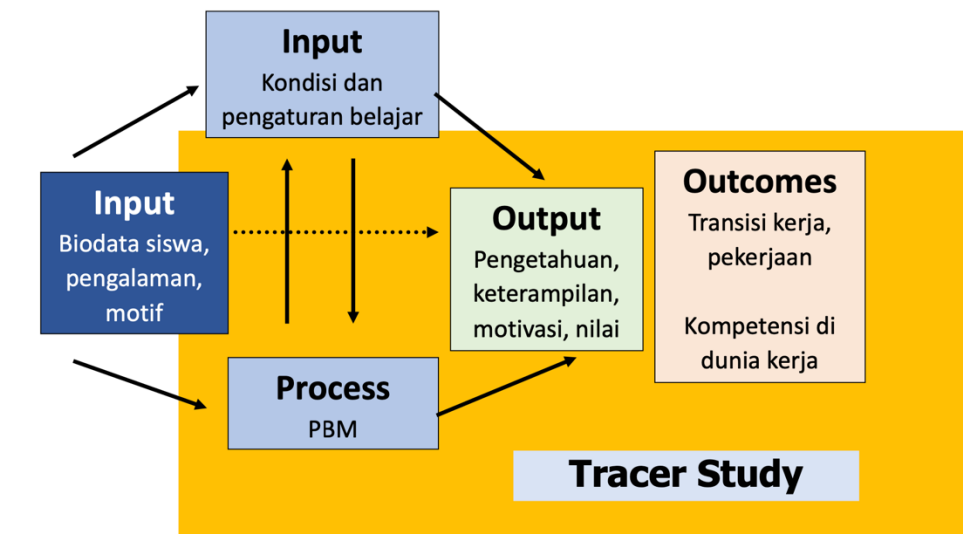
2. Tracer Study.

Tracer study online Dikti ditujukan untuk melacak jejak lulusan/alumni yang dilakukan 2 tahun setelah lulus dan bertujuan untuk mengetahui:

- a. Outcome pendidikan dalam bentuk transisi dari dunia pendidikan tinggi ke dunia kerja (termasuk masa tunggu kerja dan proses pencarian kerja pertama), situasi kerja terakhir, dan aplikasi kompetensi di dunia kerja.

- b. Output pendidikan yaitu penilaian diri terhadap penguasaan dan pemerolehan kompetensi.
- c. Proses pendidikan berupa evaluasi proses pembelajaran dan kontribusi pendidikan tinggi terhadap pemerolehan kompetensi

Hasil tracer study akan membantu PT dalam mengetahui posisi lulusan yang telah terserap dalam dunia kerja serta menyiapkan lulusan sesuai dengan kompetensi yang diperlukan di dunia kerja. Hasil tracer study yang kemudian dilaporkan ke Dikti akan membantu program Pemerintah dalam rangka memetakan kebutuhan dunia kerja dengan pembangunan pendidikan di Indonesia (Direktorat Pembelajaran dan Kemahasiswaan, 2020).



Gambar 2. 1 Framework Tracer Study

tracerstudy.kemdikbud.go.id/index.php/formkuis

f5b Apa nama perusahaan/kantor tempat Anda bekerja? :

f5c Bila berwiraswasta, apa posisi/jabatan Anda saat ini? (Apabila F8 menjawab [3] Wiraswasta) :

f5d Apa tingkat tempat kerja Anda? :

f18 Pertanyaan studi lanjut (Apabila F8 menjawab [4] Melanjutkan Pendidikan) : Sumber Biaya : Perguruan Tinggi : Program Studi : Tanggal Masuk : dd/mm/yyyy

f12 Sebutkan sumberdana dalam pembiayaan kuliah? (bukan ketika Studi Lanjut) :

f14 Seberapa erat hubungan antara bidang studi dengan pekerjaan anda? : [1] Sangat Erat [2] Erat [3] Cukup Erat [4] Kurang Erat [5] Tidak Sama Sekali

f15 Tingkat pendidikan apa yang paling tepat/sesuai untuk pekerjaan anda saat ini? : [1] Setingkat Lebih Tinggi [2] Tingkat yang Sama [3] Setingkat Lebih Rendah [4] Tidak Perlu Pendidikan Tinggi

f17 Pada saat lulus, pada tingkat mana kompetensi di bawah ini anda kuasai? (A) Pada saat ini, pada tingkat mana kompetensi di bawah ini diperlukan dalam pekerjaan? (B)

A		B	
Sangat Rendah	Sangat Tinggi	Sangat Rendah	Sangat Tinggi

Gambar 2. 2 Beberapa Pertanyaan Kuisioner Tracer Study

3. Pangkalan Data Pendidikan Tinggi (PDDikti)

Pangkalan Data Pendidikan Tinggi merupakan kumpulan data penyelenggaraan Pendidikan Tinggi seluruh Perguruan Tinggi yang terintegrasi secara nasional (Indonesia, 2012). Pangkalan Data Pendidikan Tinggi berfungsi sebagai sumber informasi bagi:

- lembaga akreditasi, untuk melakukan akreditasi Program Studi dan Perguruan Tinggi;
- Pemerintah, untuk melakukan pengaturan, perencanaan, pengawasan, pemantauan, dan evaluasi serta pembinaan dan koordinasi Program Studi dan Perguruan Tinggi; dan
- Masyarakat, untuk mengetahui kinerja Program Studi dan Perguruan Tinggi.

PDDikti menghimpun Data Pendidikan Tinggi yang dikelola dengan kaidah basis data terintegrasi. Basis data terintegrasi dimaksud mencatat hubungan antar entitas data dengan menjaga kelengkapan dan kebenaran data, sehingga hubungan antar entitas data dapat dihasilkan dari pengolahan data secara langsung tanpa melakukan pemadanan/pemetaan antar entitas data secara manual. Data Pendidikan Tinggi tersebut terdiri atas Data Pokok Pendidikan Tinggi; Data Referensi Pendidikan Tinggi; dan Data Transaksional Pendidikan Tinggi.

Perguruan Tinggi wajib menyampaikan data penyelenggaraan pendidikan tinggi yang valid ke PDDikti dan Pimpinan Perguruan Tinggi bertanggung jawab atas kelengkapan, kebenaran, ketepatan, dan kemutakhiran data penyelenggaraan pendidikan tinggi yang dilaporkan ke PDDikti. Perguruan Tinggi yang memasukkan data penyelenggaraan pendidikan tinggi yang tidak valid ke PDDikti dikenai sanksi sesuai dengan ketentuan peraturan perundang-undangan (Kementerian Pendidikan, 2016).

4. Kerjasama Perguruan Tinggi Kampus Merdeka – Merdeka Belajar

Program Kerja Sama Pendidikan Tinggi adalah program-program pendidikan tinggi yang dikembangkan oleh dua atau lebih perguruan tinggi (PT) melalui kerja sama antar perguruan tinggi di dalam negeri (PTDN) atau antara PT di dalam negeri (PTDN) dan PT di luar negeri (PTLN). Tujuan Program Kerja Sama adalah meningkatkan efektivitas, efisiensi, produktivitas, kreativitas, inovasi, mutu, dan relevansi

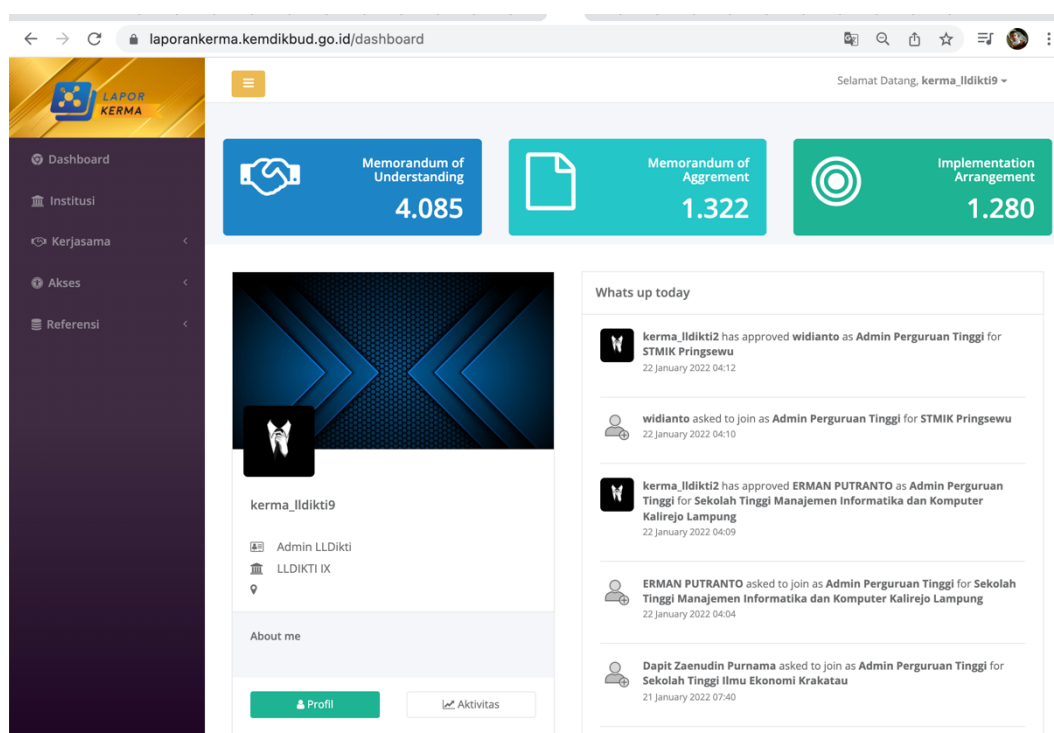
pelaksanaan Tridarma Perguruan Tinggi untuk meningkatkan daya saing bangsa (Dirjen Dikti Kemdikbud, 2018).

Merdeka Belajar – Kampus Merdeka (MBKM), merupakan kebijakan Menteri Pendidikan dan Kebudayaan, yang bertujuan mendorong mahasiswa untuk menguasai berbagai keilmuan yang berguna untuk memasuki dunia kerja. Kampus Merdeka memberikan kesempatan bagi mahasiswa untuk memilih mata kuliah yang akan mereka ambil.

Salah satu bentuk Kerjasama dalam kegiatan MBKM adalah Program magang 1-2 semester, memberikan pengalaman yang cukup kepada mahasiswa, pembelajaran langsung di tempat kerja (*experiential learning*). Selama magang mahasiswa akan mendapatkan *hardskills* (keterampilan, *complex problem solving*, *analytical skills*, dsb.), maupun *soft skills* (etika profesi/kerja, komunikasi, kerjasama, dsb.). Sementara industri mendapatkan talenta yang bila cocok nantinya bisa langsung di-recruit, sehingga mengurangi biaya recruitment dan training awal/ induksi. Mahasiswa yang sudah mengenal tempat kerja tersebut akan lebih mantab dalam memasuki dunia kerja dan karirnya. Melalui kegiatan ini, permasalahan industri akan mengalir ke perguruan tinggi sehingga meng-update bahan ajar dan pembelajaran dosen serta topik-topik riset di perguruan tinggi akan makin relevan (Dirjen Dikti Kemdikbud, 2020).

Sekretariat Direktorat Jenderal Pendidikan Tinggi mempunyai rincian tugas melaksanakan penyusunan program kerja Sekretariat Direktorat Jenderal dan konsep program kerja Direktorat Jenderal; melaksanakan pengumpulan dan pengolahan data dan informasi di bidang pendidikan

tinggi; melaksanakan penyusunan dan penyajian data dan informasi di bidang pendidikan tinggi; dan melaksanakan pengelolaan dan pengembangan system informasi manajemen di bidang pendidikan tinggi. Perguruan tinggi dapat melakukan pengelolaan laporan kerja sama melalui Sistem pelaporan kerjasama (LAPORKERMA) melalui laman website <https://laporankerma.kemdikbud.go.id/>.



Gambar 2. 3 Laman Sistem Laporankerma

Selain sistem pelaporan yang digunakan oleh perguruan tinggi melaporkan program kegiatan Kerjasama, Direktorat Pendidikan Tinggi juga menyediakan layanan data kerjasasama untuk keperluan analisis data Kerjasama yang interoperabilitas (Direktorat Kelembagaan Kemdikbud, 2020).

The screenshot shows the Postman API documentation for 'LAPORKERMA'. The sidebar on the left lists various endpoints, including 'Introduction', 'Token: Request', 'Kerma', 'Mitra', 'Instansi: PT', 'Program: Sasaran', 'Program: Indikator', 'Ref: Jenis Dokumen', 'Ref: Klasifikasi Mitra', 'Ref: Bentuk Kegiatan', 'Ref: Sumber Dana', 'Ref: Status Kerjasama', 'Ref: Status Verval', and 'Ref: Negara'. The main content area displays two endpoints:

- GET Token: Request** (Basic Auth):
 - Endpoint: `https://laporankerma.kemdikbud.go.id/api/token`
 - Description: Endpoint untuk get token
 - Authorization: Basic Auth
 - Fields: Username (<username>), Password (<password>)
- GET Kerma** (Bearer Token):
 - Endpoint: `https://laporankerma.kemdikbud.go.id/api/kerma?page=1&limit=100&jenis-dokumen=5&update-after=2021-01-01`
 - Authorization: Bearer Token

Gambar 2. 4 Laman dokumentasi penggunaan API laporan kerjasama

5. Data Wrangling

Data wrangling didefinisikan sebagai seni transformasi data atau persiapan data. Data wrangling adalah metode yang disesuaikan untuk manajemen data dasar yang akan diproses, dibentuk, dan disediakan dengan benar untuk konsumsi data yang paling nyaman oleh calon pengguna di masa depan. Data historis yang besar dikumpulkan atau disimpan sebagai fakta atau dimensi dalam gudang data untuk mengakomodasi kueri adhoc yang besar. Data wrangling memungkinkan

pemrosesan kueri yang cepat dengan solusi yang tepat bagi analis dan pengguna akhir. Wrangler menyediakan bahasa interaktif dan merekomendasikan skrip transformasi prediktif. Metode ini membantu pengguna untuk memiliki wawasan tentang pengurangan proses berulang manual (Patil and Hiremath, 2018).

Data merupakan bagian mendasar dari proses operasional di perusahaan dan organisasi ilmiah. Selain itu, mereka membentuk dasar untuk keputusan. Kualitas data yang buruk dapat memengaruhi keputusan secara negatif dan berdampak negatif pada hasil. Kualitas data sangat penting. Tujuan dari data wrangling adalah untuk menyiapkan banyak data mentah dalam keadaan aslinya sehingga dapat digunakan untuk langkah analisis lebih lanjut. Hanya dengan demikian dapat diperoleh pengetahuan yang dapat memberikan nilai tambah. Penelitian ini menunjukkan bagaimana proses data wrangling bekerja dan bagaimana hal itu dapat digunakan dalam sistem database untuk membersihkan data dari sumber data yang heterogen selama akuisisi dan integrasi (Azeroual, 2020).

6. Data Scaling (Min Max Normalization)

Data scaling memiliki peran penting dalam praproses data yang berdampak pada kinerja dari algoritme machine learning (Ambarwari et al., 2020).

Min Max Normalization adalah salah satu metode yang akan mempermudah perbandingan nilai yang memiliki satuan ukuran yang berbeda (Virmani et al., n.d.), dimana varian variabel yang ada pada

penelitian ini menggunakan satuan ukuran yang berbeda. rumus Normalisasi Min – Max.:

$$z = \frac{x - \min(x)}{[\max(x) - \min(x)]}$$

Keterangan :

z = hasil normalisasi,

x = nilai x (asli),

min(x) = nilai minimal untuk variabel x,

7. Oversampling

Oversampling merupakan metode pembangkitan data minoritas sebanyak data mayoritas. Teknik smote oversampling hanya berfungsi untuk kumpulan data dengan semua fitur berkelanjutan. Oversampling merupakan salah satu cara untuk menangani permasalahan imbalanced dengan melakukan pendistribusian data yang seimbang dengan cara replikasi instance (data sintetik) minoritas secara acak dengan melakukan iterasi (Untoro and Buliali, 2018). Penggunaan teknik oversampling merupakan salah satu cara untuk mengatasi ketidak seimbangan data, namun hanya sedikit yang digunakan untuk mengatasi ketidak seimbangan data (Wibowo and Fatichah, 2021). Penambahan teknik oversampling akan menyeimbangkan data setiap kelas untuk memberikan hasil evaluasi yang tidak bias dalam pemodelan.

8. SMOTE-NC

SMOTE-NC merupakan salah satu teknik oversampling yang berbasis SMOTE untuk melakukan keseimbangan data (Mukherjee and Khushi, 2021). SMOTE-NC merupakan SMOTE untuk data yang terdiri dari peubah numerik dan kategorik (SMOTE-Nominal Continuous). Penentuan jarak terdekat dihitung dengan jarak Euclidean dengan nilai median dari simpangan baku semua peubah numerik kelas minoritas sebagai selisih nilai peubah kategorik (Wijaya et al., 2018). Pada SMOTE-NC, jika label atribut kategoris berbeda antara instance dan tetangga terdekatnya, maka nilai tetap, yaitu, median standar deviasi fitur kontinu, ditambahkan dalam perhitungan jarak. Oleh karena itu, dalam metode ini, untuk fitur kategoris multi-label, jarak antara dua label adalah sama.

Berdasarkan penelitian (Chawla et al., 2002) menggunakan data Adult dari repository UCI, dataset tersebut mempunyai 6 kontinyu variabel dan 8 nominal variabel. Algoritma SMOTE dan SMOTE-NC digunakan untuk mengevaluasi dataset tersebut. Algoritma SMOTE-NC dijelaskan sebagai berikut:

- a. Perhitungan Median: Hitung median standar deviasi semua variabel kontinyu untuk kelas minoritas. Jika variabel nominal berbeda antara sampel dan potensi tetangga terdekatnya, maka median ini termasuk dalam jarak Euclidean. Median digunakan untuk mengesampingkan perbedaan variabel nominal dengan suatu jumlah yang terkait dengan perbedaan khas dalam variabel kontinyu.

- b. Perhitungan Tetangga Terdekat: Hitung jarak Euclidean antara variabel vektor yang mana k tetangga terdekat yang sedang diidentifikasi (sampel kelas minor) dan variabel vektor lainnya (sampel kelas minoritas) menggunakan ruang variabel kontinyu. Untuk setiap variabel nominal yang berbeda antara variabel vektor yang dipertimbangkan dan calon tetangga terdekatnya, termasuk median standar deviasi yang dihitung sebelumnya, dalam perhitungan jarak Euclidean. Gambar 2.2 menunjukkan contoh perhitungannya.
- c. Membuat Synthetic Sampel: Variabel kontinyu dari data synthetic yang baru untuk kelas minoritas dibuat menggunakan pendekatan SMOTE yang sama seperti penjelasan sebelumnya. Variabel nominal diberikan nilai yang sering terjadi di sebagian besar k tetangga terdekat.

Misalkan diberikan sampel untuk menghitung tetangga terdekat

F1 = 1 2 3 A B C

F2 = 4 6 5 A D E

F3 = 3 5 6 A B K

Maka, jarak Euclidean antara F2 dan F1 adalah sebagai berikut:

$$Euclidean = \sqrt{(4 - 1)^2 + (6 - 2)^2 + (5 - 3)^2 + Med^2 + Med^2}$$

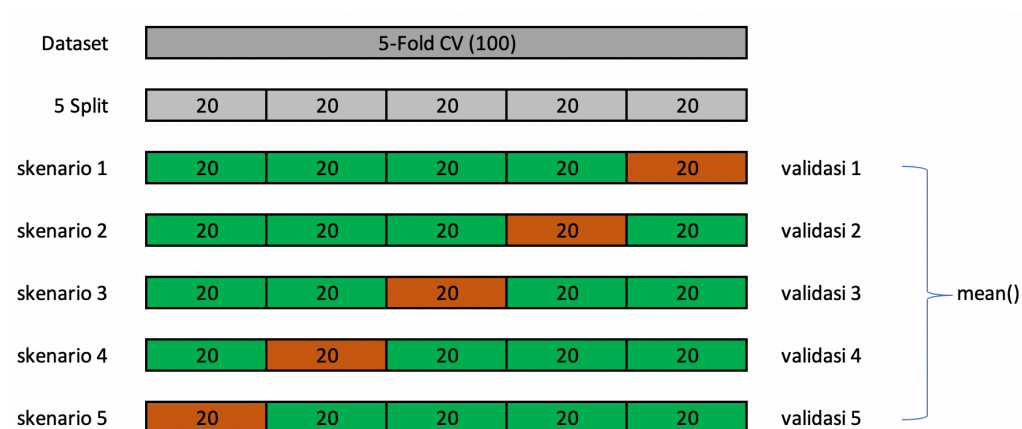
Med adalah nilai median dari nilai standar deviasi variabel kontinyu dari kelas minoritas. Istilah median disebut dua kali untuk variabel nomer 5:

B -> D, dan nomer 6: C -> E yang mana berbeda untuk dua vektor variabel F1 dan F2.

9. Cross Validation

Cross Validation merupakan metode tambahan dari teknik data mining yang mempunyai tujuan untuk menghasilkan akurasi yang lebih maksimal, metode ini juga sering disebut *K-Fold Cross Validation* dimana percobaan dimana percobaan sebanyak K Kalo untuk satu model dengan parameter yang sama. *k-Fold Cross Validation (k-FCV)*. Validasi silang bersifat statistik, dapat digunakan dalam memilih model untuk memprediksi kesalahan uji model prediksi yang lebih baik. *k-Fold Cross Validation* merupakan teknik terbaik yang dapat digunakan pada setiap kasus dan melengkapi pilihan metode untuk iterasi adaptasi selanjutnya (Tembusai et al., 2021).

Konsep *K-Fold Cross Validation* dapat dilihat pada gambar 2.5:



Gambar 2. 5 Konsep 5-Fold Cross Validation

Gambar 2.5 menunjukkan jumlah dataset sebanyak 100 data yang akan divalidasi sebanyak 5 kali, sehingga data akan dipecah dalam 4 kelompok data. Dalam proses pelatihan pembagian data latih dan data uji akan dilakukan secara bergantian pada 5 skenario. Kelompok data berwarna hijau merupakan data latih dan kelompok data berwarna oranye merupakan data uji. Setiap skenario akan dilakukan proses validasi, dan nilai skor (CV score) dari proses ini ditentukan melalui nilai rata-rata (mean) dari 5 hasil validasi dari setiap skenario yang dilakukan.

10. Jaringan Saraf Tiruan (Neural Network)

Menurut (Sinaga, 2012) Jaringan Saraf Tiruan adalah merupakan salah satu representasi buatan dari otak manusia yang selalu mencoba untuk mensimulasikan proses pembelajaran otak manusia tersebut. Jaringan syaraf tiruan (JST) atau yang biasa disebut Artificial Neural Network (ANN) atau Neural Network (NN) saja, merupakan sistem pemroses informasi yang memiliki karakteristik mirip dengan jaringan syaraf pada makhluk hidup. Neural network berupa suatu model sederhana dari suatu syaraf nyata dalam otak manusia seperti suatu unit threshold yang biner. Jaringan Syaraf Tiruan tercipta sebagai suatu generalisasi model matematis dari pemahaman manusia (human cognition) yang didasarkan atas asumsi sebagai berikut:

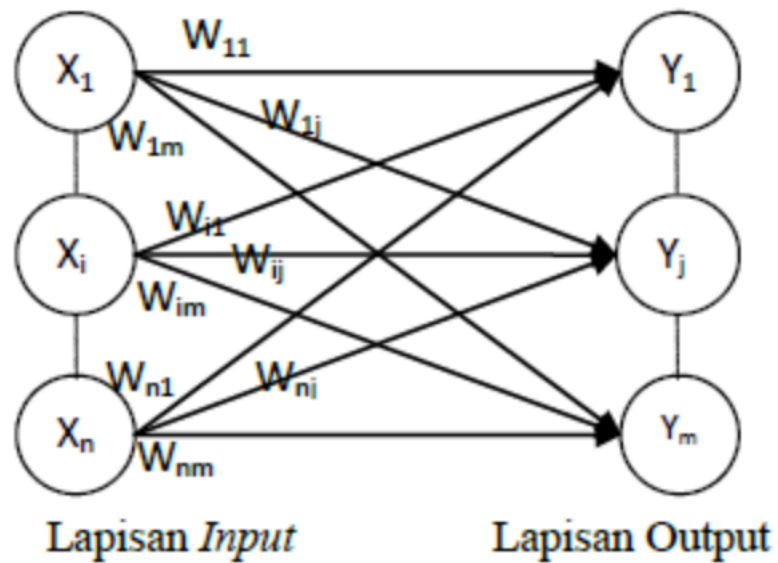
- a. Pemrosesan informasi terjadi pada elemen sederhana yang disebut neuron.
- b. Isyarat mengalir di antara sel syaraf atau neuron melalui penghubung.

- c. Setiap sambungan penghubung memiliki bobot yang bersesuaian.
- d. Setiap sel saraf akan merupakan fungsi aktivasi terhadap isyarat hasil penjumlahan berbobot yang masuk kepadanya untuk menentukan isyarat keluarannya.

Pada Artificial Neural Network, neuron diasumsikan dapat dikelompokkan dalam layer, dan umumnya neuron yang berada dalam satu layer akan memiliki kelakuan yang sama. Pengelompokan neuron ke dalam layer dan pola yang saling terhubung antar layer dikenal sebagai arsitektur jaringan. Arsitektur dari Artificial Neural Network antara lain

a) Jaringan lapis tunggal (single layer network)

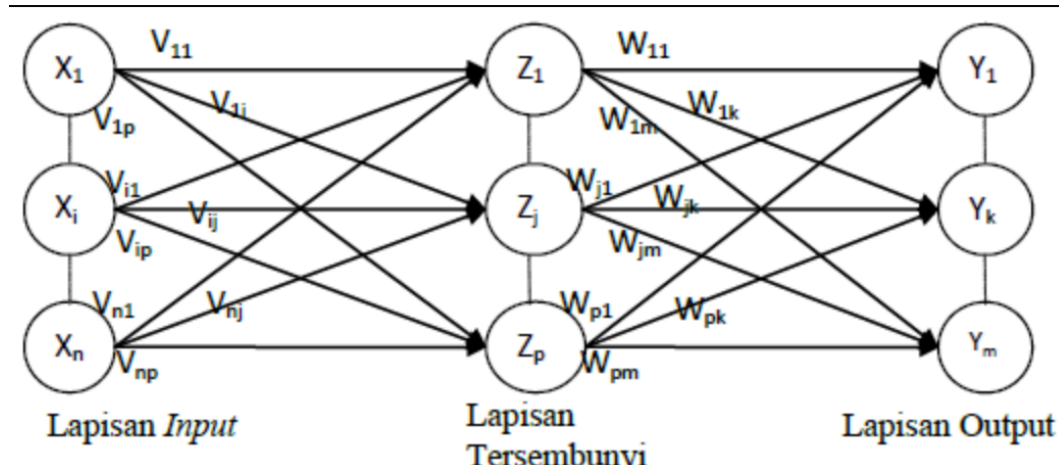
Jaringan lapis tunggal terdiri dari satu layer dari bobot yang saling terhubung. Pada jaringan lapis tunggal sinyal masukan yang berasal dari dunia luar akan langsung diproses menjadi sinyal keluaran yang berisi respon tanpa adanya layer tersembunyi. Arsitektur Jaringan Lapis Tunggal dapat dilihat pada gambar 2.6.



Gambar 2. 6 Arsitektur Jaringan Lapis Tunggal

b) Jaringan lapis jamak (multilayer network)

Jaringan lapis jamak terdiri dari satu atau lebih layer yang sering disebut layer tersembunyi (hidden layer). Layer ini terletak antara layer masuk dan layer keluar. Jaringan lapis jamak dengan menggunakan fungsi aktivasi nonlinear dapat menyelesaikan berbagai macam tipe permasalahan, tetapi proses pembelajaran yang dilakukan akan lebih sulit dan kompleks. Arsitektur jaringan lapis jamak dapat dilihat pada gambar 2.7



Gambar 2. 7 Arsitektur Jaringan Lapis Jamak

11. Algoritma C45

Algoritma C4.5 adalah algoritma yang sudah cukup terkenal dan merupakan salah satu algoritma dalam Data Mining. Salah satu kegunaannya adalah untuk membentuk sebuah pohon keputusan (Decision Tree). Pohon Keputusan sendiri merupakan metode klasifikasi dan prediksi yang digunakan untuk mencari data dan menemukan hubungan yang tersimpan dari variabel atau atribut yang digunakan dan sebuah variabel target yang biasa disebut class atau label. Algoritma yang sering digunakan untuk membuat sebuah pohon keputusan adalah ID3, algoritma C4.5, dan CART (Saputra Ginting and Taufiq Luthfi, 2020).

Algoritma C4.5 sendiri merupakan hasil dari pengembangan algoritma ID3, dimana proses pada algoritma C4.5 akan membentuk sebuah pohon keputusan, mengubah model pohon menjadi rules, kemudian rules tersebut disederhanakan kembali. Pembuatan decision

tree dengan algoritma C4.5 digunakan untuk membangun sebuah pohon keputusan yang dimulai dari pemilihan variabel atau atribut sebagai akar, membangun cabang untuk tiap nilai, membagi kasus dalam cabang kemudian melakukan pengulangan proses untuk setiap cabang sampai seluruh kasus pada cabang mempunyai kelas yang sama

12. Proses Klasifikasi

Klasifikasi dokumen adalah pemberian kategori yang telah didefinisikan kepada dokumen yang belum memiliki kategori (Goller et al., 2000). Mengklasifikasi dokumen merupakan salah satu cara untuk mengorganisasikan dokumen. Dokumen-dokumen yang memiliki isi yang sama akan dikelompokkan ke dalam kategori yang sama. Dengan demikian, orang-orang yang melakukan pencarian informasi dapat dengan mudah melewatkan kategori yang tidak relevan dengan informasi yang dicari atau yang tidak menarik perhatian (Feldman and Gross, 2004). Pada penelitian ini, klasifikasi diterapkan untuk mengkategorikan data mahasiswa yang berpotensi memiliki kinerja akademik baik dan tidak baik.

Klasifikasi adalah proses untuk menemukan model atau fungsi yang menjelaskan atau membedakan konsep atau kelas data dengan tujuan untuk memperkirakan kelas yang tidak diketahui dari suatu objek. Adapun dalam pengklasifikasian data terdapat dua proses yang dilakukan yaitu:

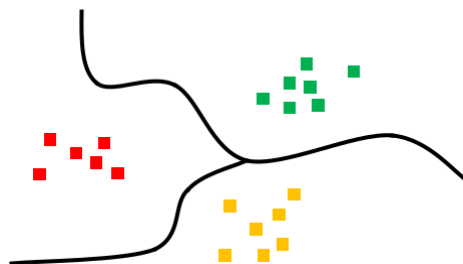
1. Proses *Training*

Pada proses *training* digunakan *training* set yang telah diketahui label-labelnya untuk membangun model atau fungsi.

2. Proses *Testing*

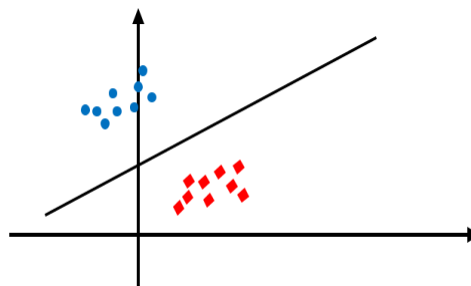
Untuk mengetahui keakuratan model atau fungsi yang akan dibangun pada proses training, maka digunakan data yang disebut dengan *testing set* untuk memprediksi label-labelnya.

Menurut Santosa (2007), klasifikasi dibagi menjadi dua macam yakni klasifikasi dua kelas dan multi kelas. Yang pertama, klasifikasi dua kelas dapat dijelaskan sebagai berikut. Misalkan kita memiliki set data training $(x_i, y_i), i = 1, \dots, n$, dengan data input $x = \{1, 2, \dots, n\} \subseteq \mathbb{R}$ dan output yang bersangkutan $y = \{1, \dots, n\} \subseteq \{\pm\}$. Tujuan dari klasifikasi dua kelas adalah menemukan suatu fungsi keputusan (decision function) $f(x)$ yang secara akurat memprediksi kelas dari data test (x, y) yang berasal dari fungsi distribusi yang sama dengan data untuk training, lihat Gambar 2.8. Set data $(x_i, y_i), i = 1, \dots, n$, biasa dinamakan set training, dimana x_i berkaitan dengan parameter input dan y_i menunjukkan parameter output (Santosa, 2007).



Gambar 2. 8 Ilustrasi Kelas Klasifikasi Kelas

Untuk klasifikasi Multi Kelas, misalkan kita memiliki set data untuk training $(x_i, y_i), i = 1, \dots, n$, dengan data input $x_i = \{x_1, x_2, \dots, x_n\} \subseteq \mathbb{R}$ dan output yang bersangkutan $y_i = \{y_1, y_2, \dots, y_n\} \subseteq \{1, 2, \dots, k\}$. Terlihat bahwa output Y tidak lagi terbatas ± 1 seperti dalam kasus dua kelas.



Gambar 2. 9 Ilustrasi Klasifikasi Multi Kelas

Output dari data kita bisa 1,2,3,4 atau bahkan 10. Ilustrasi ditampilkan dalam Gambar 2.9 Dalam hal ini kita harus mengelompokkan objek yang kita pelajari ke dalam lebih dari dua kelas atau ke dalam k kelas dimana nilai k lebih dari 2. Dalam kasus demikian tugas klasifikasi menjadi lebih rumit dan perlu teknik khusus untuk mengatasinya.

Peningkatan jumlah dokumen yang begitu pesat telah mendorong berkembangnya metode pengklasifikasian secara otomatis. Metode-metode klasifikasi ini dapat melakukan klasifikasi dengan cara belajar dari sekumpulan contoh dokumen yang telah diklasifikasi sebelumnya. Keuntungan dari metode ini adalah dapat menghemat waktu kerja dan memperoleh hasil yang lebih baik.

13. Evaluasi Performansi Metode Klasifikasi

Akurasi klasifikasi merupakan ukuran ketepatan klasifikasi yang menunjukkan performansi teknik klasifikasi secara keseluruhan (Nugroho et al., 2003). Semakin tinggi akurasi klasifikasi berarti performansi teknik klasifikasi juga semakin baik. Permasalahan pada klasifikasi biner, akurasi klasifikasi disajikan pada:

Tabel 2. 1 Akurasi Validasi

Aktual	Prediksi	
	Positif	Negatif
Positif	TP	FN
Negatif	FP	TN

Keterangan :

TP : True Positive (Jumlah prediksi benar pada kelas positif)

FP : False Positive (Jumlah prediksi salah pada kelas positif)

FN : False Negative (Jumlah prediksi salah pada kelas negatif)

TN : True Negative (Jumlah prediksi benar pada kelas negatif)

Pada Tabel 2.1 nilai TP (true positive) dan TN (true negative) menunjukkan tingkat ketepatan klasifikasi. Umumnya semakin tinggi nilai TP dan TN semakin baik pula tingkat klasifikasi dari akurasi, presisi, dan recall. Jika label prediksi keluaran bernilai benar (true) dan nilai sebenarnya bernilai salah (false) disebut sebagai false positive (FP). Sedangkan jika prediksi label keluaran bernilai salah (false) dan nilai sebenarnya bernilai benar (true) maka hal ini disebut sebagai false

negative (FN). Berdasarkan Tabel 2.1 perhitungan evaluasi performansi metode klasifikasi dapat dilakukan dengan formula sebagai berikut:

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

$$\text{Presisi} = \frac{TP}{TP+FP} \times 100\%$$

$$\text{Recall} = \frac{TP}{TP+FN} \times 100\%$$

14. Python

Python merupakan sebuah bahasa pemrograman yang cukup terkenal yang memiliki banyak manfaat untuk mendukung pemrograman yang berorientasi objek dan dapat berjalan diberbagai macam platform sistem operasi, seperti PCs, Macintosh, UNIX. Beberapa kelebihan dari bahasa pemrograman python antara lain:

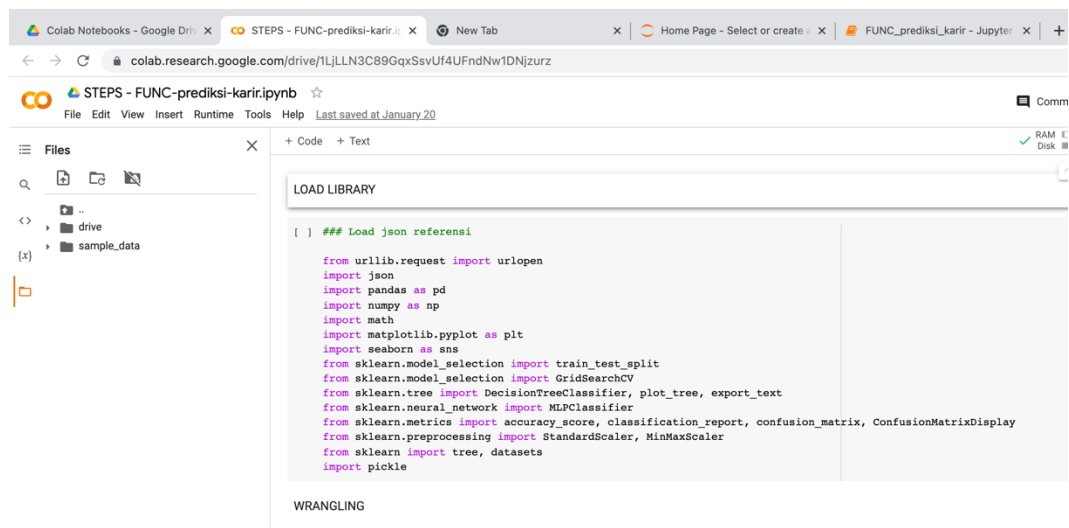
- a. Pengembangan program dilakukan dengan cepat dan coding yang lebih sedikit
- b. Mendukung Multi Platform
- c. Memiliki sistem pengelolaan memori yang otomatis
- d. Python bersifat Object Oriented Programming (OOP)

15. Google Colaboratory

Google Colaboratory atau disebut juga Google Colab adalah *tools* yang dikeluarkan oleh *Google Internal Research*. *Tools* ini dibuat untuk membantu para Peneliti dalam mengolah data, khususnya di bidang

Machine Learning. Dengan Google Colab pengguna dapat mengakses sumber daya yang powerful melalui browser secara bebas untuk mengeksekusi kode dan membangun aplikasi menggunakan berbagai *library* yang tersedia seperti Numpy, OpenCV, Keras, TensorFlow, dan lain-lain.

Google Colab seperti terlihat pada Gambar 2.10 memiliki tampilan yang mirip dengan Jupyter Notebook dan dibuat di atas *environment* Jupyter yang tidak memerlukan pengaturan terlebih dahulu sebelum digunakan dan berjalan sepenuhnya pada *cloud* dengan memanfaatkan media penyimpanan Google Drive (Digmi, 2018).



Gambar 2. 10 Tampilan Google Colaboratory

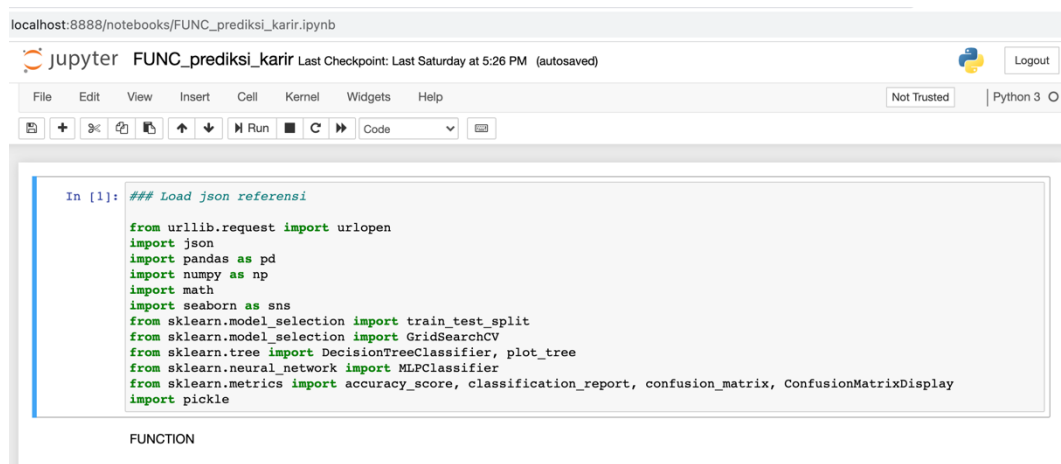
Selain bahasa python, terdapat beberapa perintah yang bisa digunakan di Google Colab seperti perintah git, pip dan sed. Git merupakan fasilitas *version control system* yang banyak digunakan oleh programmer untuk berbagi paket program dan membangun aplikasi secara bersama-sama dengan menggunakan perintah dasar seperti git

init (membuat repositori/tempat penyimpanan paket program), git add (menambah file baru pada repositori), git clone (menyalin repositori), dll (“Git,” n.d.).

Pip adalah salah satu pengelola paket python. Perintah pip merupakan baris program yang dapat dieksekusi di command prompt/penerjemah baris perintah yang ada di Windows ataupun python interpreter. Adapun sed merupakan stream editor dari Unix yang digunakan untuk manipulasi *script* tanpa harus membuka file sumbernya.

16. Jupyter Notebook

Jupyter Notebook merupakan aplikasi yang digunakan untuk membangun aplikasi untuk keperluan *scientific computation* seperti *data science*, *machine learning*, *data processing*, analisis, prediksi, dan lain sebagainya menggunakan bahasa pemrograman python. Jupyter Notebook memungkinkan penggunaannya untuk mengintegrasikan antara kode dengan output di dalam satu dokumen secara interaktif (Faqih, 2018)



The image shows a screenshot of a Jupyter Notebook interface. The browser address bar indicates the URL is localhost:8888/notebooks/FUNC_prediksi_karir.ipynb. The notebook title is 'FUNC_prediksi_karir' and it shows the last checkpoint was on Saturday at 5:26 PM. The interface includes a menu bar (File, Edit, View, Insert, Cell, Kernel, Widgets, Help) and a toolbar with icons for running, saving, and other actions. The main content area shows a code cell with the following Python code:

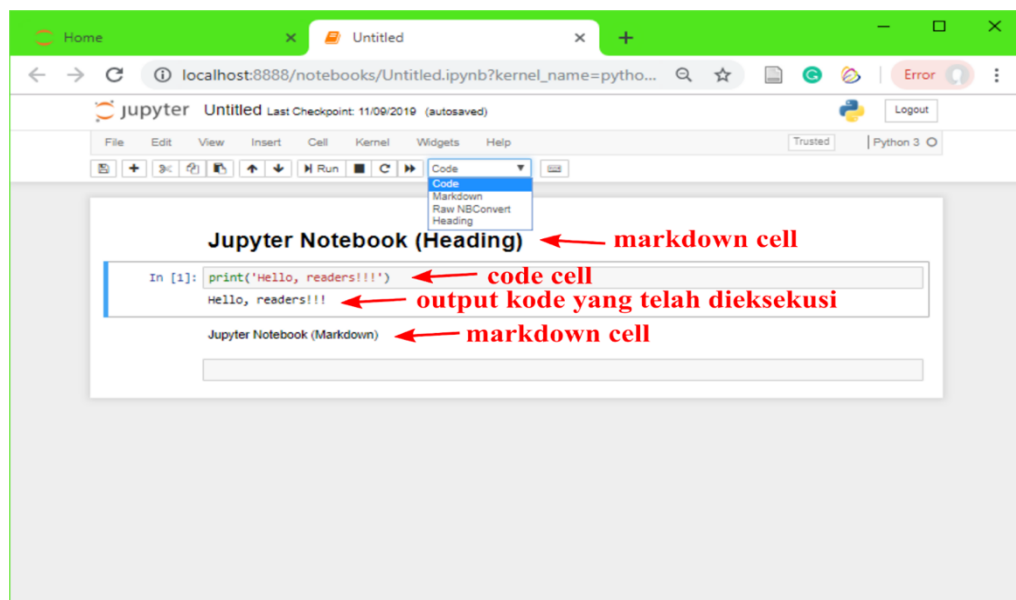
```
In [1]: ### Load json referensi
from urllib.request import urlopen
import json
import pandas as pd
import numpy as np
import math
import seaborn as sns
from sklearn.model_selection import train_test_split
from sklearn.model_selection import GridSearchCV
from sklearn.tree import DecisionTreeClassifier, plot_tree
from sklearn.neural_network import MLPClassifier
from sklearn.metrics import accuracy_score, classification_report, confusion_matrix, ConfusionMatrixDisplay
import pickle
```

Below the code cell, the text 'FUNCTION' is visible, indicating the start of a function definition.

Gambar 2. 11 Tampilan halaman utama Jupyter Notebook

Jupyter Notebook merupakan aplikasi bawaan dari *platform* Anaconda yang berjalan di halaman *browser*. Anaconda sendiri merupakan platform berbasis *open source* untuk bahasa pemrograman Python dan R. Anaconda diciptakan untuk memudahkan pengguna dalam mengelola paket Python yang terdapat di system (Wljaya, 2019).

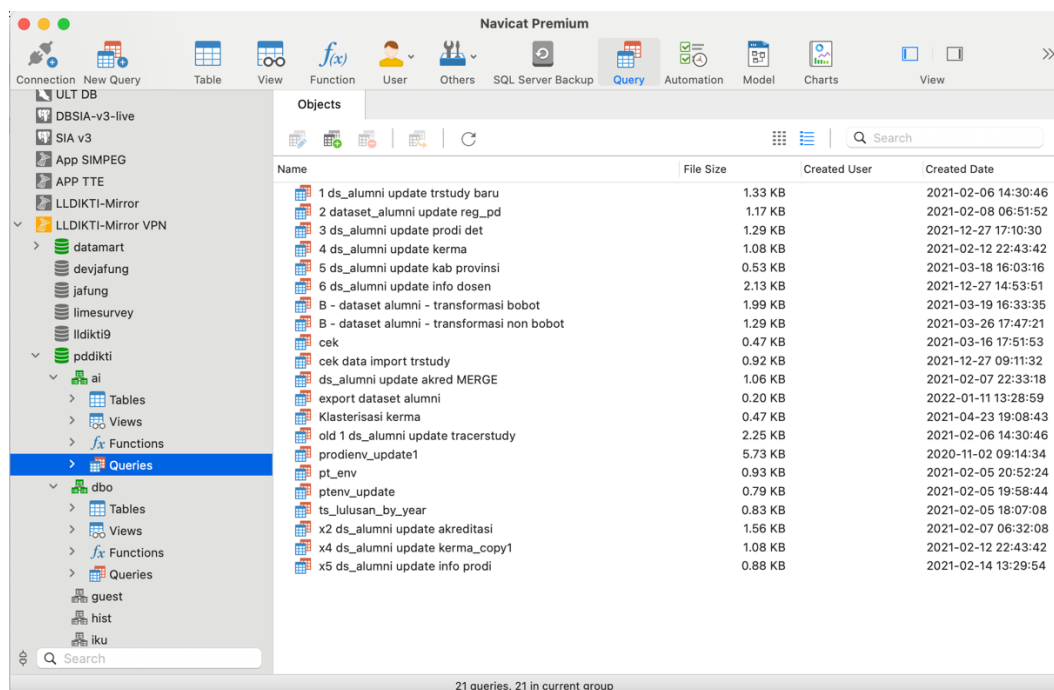
Saat membuka Jupyter Notebook maka yang pertama tampil adalah halaman utama sebagaimana terlihat pada Gambar 2.11. Adapun antarmuka dari Jupyter Notebook terdiri dari *cell* yang merupakan area di mana kode dan teks dituliskan. Untuk mengeksekusi kode dan menampilkan *output* dapat dilakukan dengan menekan SHIFT+ENTER pada *keyboard* atau tombol RUN di Toolbar pada *cell* aktif (Faqih, 2019). Pada Gambar 2.12 dapat dilihat bahwa *cell* untuk menulis dan mengeksekusi kode python disebut *code cell* sedangkan *cell* untuk menuliskan teks yang terformat disebut *markdown cell*.



Gambar 2. 12 Tampilan antarmuka Jupyter Notebook

17. Navicat

Navicat Premium adalah software yang dijadikan sebagai alat pengembangan database yang memungkinkan pengguna untuk terhubung secara bersamaan ke database MySQL, MariaDB, MongoDB, SQL Server, Oracle, PostgreSQL, dan SQLite dari satu aplikasi. Software ini kompatibel dengan database cloud seperti Amazon RDS, Amazon Aurora, Amazon Redshift, Microsoft Azure, Oracle Cloud, Google Cloud, dan MongoDB Atlas. Pengguna dapat dengan cepat dan mudah membangun, mengelola, dan memelihara database. (Cybertech Ltd, 2020)



Gambar 2. 13 Tampilan antar muka Navicat

B. Penelitian Terkait

Beberapa penelitian terkait untuk memprediksi karir alumni di antaranya sebagai berikut:

1. (Rangnekar et al., 2018b) Melakukan Penelitian Model prediksi karir menggunakan *data mining* dan klasifikasi linear. Kepribadian siswa ditentukan dengan menggunakan akun media sosial mereka melalui Facebook Graph API. Jawabannya kemudian dimasukkan ke dalam model yang dibuat untuk memprediksi karir yang sesuai dengan bakat dan kepribadian siswa. Model ini memberikan akurasi rata-rata 77,41% untuk bakat, 75,4% untuk kepribadian, dan 60,09% untuk informasi latar belakang.
2. (Katore et al., 2015) Melakukan penelitian tentang rekomendasi karir siswa berdasarkan analisis kondisi psikologis. Sistem rekomendasi dikembangkan dengan bantuan Algoritma C4. 5 dan pemisahan node didasarkan pada perolehan informasi Keakuratan sistem adalah 86%. Selain Algoritma C4.5, metode yang digunakan sebagai pembanding adalah simple chart, K Chart dan Naïve Bayes namun hanya memperoleh akurasi di bawah 85%.
3. (Xinjuan and Jing, 2013b) Penelitian tentang Pengambilan Keputusan Karir Pasca sarjana Berdasarkan Metode Rough Set Theory dan Decision Tree. Penelitian ini menggabungkan situasi aktual universitas, dan menerapkan kumpulan data dan metode pohon keputusan untuk membangun pohon keputusan. Sebelum membangun pohon

keputusan, makalah ini mengambil pengurangan atribut berdasarkan teori himpunan kasar, yang mengurangi biaya komputasi algoritma ID3.

4. (Ambarwari et al., 2020) Penelitian ini bertujuan menganalisis pengaruh teknik normalisasi min-max dan standarisasi (normalisasi zero-mean) terhadap performa algoritme machine learning. Hasil dari dataset yang telah dinormalisasi, kemudian diujikan ke empat algoritme machine learning antara lain KNN, Naïve Bayesian, ANN, SVM dengan kernel RBF dan kernel linear.
5. (Goldman, Ian. and Pabari, 2021) penelitian dengan judul Applying Smote-Nc On Cart Algorithm To Handle Imbalanced Data In Customer Churn Prediction: A Case Study Of Telecommunications Industry. Dalam pengolahan data untuk prediksi churn pelanggan terdapat data yang tidak seimbang yang berpotensi terjadinya kesalahan prediksi dan menimbulkan masalah. Oleh karena itu, untuk mengatasi masalah tersebut, penelitian ini menggabungkan Teknik Synthetic Minority Oversampling – Nominal Continuous (SMOTE-NC) dengan Classification and Regression Trees (CART).
6. (Gök and Olgun, 2021) Melakukan penelitian untuk meningkatkan akurasi sistem diagnosis berdasarkan sampel darah untuk memprediksi tingkat keparahan pasien covid 19. Salah satu masalah dalam penelitian ini adalah adanya data yang tidak seimbang, sehingga di tahap preprosesing digunakan teknik SMOTE-NC untuk mengatasi hal tersebut. Pada akhir penelitian didapatkan akurasi 0,98.

7. (Hayaty et al., 2021) Melakukan prediksi donor darah yang layak dan tidak layak dengan dataset yang memiliki komposisi kelas tidak seimbang. Metode SMOTE-Oversampling dan Random Oversampling untuk mereplikasi data sampel pada kelas minoritas berhasil meningkatkan akurasi pengenalan kelas yang tidak sesuai dari 12% menjadi 100% untuk algoritma KNN dan 89% untuk algoritma Naïve Bayes.

C. State Of The Art

Rangnekar, dkk melakukan penelitian prediksi karir dengan melihat kepribadian alumni sebagai fitur. Katore, dkk juga melakukan penelitian terkait rekomendasi karir dengan factor psikologis siswa sebagai fitur. Sedangkan Xinjuan, dkk melakukan penelitian terkait karir alumni dengan menggabungkan situasi universitas sebagai fitur. Ketiga penelitian di atas memiliki relevansi terhadap penentuan fitur dataset pada penelitian ini yaitu terkait dengan kepribadian dan psikologi alumni yang dapat dibentuk dari penyelenggaraan akademik program studi dalam menentukan porsi perkuliahan, diskusi, demonstrasi, magang, praktikum lab dan lapangan dan keterlibatan dalam proyek. Sedangkan kondisi universitas meliputi umur program studi, akreditasi, kondisi dosen, serta berbagai dukungan mitra yang dibangun dari kerja sama.

Penelitian yang dilakukan oleh Ambawari, dkk terkait normalisasi data juga memiliki relevansi terhadap penelitian ini karena terdapat variable yang

memiliki rentang nilai yang sangat jauh seperti variabel umur program studi dengan jumlah dosen sebidang maupun jumlah mahasiswa aktif. Sehingga normalisasi dengan metode min max dibutuhkan untuk mengubah nilai numerik ke skala umum yaitu berada di rentang 0 – 1.

Penelitian Goldmen,dkk, Gok dan Olgun serta Hayati,dkk memiliki karakteristik dataset yang sama yaitu terjadi ketidakseimbangan kelas data sehingga dalam meningkatkan akurasi terhadap proses klasifikasi untuk mendapatkan prediksi yang akurat digunakan Teknik SMOTENC Oversampling untuk mendapatkan komposisi kelas dataset yang seimbang. Penelitian yang dilakukan oleh penulis juga memiliki komposisi kelas dataset yang tidak seimbang yaitu kelas dengan nilai baik pada target masa tunggu kerja, hubungan horizontal dan vertikal memiliki populasi data mayoritas. Hal ini tidak dapat dihindari oleh karena perguruan tinggi akan memprioritaskan pendataan terhadap alumni yang memiliki tingkat keberhasilan yang baik dalam karir mereka.

Relevansi penelitian ini terhadap penelitian yang dilakukan penulis adalah beberapa data numerik pada dataset seperti umur, jumlah mahasiswa aktif, jumlah data Kerjasama, dan sejenisnya memiliki rentang nilai yang sangat jauh.

Tabel 2. 2 State of the art

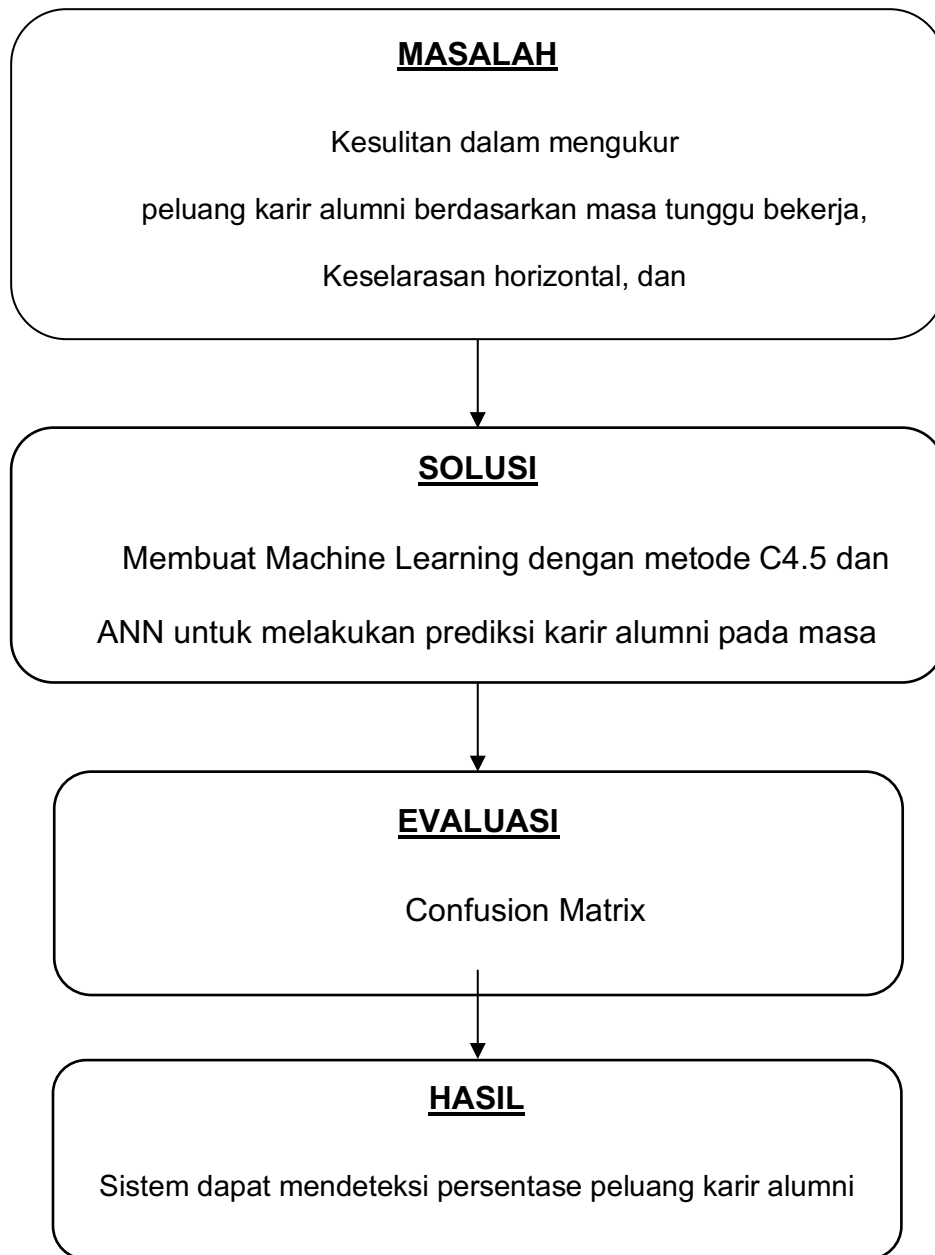
No	Judul	Penerbit, Penulis & Tahun	Metode	Hasil
1.	<i>Career Prediction Model Using Data Mining and Linear Classification</i>	(Rangnekar et al., 2018) IEEE 2018 Fourth International Conference on Computing Communication Control and Automation (ICCCUBEA)	<i>Stochastic Gradient Descent K-Nearest Neighbors Logistic Regression Random Forest</i>	Accuracy 77,41%
2.	<i>Novel professional career prediction and recommendation method for individual through analytics on personal traits using C4.5 algorithm</i>	(Katore et al., 2015) IEEE 2015 Global Conference on Communication Technologies (GCCT)	<i>C4.5 algorithm Advantages of C4.5 High Score</i>	Accuracy 86%
3.	<i>Research on the Graduate Career Decision-making Based on Rough Set Theory and Decision Tree Method)</i>	(Xinjuan Zhang, 2013) AISS INTERNATIONAL JOURNAL ON Advances in Information Sciences and Service Sciences	<i>Algoritma Levenberg- Marquardt 3D reconstruction model</i>	

4.	Analysis of the Effect of Data Scaling on the Performance of the Machine Learning Algorithm for Plant Identification	(Ambarwari A, Jafar Adrian Q, Herdiyeni Y, 2020) Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)	KNN, Naïve Bayesian, ANN, SVM	ANN Accuracy 81,34% Precision 82% Recall 81%
5.	Applying Smote-Nc On Cart Algorithm To Handle Imbalanced Data In Customer Churn Prediction: A Case Study Of Telecommunications Industry	(Goldman, Ian. And Pabari M, 2013) Syntax Literate: Jurnal Ilmiah Indonesia	SMOTE-NC CART	
6.	SMOTE-NC and gradient boosting imputation based random forest classifier for predicting severity level of covid-19 patients with blood samples	(Gök E, Olgun M) Neural Computing and Applications (2021) 33(22) 15693-15707	SMOTE-NC CART	Accuracy 75%
7.	Random and Synthetic Over-Sampling Approach to Resolve Data Imbalance in Classification	(Hayaty M, Muthmainah S, Ghufran S, 2020) International Journal Of Artificial Intelligence Research	SMOTE-OverSampling(SOS) and Random Oversampling (ROS) methods	
8.	SMOTE-ENC: A Novel SMOTE-Based Method to Generate Synthetic Data for Nominal and Continuous Features Mimi	(Mimi Mukherjee and Matloob Khushi, 2021) School of Computer Science, University of Sydney	SMOTE SMOTE-NC SMOTE-ENC	

9.	<i>Predicting academic performance of students from VLE big data using deep learning models</i>	(Wahheed et al., 2020) Computers in Human Behavior	Computers in Human Behavior	Virtual Learning Environment
10.	<i>Predicting at-risk university students in a virtual learning environment via a machine learning algorithm</i>	(Chui et al., 2018) Computers in Human Behavior	RTV-SVM	93,5%.
11.	<i>Student Performance Prediction with Optimum Multilabel Ensemble Model</i>	(Yekun and Teklay, 2019) arXiv:1909.07444 [cs, stat] Mekelle University	- SVM - Random forest - Classifier - K-nearest neighbor	82,2% F1 micro 81,5% F1 macro
12.	<i>Utilizing early engagement and machine learning to predict student outcomes</i>	(Gray and Perkins, 2019) Computers & Education	C4.5 trees	Accuracy 86,20%
13.	<i>An in-depth performance analysis of the oversampling techniques for high-class imbalanced dataset</i>	(Prasetyo Wibowo, Chastine Fatchah, 2021) Department of Informatics, Institut Teknologi Sepuluh Nopember	ROS ADSYN SMOTE KNN	accuracy value of 0.9997, precision, 0.8571 recall, 0.9000 F1-score, 0.9388 ROC-AUC, and 0.8581

14.	K-Nearest Neighbor with K-Fold Cross Validation and Analytic Hierarchy Process on Data Classification	(Zoelkarnain Tembusai, Mawengkang, Zarlis, 2021)	Rinanda Herman *Muhammad	KNN	accuracy rate of 95%
-----	---	--	--------------------------	-----	----------------------

D. Kerangka Pikir



Gambar 2. 14 Kerangka Pikir

Berdasarkan kerangka pemikiran, tahap pertama menunjukkan permasalahan yang dialami program studi dalam mengukur peluang karir alumni pasca lulus berdasarkan masa tunggu memperoleh pekerjaan,

keselarasan bidang ilmu terhadap bidang pekerjaan (keselarasan horisontal) dan keselarasan jenjang pendidikan program studi terhadap tingkatan jabatan dalam pekerjaan (keselarasan horisontal). Kemudian solusi yang ditawarkan adalah membuat sistem menggunakan metode decision tree C4.5 (pohon keputusan) dan Neural Network (jaringan syaraf tiruan) yang dapat digunakan untuk melakukan prediksi terhadap peluang karir alumni untuk mengukur lebih dini tingkat peluang penyerapan alumni berdasarkan masa tunggu bekerja, keselarasan horisontal dan vertikal sehingga dapat dijadikan acuan dalam peningkatan kinerja program studi. Evaluasi dilakukan menggunakan Confusion Matrix untuk melihat persentase keberhasilan model yang dihasilkan.