KLASIFIKASI KATEGORI IQ ANAK BERDASARKAN TINGKAT PENDIDIKAN ORANG TUA MENGGUNAKAN *MULTILAYER PERCEPTRON*

SKRIPSI

Oleh :
‘IZZAN NUHA ZAMRONI
220605110082

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

KLASIFIKASI KATEGORI IQ ANAK BERDASARKAN TINGKAT PENDIDIKAN ORANG TUA MENGGUNAKAN *MULTILAYER PERCEPTRON*

SKRIPSI

Diajukan kepada:
Universitas Islam Negeri Maulana Malik Ibrahim Malang
Untuk memenuhi Salah Satu Persyaratan dalam
Memperoleh Gelar Sarjana Komputer (S.Kom)

Oleh :
'IZZAN NUHA ZAMRONI
220605110082

PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS SAINS DAN TEKNOLOGI
UNIVERSITAS ISLAM NEGERI MAULANA MALIK IBRAHIM
MALANG
2026

HALAMAN PERSETUJUAN

KLASIFIKASI KATEGORI IQ ANAK BERDASARKAN TINGKAT PENDIDIKAN ORANG TUA MENGGUNAKAN *MULTILAYER PERCEPTRON*

SKRIPSI

Oleh:
'IZZAN NUHA ZAMRONI
220605110082

Telah Diperiksa dan Disetujui untuk Diuji:
Tanggal: 12 Juni 2026

Pembimbing I,	Pembimbing II,
Dr. Zainal Abidin, M.Kom	Roro Inda Melani, M.T, M.Sc
NIP. 19760613 200501 1 004	NIP. 19780925 200501 2 008

Mengetahui,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

HALAMAN PENGESAHAN

KLASIFIKASI KATEGORI IQ ANAK BERDASARKAN TINGKAT PENDIDIKAN ORANG TUA MENGGUNAKAN *MULTILAYER PERCEPTRON*

SKRIPSI

Oleh:
'IZZAN NUHA ZAMRONI
NIM. 220605110082

Telah Dipertahankan di Depan Dewan Penguji Skripsi
dan Dinyatakan Diterima Sebagai Salah Satu Persyaratan
Untuk Memperoleh Gelar Sarjana Komputer (S.Kom)
Pada Tanggal: 12 Juni 2026

Susunan Dewan Penguji

Ketua Penguji	: Fatchurrohman, M.Kom. NIP. 19700731 200501 1 002	( )
Anggota Penguji I	: Nurizal Dwi Priandani, M.Kom. NIP. 199208302022031001	( )
Anggota Penguji II	: Dr. Zainal Abidin, M.Kom. NIP. 19760613 200501 1 004	( )
Anggota Penguji III	: Roro Inda Melani, M.T, M.Sc NIP. 19780925 200501 2 008	( )

Mengetahui dan Mengesahkan,
Ketua Program Studi Teknik Informatika
Fakultas Sains dan Teknologi
Universitas Islam Negeri Maulana Malik Ibrahim Malang

Supriyono, M.Kom
NIP. 19841010 201903 1 012

PERNYATAAN KEASLIAN TULISAN

Saya yang bertanda tangan di bawah ini:

Nama	: 'Izzan Nuha Zamroni
NIM	: 220605110082
Fakultas / Program Studi	: Sains dan Teknologi / Teknik Informatika
Judul Skripsi	: Klasifikasi Kategori Iq Anak Berdasarkan Tingkat Pendidikan Orang Tua Menggunakan *Multilayer Perceptron*

Menyatakan dengan seberanya bahwa Skirpsi yang saya tulis ini benar-benar merupakan hasil karya sendiri, bukan merupakan pengambil alihan data, tulisan, atau pikiran orang lain yang saya akui sebagai hasil tulisan atau pikiran saya sendiri, kecuali dengan mencantumkan sumber cuplikan pada daftar pustaka.

Apabila dikemudian hari terbukti atau dapat dibuktikan skripsian ini merupakan hasil jiplakan, maka saya bersedia menerima sanksi atas perbuatan tersebut.

Malang, 12 Juni 2026
Yang membuat pernyatan,

'Izzan Nuha Zamroni
NIM.220605110082

MOTTO

"People come and go in your life, but the right ones will always stay"

"Pelajaran tanpa rasa sakit tidak berarti apa-apa"
(lolita)

HALAMAN PERSEMBAHAN

Dengan penuh rasa syukur kepada Allah SWT atas limpahan rahmat dan kemudahan-Nya, akhirnya skripsi ini dapat terselesaikan dengan baik. Karya ini penulis persembahkan kepada:

Ayah, Ibu, kakak, dan adik

atas doa, dukungan, dan kasih sayang tanpa henti yang selalu menjadi sumber kekuatan dan motivasi.

Dosen pembimbing dan para pengajar

atas bimbingan, arahan, serta ilmu yang sangat berarti bagi penulis.

Teman-teman yang penulis temui di tahun 2022 - 2026

yang telah menjadi bagian dari perjalanan ini melalui kebersamaan dan semangat yang tidak pernah padam.

Dan yang terakhir, untuk diri sendiri

Terima kasih telah bertahan, tetap kuat di saat lelah, dan tidak menyerah dalam menghadapi setiap proses.

KATA PENGANTAR

Assalamu'alaikum Warahmatullahi Wabarakatuh

Alhamdulillahirabbil'alamin, segala puji syukur penulis panjatkan ke hadirat Allah SWT. atas limpahan rahmat, nikmat, dan hidayah-Nya, sehingga penulis dapat menyelesaikan tugas skripsi yang berjudul "KLASIFIKASI KATEGORI IQ ANAK BERDASARKAN TINGKAT PENDIDIKAN ORANG TUA MENGGUNAKAN *MULTILAYER PERCEPTRON*" dengan baik dan lancar. Shalawat serta salam tak lupa penulis sampaikan kepada Nabi Muhammad ShallAllahu 'Alaihi Wasallam, sosok teladan mulia yang telah menuntun kita dari zaman kegelapan menuju zaman kebenaran, Islam, dan ilmu pengetahuan yang kita rasakan manfaatnya hingga hari ini.

Tugas skripsi ini disusun dalam rangka memenuhi salah satu syarat kelulusan untuk memperoleh gelar Sarjana Komputer pada Program Studi Teknik Informatika, Fakultas Sains dan Teknologi, UIN Maulana Malik Ibrahim Malang. Dalam proses penyusunan tugas ini, penulis mendapatkan bantuan, dukungan, serta doa dari banyak pihak, baik secara langsung maupun tidak langsung. Untuk itu, penulis ingin menyampaikan rasa terima kasih yang sebesar-besarnya kepada:

1. Prof. Dr. Hj. Ilfi Nur Diana, M.Si., selaku Rektor Universitas Islam Negeri (UIN) Maulana Malik Ibrahim Malang.
2. Dr. H. Agus Mulyono, S.Pd., M.Kes, selaku Dekan Fakultas Sains dan Teknologi UIN Maulana Malik Ibrahim Malang.
3. Supriyono, M.Kom., selaku Ketua Program Studi S1 Teknik Informatika UIN Maulana Malik Ibrahim Malang.

4. Dr. Zainal Abidin, M.Kom., selaku Dosen Pembimbing I, yang dengan kesabaran dan dedikasinya membimbing penulis selama penyusunan skripsi.
5. Roro Inda Melani, M.T, M.Sc, selaku Dosen Pembimbing II, yang senantiasa memberikan arahan dan masukan yang sangat berharga dalam pengerjaan skripsi ini.
6. Fatchurrohman, M.Kom., selaku Dosen Penguji sekaligus Dosen wali yang telah memberikan kritik, saran, dan bimbingan untuk menyempurnakan skripsi.
7. Nurizal Dwi Priandani, M.Kom., selaku Dosen Penguji yang telah memberikan kritik, saran, dan bimbingan untuk menyempurnakan skripsi ini.
8. Nia Faricha, S.Si., admin Program Studi Teknik Informatika, yang membantu dalam urusan administrasi dan selalu mengingatkan tentang kelengkapan berkas.
9. Bapak dan Ibu tercinta, serta kakak saya, dan adik saya, yang senantiasa menjadi sumber kekuatan melalui doa, kasih sayang, dan dukungan tanpa henti.
10. Teman dekat kamar ma'had dan pondok, yang selalu ada ketika penulis sedang dalam kesulitan dalam hal apapun.
11. Teman-teman 2RU, yang senantiasa memberikan waktu luang kepada penulis untuk sejenak menghibur diri.
12. Seluruh keluarga, teman, sahabat, dan kerabat penulis yang tidak dapat penulis sebutkan satu per satu, yang turut memberikan bantuan, semangat, dukungan, dan doa untuk penulis.

Dan terakhir tak lupa terimakasih kepada diri sendiri, yang dengan tekad kuat mampu melewati setiap tantangan, rintangan, dan kesulitan selama perjalanan

ini, membuktikan bahwa kerja keras dan usaha tidak pernah sia-sia. Penulis menyadari bahwa penelitian dalam tugas skripsi ini masih jauh dari kata sempurna dan memiliki berbagai keterbatasan. Oleh karena itu, dengan penuh kerendahan hati, penulis membuka diri untuk menerima kritik dan saran yang membangun dari para pembaca guna menjadi bahan evaluasi dan pengembangan di masa mendatang. Penelitian ini juga memiliki potensi untuk dilanjutkan dan dikembangkan lebih jauh dalam penelitian berikutnya, sehingga dapat melengkapi kekurangan yang ada. Penulis berharap, karya ini tidak hanya memberikan manfaat bagi pembaca, tetapi juga dapat berkontribusi positif bagi masyarakat secara luas.

Wassalamu'alaikum Warahmatullahi Wabarakatuh

Malang, 12 Juni 2026

Penulis

DAFTAR ISI

HALAMAN PERSETUJUAN ..... **iii**
HALAMAN PENGESAHAN ..... **iv**
PERNYATAAN KEASLIAN TULISAN ..... **v**
MOTTO ..... **vi**
HALAMAN PERSEMBAHAN ..... **vii**
KATA PENGANTAR ..... **viii**
DAFTAR ISI ..... **xi**
DAFTAR GAMBAR ..... **xiii**
DAFTAR TABEL ..... **xiv**
ABSTRAK ..... **xv**
ABSTRACT ..... **xvi**
مستخلص البحث ..... **xvii**
BAB I PENDAHULUAN **1**
1.1 Latar Belakang 1
1.2 Rumusan Masalah 4
1.3 Batasan Masalah 4
1.4 Tujuan Penelitian 5
1.5 Manfaat Penelitian 5
BAB II STUDI PUSTAKA **6**
2.1 Stanford-Binet Intelligence Scales Fifth Edition (SB5) 6
2.2 Pendidikan Orang Tua 9
2.3 Machine Learning dan Multilayer Perceptron (MLP) 10
2.4 Penelitian Terkait 18
BAB III DESAIN DAN IMPLEMENTASI **26**
3.1 Desain Penelitian 26
3.1.1 Pengumpulan Data 27
3.1.2 Prapemrosesan Data 27
3.1.3 Arsitektur Model *Multilayer* Perceptron (MLP) 35
3.1.4 Loss Function 43

3.1.5 Backpropagation 44
3.1.6 Pembaruan Bobot dan Bias 45
3.1.7 Skenario Eksperimen 47
3.2 Implementasi 50
3.2.1 Implementasi User Interface (UI) 51
3.2.2 Implementasi Model *Multilayer Perceptron* 55
3.2.3 Alur Implementasi Sistem 57
BAB IV HASIL DAN PEMBAHASAN 59
4.1 Hasil Eksperimen 59
4.1.1 Arsitektur Hidden Layer (6, 5) 59
4.1.2 Arsitektur Hidden Layer (64, 64) 70
4.1.3 Analisis Perbandingan Model Arsitektur 75
4.3 Impelemtasi GUI 78
4.4 Pembahasan dan Integrasi Islam 86
BAB V KESIMPULAN DAN SARAN 90
5.1 Kesimpulan 90
5.2 Saran 92
DAFTAR PUSTAKA
LAMPIRAN-LAMPIRAN

DAFTAR GAMBAR

Gambar 2. 1 Contoh gambar arsitektur mlp 12
Gambar 3. 1 Alur Penelitian 26
Gambar 3. 2 Tahapan prapemrosesan 28
Gambar 3. 3 Dataset stanford binet intelligence scales fifth edition 29
Gambar 3. 4 Arsitektur model multilayer perceptron 35
Gambar 3. 5 Skenario pengujian 49
Gambar 3. 6 Contoh halaman single prediction 52
Gambar 3. 7 Contoh halaman bulk prediction 53
Gambar 3. 8 Sourcecode input data pada halaman single prediction 53
Gambar 3. 9 Sourcecode input data pada halaman bulk prediction 54
Gambar 3. 10 Kode program pemanggilan model 55
Gambar 3. 11 Kode program prediksi menggunakan mlp 56
Gambar 3. 12 . Integrasi model mlp dengan streamlit 57
Gambar 3. 13 Alur implementasi sistem 57
Gambar 4. 1 Confusion matrix (model terbaik) 66
Gambar 4. 2 Confusion matrix (model 17) 72
Gambar 4. 3 Tampilan formulir input single prediction 81
Gambar 4. 4 Tampilan hasil prediksi single prediction 82
Gambar 4. 5 Tampilan antarmuka input bulk prediction 84
Gambar 4. 6 Tampilan hasil bulk prediction 85

DAFTAR TABEL

Tabel 2. 1 Kategorisasi iq anak 6
Tabel 2. 2 Ringkasan jenis soal berdasarkan bentuk verbal dan nonverbal 9
Tabel 2. 3 Confusion matrix 15
Tabel 2. 4 Penelitian terkait tes SB5 22
Tabel 3. 1 Ringkasan variabel dan metode korelasi 30
Tabel 3. 2 Variabel input 31
Tabel 3. 3 Variabel output 31
Tabel 3. 4 Encoding variabel education_father & education_mother 33
Tabel 3. 5 Encoding variabel gender 33
Tabel 3. 6 Skenario uji coba 49
Tabel 3. 7 Komponen utama ui streamlit 54
Tabel 4. 1 Hasil uji coba model dengan full preprocessing 61
Tabel 4. 2 Hasil uji coba model tanpa standarisasi 62
Tabel 4. 3 Hasil uji coba model tanpa smote 62
Tabel 4. 4 Hasil uji coba model tanpa smote dan standarisasi 63
Tabel 4. 5 Perbandingan pengaruh smote dan standarisasi pada arsitektur (6, 5) 64
Tabel 4. 6 Perbandingan pengaruh variabel pendidikan orang tua pada arsitektur (6, 5) 67
Tabel 4. 7 Ringkasan hasil skenario arsitektur (6, 5) 69
Tabel 4. 8 Perbandingan pengaruh smote dan standarisasi pada arsitektur (64, 64) 70
Tabel 4. 9 Perbandingan pengaruh variabel pendidikan orang tua pada arsitektur (64, 64) 73
Tabel 4. 10 Ringkasan hasil skenario arsitektur (64, 64) 75
Tabel 4. 11 Perbandingan model terbaik (6, 5 dan 64, 64) 76
Tabel 4. 12 Ringkasan hasil seluruh model 77

ABSTRAK

Zamroni, ‘Izzan Nuha. 2026. **Klasifikasi Iq Anak Berdasarkan Tingkat Pendidikan Orang Tua Menggunakan *Multilayer Perceptron***. Skripsi. Program Studi Teknik Informatika Fakultas Sains dan Teknologi Universitas Islam Negeri Maulana Malik Ibrahim Malang. Pembimbing: (I) Dr. Zainal Abidin, M.Kom (II) Roro Inda Melani, M.T, M.Sc.

Kata Kunci: *Multilayer Perceptron*, klasifikasi IQ anak, pendidikan orang tua, Stanford-Binet Fifth Edition, SMOTE

Penelitian ini bertujuan mengimplementasikan metode *Multilayer Perceptron* (MLP) untuk mengklasifikasikan kemampuan intelektual anak ke dalam lima kategori berdasarkan *Stanford-Binet Intelligence Scales Fifth Edition* (SB5), yakni *Moderate Intellectual Disability*, *Mild Intellectual Disability*, *Below Average Intelligence*, *Average Intelligence*, dan *Above-Average Intelligence*. Variabel prediktor yang digunakan meliputi tingkat pendidikan ayah, tingkat pendidikan ibu, usia anak, dan jenis kelamin. Penelitian melibatkan 32 skenario eksperimen yang dikombinasikan dari dua arsitektur *hidden layer*, yaitu (6, 5) dan (64, 64), dua teknik prapemrosesan (*SMOTE* dan standardisasi), serta empat kombinasi variabel pendidikan orang tua. Evaluasi dilakukan menggunakan *Stratified 5-Fold Cross Validation* dengan metrik akurasi, *precision*, *recall*, *Macro F1-Score*, dan *AUC-ROC*. Hasil eksperimen menunjukkan bahwa penerapan *SMOTE* memberikan pengaruh terbesar terhadap peningkatan keseimbangan klasifikasi, sementara variabel *education_father* memberikan kontribusi lebih signifikan dibandingkan *education_mother*. Model terbaik (Model 17) dengan arsitektur (64, 64), *SMOTE*, standardisasi, dan seluruh variabel input menghasilkan akurasi 28,66%, *Macro F1-Score* 26,30%, dan *AUC-ROC* 0,63–0,79. Performa model yang masih moderat mencerminkan keterbatasan informasi pada variabel yang digunakan, bukan kegagalan model, mengingat kecerdasan anak dipengaruhi oleh faktor multidimensional yang tidak sepenuhnya terukur melalui latar belakang pendidikan orang tua semata.

ABSTRACT

Zamroni, 'Izzan Nuha. 2026. **Classification of Children's IQ Based on Parents' Educational Level Using a Multilayer Perceptron**. Undergraduate thesis. Department of Computer Science, Faculty of Science and Technology, Maulana Malik Ibrahim State Islamic University, Malang. Supervisors: (I) Dr Zainal Abidin, M.Kom (II) Roro Inda Melani, M.T, M.Sc.

Keywords: *Multilayer Perceptron, children's IQ classification, parental education, Stanford-Binet Fifth Edition, SMOTE*

This study aims to implement the Multilayer Perceptron (MLP) method to classify children's intellectual abilities into five categories based on the Stanford-Binet Intelligence Scales Fifth Edition (SB5), namely Moderate Intellectual Disability, Mild Intellectual Disability, Below-Average Intelligence, Average Intelligence, and Above-Average Intelligence. The predictor variables used include the father's educational level, the mother's educational level, the child's age, and gender. The study involved 32 experimental scenarios combining two hidden layer architectures, namely (6, 5) and (64, 64), two pre-processing techniques (SMOTE and standardisation), and four combinations of parental educational variables. Evaluation was carried out using Stratified 5-Fold Cross-Validation with the metrics of accuracy, precision, recall, Macro F1-Score, and AUC-ROC. The experimental results showed that the application of SMOTE had the greatest impact on improving classification balance, whilst the variable 'education_father' made a more significant contribution than 'education_mother'. The best model (Model 17), featuring an architecture of (64, 64), SMOTE, standardisation, and all input variables, achieved an accuracy of 28.66%, a Macro F1-Score of 26.30%, and an AUC-ROC of 0.63–0.79. The model's still moderate performance reflects the limitations of the information contained in the variables used, rather than a failure of the model, given that a child's intelligence is influenced by multidimensional factors that cannot be fully measured by parental educational background alone.

مستخلص البحث

زمروني، عزان نوها. 2026. تصنيف الذكاء لدى الأطفال بناءً على مستوى تعليم الوالدين باستخدام شبكة بيرسيبترون متعددة الطبقات. أطروحة تخرج. برنامج هندسة المعلوماتية، كلية العلوم والتكنولوجيا، الجامعة الإسلامية الحكومية «مولانا مالك إبراهيم» في مالانغ. المشرفان (I) :د. زينال أبيدين، ماجستير في علوم الحاسوب (II)رورو إندا ميلاني، ماجستير في الهندسة، ماجستير في العلوم .

الكلمات المفتاحية: بيرسيبترون متعدد الطبقات، تصنيف معدل الذكاء لدى الأطفال، تثقيف الوالدين، اختبار ستانفورد–بينيت الطبعة الخامسة، SMOTE

يهدف هذا البحث إلى تطبيق طريقة «البيرسيبترون متعدد الطبقات (MLP) «لتصنيف القدرات الذهنية للأطفال إلى خمس فئات استنادًا إلى مقياس ستانفورد–بينيت للذكاء، الطبعة الخامسة (SB5) ، وهي: الإعاقة الذهنية المتوسطة، والإعاقة الذهنية الخفيفة، والذكاء الأقل من المتوسط، والذكاء المتوسط، والذكاء فوق المتوسط. وتشمل المتغيرات التنبؤية المستخدمة مستوى تعليم الأب، ومستوى تعليم الأم، وعمر الطفل، والجنس. تضمن البحث 32 سيناريو تجريبيًا تم تكوينها من مزيج بين بنيتين للطبقة المخفية، وهما (6، 5) و(64، 64)، وتقنيتين للمعالجة المسبقة SMOTE) والتوحيد القياسي(، بالإضافة إلى أربعة مجموعات من متغيرات مستوى تعليم الوالدين. أُجري التقييم باستخدام «التحقق المتقاطع الخماسي الطبقات (*Stratified 5-Fold Cross Validation*) «مع مقاييس الدقة، والدقة النوعية، ومعدل الاسترجاع، ومؤشر F1 الكلي، ومنطقة تحت منحنى .(ROC (AUC-ROC أظهرت نتائج التجارب أن تطبيق تقنية SMOTE كان له التأثير الأكبر على تحسين توازن التصنيف، في حين أن متغير «مستوى تعليم الأب (*education_father*) «قدم مساهمة أكثر أهمية مقارنة بمتغير «مستوى تعليم الأم .(*education_mother*) «أفضل نموذج (النموذج 17) الذي يعتمد على بنية (64، 64)، وتقنيةSMOTE ، والتوحيد القياسي، وجميع متغيرات الإدخال، حقق دقة بنسبة 28,66٪، و Macro F1-Score بنسبة 26,30٪، و AUC-ROC في نطاق 0,63–0,79. ويعكس الأداء المتوسط للنموذج محدودية المعلومات المتوفرة حول المتغيرات المستخدمة، وليس فشل النموذج، نظراً لأن ذكاء الطفل يتأثر بعوامل متعددة الأبعاد لا يمكن قياسها بالكامل من خلال الخلفية التعليمية للوالدين وحدها.

BAB I

PENDAHULUAN

1.1 Latar Belakang

Kategorisasi IQ anak penting dalam bidang pendidikan dan psikologi karena membantu mengidentifikasi kemampuan kognitif anak untuk mendukung proses pembelajaran dan evaluasi perkembangan secara lebih tepat. Tingkat kecerdasan anak berkaitan dengan perkembangan akademik, kemampuan adaptasi sosial, serta kemampuan menerima pembelajaran (Reinaldi & Hidayat, 2021a). Penelitian menunjukkan bahwa kemampuan intelektual anak dipengaruhi tidak hanya oleh faktor biologis, tetapi juga oleh lingkungan keluarga, pendidikan orang tua, pola pengasuhan, serta perkembangan memori dan adaptasi anak (Sajewicz-Radtke, Łada-Maśko, Olech, Leoniak, et al., 2025). Selain itu, setiap anak memiliki profil kognitif yang berbeda sehingga diperlukan proses kategorisasi IQ yang sesuai dengan kondisi masing-masing anak (Sajewicz-Radtke et al., 2022). Oleh karena itu, kategorisasi IQ dapat membantu identifikasi kemampuan anak secara lebih objektif. Proses tersebut juga mendukung pengambilan keputusan pendidikan yang lebih tepat dan terarah.

Berbagai penelitian telah menganalisis kemampuan intelektual anak serta faktor-faktor yang memengaruhi hasil IQ menggunakan instrumen Stanford-Binet Fifth Edition (SB-5). Penelitian (Stephenson et al., 2023) menggunakan SB-5 untuk mengukur IQ anak ASD dan ADHD, sedangkan penelitian (Reinaldi & Hidayat, 2021b) menggunakan Stanford-Binet untuk memprediksi kemampuan intelektual

dan performa akademik anak. Faktor lingkungan keluarga dan tingkat pendidikan orang tua juga memiliki hubungan terhadap perkembangan kemampuan intelektual anak (McKinney et al., 2025), Sajewicz-Radtke, Łada-Maśko, Olech, Jurek, et al., 2025) Penelitian tersebut menunjukkan bahwa pendidikan orang tua, terutama ibu, berpengaruh terhadap perkembangan IQ anak. Namun, sebagian besar penelitian sebelumnya masih menggunakan pendekatan asesmen konvensional dan analisis statistik dasar serta belum mengembangkan sistem klasifikasi IQ otomatis berbasis machine learning. Oleh karena itu, diperlukan metode machine learning seperti Multilayer Perceptron (MLP) untuk membantu klasifikasi kategorisasi IQ anak secara lebih akurat dan efisien.

Penelitian ini bertujuan membangun model klasifikasi IQ anak berdasarkan pendidikan orang tua menggunakan metode *Multilayer Perceptron* (MLP), yang diharapkan dapat membantu proses identifikasi kemampuan intelektual anak secara lebih terstruktur dan efisien. Metode *Multilayer Perceptron* (MLP) dipilih karena *Artificial Neural Network* memiliki kemampuan dalam mengenali hubungan data yang kompleks serta bersifat nonlinear, yang menunjukkan bahwa *Artifical Neural Network* mampu memberikan hasil klasifikasi yang baik pada data kompleks melalui proses pelatihan model dan optimasi hyperparameter (Wang, 2025). Penelitian (Kufel et al., 2023) menunjukkan bahwa penerapan *machine learning* dan *deep learning* mampu meningkatkan kemampuan analisis data dan pengenalan pola dibanding pendekatan konvensional. Arsitektur *Artificial Neural Network* menunjukkan bahwa metode Multilayer Perceptron (MLP) mampu mempelajari

hubungan data yang bersifat nonlinear melalui proses training (*feedforward*) dan mekanisme backpropagation (Freire et al., 2023).

Dalam perspektif keilmuan modern, klasifikasi berfungsi untuk mengidentifikasi pola dan mengelompokkan individu berdasarkan karakteristik tertentu secara objektif. Prinsip dasar ini sejalan dengan nilai-nilai Islam mengenai keadilan dan penghargaan terhadap keberagaman manusia, sebagaimana ditegaskan dalam QS. Al-Hujurāt [49]:13:

يٰٓاَيُّهَا النَّاسُ اِنَّا خَلَقْنٰكُمْ مِّنْ ذَكَرٍ وَّاُنْثٰى وَجَعَلْنٰكُمْ شُعُوْبًا وَّقَبَاۤىِٕلَ لِتَعَارَفُوْاۚ اِنَّ اَكْرَمَكُمْ عِنْدَ اللّٰهِ اَتْقٰىكُمْۗ اِنَّ اللّٰهَ عَلِيْمٌ خَبِيْرٌ

"Wahai manusia, sesungguhnya Kami telah menciptakan kamu dari seorang laki-laki dan perempuan. Kemudian, Kami menjadikan kamu berbangsa-bangsa dan bersuku-suku agar kamu saling mengenal. Sesungguhnya yang paling mulia di antara kamu di sisi Allah adalah orang yang paling bertakwa. Sesungguhnya Allah Maha Mengetahui lagi Mahateliti." (QS. Al-Hujurat: 13).

Menurut Tafsir Ibn Kathir, QS. Al-Hujurāt [49]:13 menjelaskan bahwa seluruh manusia berasal dari satu asal, yaitu Adam dan Hawa, kemudian berkembang menjadi berbagai bangsa dan suku agar saling mengenal, bukan untuk saling merendahkan. Perbedaan di antara manusia tidak dimaksudkan sebagai dasar keunggulan, karena kemuliaan sejati hanya diukur dari tingkat ketakwaan dan ketaatan kepada Allah, bukan dari garis keturunan, kekayaan, maupun status sosial. Nilai moral yang terkandung dalam ayat tersebut memberikan fondasi etis bagi konsep klasifikasi dalam ilmu pengetahuan modern bahwa proses pengelompokan seharusnya tidak bersifat diskriminatif atau hierarkis, melainkan berorientasi pada pemahaman terhadap keragaman manusia demi kemaslahatan bersama. Penelitian ini penting untuk dilakukan karena proses kategorisasi IQ anak masih banyak menggunakan metode konvensional yang memiliki keterbatasan dalam mengenali

pola hubungan kompleks antarvariabel kognitif. Selain itu, sebagian besar penelitian terdahulu berfokuskan pada pengukuran IQ, analisis psikologis, serta pengaruh faktor lingkungan terhadap kemampuan intelektual anak, yang belum mengembangkan ke sistem klasifikasi berbasis machine learning. Seiring berkembangnya teknologi Artificial Intelligence, menunjukkan bahwa metode machine learning mampu meningkatkan proses klasifikasi dan prediksi data kompleks secara lebih optimal (Mohammad & Al Mansoor, 2024). Penelitian ini menegaskan bahwa klasifikasi kategori kemampuan intelektual anak tidak dimaksudkan untuk memberi label atau membeda-bedakan anak, tetapi untuk mengenali kebutuhan dan potensi mereka secara objektif agar intervensi pendidikan serta kebijakan sosial dapat dilakukan secara lebih adil, inklusif, dan tepat sasaran.

1.2 Rumusan Masalah

Bagaimana performa model *Multilayer Perceptron* (MLP) dalam mengklasifikasikan kategori kemampuan intelektual anak berdasarkan pendidikan orang tua, serta seberapa akurat model tersebut ditinjau dari metrik Accuracy, F1-Score, dan AUC-ROC?

1.3 Batasan Masalah

Penelitian ini menggunakan data sekunder hasil tes *Stanford Binet Intelligence Scales Fifth Edition* (SB5) pada anak usia 3-18 tahun yang tercatat dalam sistem asesmen psikologis di Polandia.

1.4 Tujuan Penelitian

Menerapkan algoritma Multilayer Perceptron (MLP) untuk mengklasifikasikan kategori kemampuan intelektual anak berdasarkan data pendidikan orang tua, serta mengevaluasi performa model menggunakan metrik Accuracy, F1-Score, dan AUC-ROC.

1.5 Manfaat Penelitian

Dengan dilakukannya penelitian ini diharapkan dapat memberikan manfaat di masa depan, yaitu:

a. Penelitian ini diharapkan memberikan kontribusi ilmiah melalui penerapan *machine learning* untuk klasifikasi profil kognitif, yang dapat menjadi landasan bagi studi selanjutnya mengenai penerapan *machine learning* di bidang asesmen psikologi pendidikan.
b. Bagi praktisi pendidikan dan psikologi, penelitian ini diharapkan dapat membantu guru, konselor, dan psikolog dalam melakukan deteksi sejak dini serta klasifikasi kategori kemampuan intelektual anak secara lebih cepat dan efisien, dengan memanfaatkan data hasil tes psikometrik seperti *Stanford Binet* 5 (SB5) sebagai dasar pengambilan keputusan.

BAB II

STUDI PUSTAKA

2.1 Stanford-Binet Intelligence Scales Fifth Edition (SB5)

SB5 merupakan salah satu alat ukur tes kecerdasan yang banyak digunakan dalam asesmen anak dan remaja, termasuk dalam sistem layanan psikologis pendidikan di Polandia (Olech et al., 2024). Dalam praktik klinis, hasil SB5 umumnya diinterpretasikan berdasarkan kategori kemampuan intelektual anak, mulai dari rentang disabilitas intelektual hingga kemampuan tinggi yang relevan untuk program pendidikan. Sebagai contoh, studi (Sajewicz-Radtke, Łada-Maśko, Olech, Jurek, et al., 2025) mendefinisikan kategori IQ anak seperti pada tabel 2.1, yang digunakan untuk menggambarkan distribusi kemampuan kognitif anak yang menjadi peserta penelitian.

Tabel 2. 1 Kategorisasi iq anak

Kategori IQ	**Rentang Skor IQ**
Moderate intellectual disability	35 – 54
Mild intellectual disability	55 – 69
Below average intelligence	70 – 84
Average intelligence	85 – 114
Above-average intelligence	> 114

Klasifikasi ini didasarkan pada teori kecerdasan fluid dan *crystallized* yang dikemukakan oleh (Chase & Chute, 2005). SB5 terdiri atas lima kategori (*Fluid Reasoning (FR), Knowledge (KN), Quantitative Reasoning (QR), Visual-Spatial Processing (VS), dan Working Memory (WM)*) yang merepresentasikan proses kognitif utama, dan dua subkategori (*Non verbal dan Verbal*) yang menunjukkan jenis media stimulus yang digunakan. Penjelasan terkait lima kategori dan dua subkategori sebagai berikut:

1. *Fluid reasoning* (FR) adalah kategori soal yang digunakan untuk menemukan hubungan, pola, atau aturan tersembunyi, kemudian mencari kesimpulan untuk menjawab dengan tepat. Pada soal FR nonverbal, soal disajikan dalam bentuk visual seperti gambar, bentuk, atau pola abstrak. Peserta diminta memilih gambar yang paling tepat untuk melanjutkan pola tersebut guna mengukur kemampuan penalaran visual. Untuk soal (FR) *verbal* disajikan dalam bentuk bahasa atau simbol linguistic. Peserta diminta menentukan hubungan yang paling sesuai atau melanjutkan hubungan tersebut secara logis, yang bertujuan untuk mengukur kemampuan penalaran logis tanpa bergantung pada pengetahuan factual. Perbedaan antara kategori FR verbal dan nonverbal terletak pada media penyajian, sedangkan proses kognitif yang diukur tetap sama.
2. *Knowledge* (KN) adalah kategori tes kognitif yang bertujuan untuk mengukur nilai pengetahuan yang diperoleh dari pengalaman atau hasil belajarnya. Untuk soal tes kognitif (KN) *nonverbal*, peserta ditunjukkan gambar objek. Kemudian peserta diminta mengidentifikasi atau mengenali informasi yang sesuai dengan gambar tersebut. Pada soal KN *verbal*, peserta diberikan pertanyaan berbentuk bahasa yang menguji pemahaman kosakata, fakta, atau informasi umum yang diperoleh dari pengalaman dan hasil belajarnya. Tes ini bertujuan mengukur tingkat pemahaman dan pengetahuan yang diperoleh peserta melalui pengalaman belajar.
3. *Quantitative Reasoning* (QR) adalah kategori tes kognitif yang bertujuan untuk mengukur kemampuan individu dalam memahami, menalar, dan

memecahkan masalah yang berkaitan dengan konsep kuantitas, jumlah, dan hubungan numerik. Untuk soal (QR) *nonverbal*, peserta diminta untuk menalar hubungan kuantitatif secara visual, misalnya dengan membandingkan, mengurutkan, atau melanjutkan pola jumlah. Pada soal QR *verbal*, peserta diminta menalar hubungan *kuantitatif* melalui penyajian matematis. Pada tes kognitif (QR) mengukur nilai logika *kuantitatif*, seperti membandingkan besaran, mengenali pola jumlah, atau menentukan hubungan numerik tertentu, bukan sekadar melakukan perhitungan mekanis.

4. *Visual-Spatial Processing* (VS) adalah kategori soal tes kognitif yang bertujuan untuk memahami posisi, arah, dan bentuk benda, serta bagaimana benda tersebut berubah jika diputar atau dipindahkan. VS dibagi menjadi dua kategori yaitu verbal dan nonverbal. Pada kedua kategori tersebut, peserta diminta memahami perubahan posisi atau pergerakan suatu objek, baik melalui penjelasan verbal maupun representasi visual.
5. *Working Memory* (WM) adalah kategori soal tes kognitif yang mengandalkan memori dan ketelitian seseorang dalam waktu singkat, kategori ini memiliki dua tipe soal yaitu, *Working Memory* (WM) *verbal* dan *Working Memory* (WM) *nonverbal*. Pada kategori ini, peserta diminta mengingat informasi dalam bentuk kata, angka, atau instruksi lisan maupun visual dalam jangka waktu singkat. Dengan tujuan untuk menilai kapasitas memori jangka pendek.

Untuk mempermudah pemahaman mengenai perbedaan karakteristik soal pada masing-masing kategori kemampuan kognitif, contoh soal terdapat di halaman lampiran, dan berikut ringkasan ciri utama soal disajikan dalam Tabel 2.2

Tabel 2. 2 Ringkasan jenis soal berdasarkan bentuk *verbal* dan *nonverbal*

No	Kategori	*Nonverbal*	*Verbal*	Tujuan
1	*Fluid Reasoning (FR)*	Menemukan pola atau hubungan dari gambar	Menarik kesimpulan dari kalimat atau kata	Menilai kemampuan berpikir dan memecahkan masalah baru
2	*Knowledge (KN)*	Mengenali benda atau informasi dari gambar	Menjawab arti kata, fakta, atau informasi umum	Menilai apa yang sudah diketahui atau dipelajari
3	*Quantitative Reasoning (QR)*	Membandingkan atau memahami jumlah lewat gambar	Menghitung atau memahami soal angka	Menilai pemahaman tentang jumlah dan angka
4	*Visual-Spatial Processing (VS)*	Memahami bentuk, posisi, atau putaran dari gambar	Memahami posisi/arah dari penjelasan kata	Menilai pemahaman posisi, arah, dan bentuk
5	*Working Memory (WM)*	Mengingat urutan atau posisi gambar	Mengingat kata, angka, atau instruksi	Menilai kemampuan mengingat dalam waktu singkat

2.2 Pendidikan Orang Tua

Pendidikan orang tua dipandang sebagai salah satu indikator penting dalam merepresentasikan status sosial ekonomi keluarga dan berkaitan dengan perkembangan kognitif anak. Penelitian (Klein & Kühhirt, 2023) menunjukkan bahwa anak dari orang tua dengan tingkat pendidikan lebih tinggi cenderung memiliki capaian kemampuan kognitif yang lebih baik dibandingkan dengan anak dari latar pendidikan orang tua yang lebih rendah. Temuan tersebut mengindikasikan bahwa pendidikan orang tua berperan dalam membentuk kualitas lingkungan belajar di rumah, intensitas stimulasi kognitif, serta pola interaksi yang mendukung perkembangan intelektual anak. Hubungan ini diperkuat oleh penelitian berbasis tes IQ terstandar menggunakan *Wechsler Intelligence Scale*.

Hasilnya menunjukkan bahwa tingkat pendidikan ayah dan ibu berkorelasi dengan skor Verbal IQ, Performance IQ, dan Full-Scale IQ anak. Pendidikan yang lebih rendah cenderung berkaitan dengan skor IQ yang lebih rendah (Cermakova et al., 2023). Selain itu, latar belakang pendidikan keluarga juga berhubungan dengan kemampuan kognitif dasar, seperti *working memory*, yang merupakan komponen penting dalam pembentukan fungsi intelektual (Shatskaya et al., 2025).

Dengan mempertimbangkan teori dan hasil penelitian terdahulu, penelitian ini merumuskan hipotesis bahwa pendidikan orang tua digunakan sebagai salah satu variabel prediktor dalam pemodelan klasifikasi kategori IQ anak. Hipotesis tersebut diuji melalui pendekatan pemodelan berbasis machine learning menggunakan *Multilayer Perceptron* (MLP). Data pendidikan ayah dan ibu digunakan sebagai variabel *input*, sedangkan kategori IQ anak yang diperoleh dari *Stanford–Binet Intelligence Scales Fifth Edition* (SB5) digunakan sebagai variabel target. Pendekatan ini cocok dengan bukti statistik mengenai keterkaitan pendidikan orang tua dan skor IQ anak dengan memanfaatkan MLP dalam memodelkan hubungan nonlinier untuk mengidentifikasi pola kompleks sehingga diharapkan menghasilkan akurasi klasifikasi yang lebih optimal.

2.3 Machine Learning dan Multilayer Perceptron (MLP)

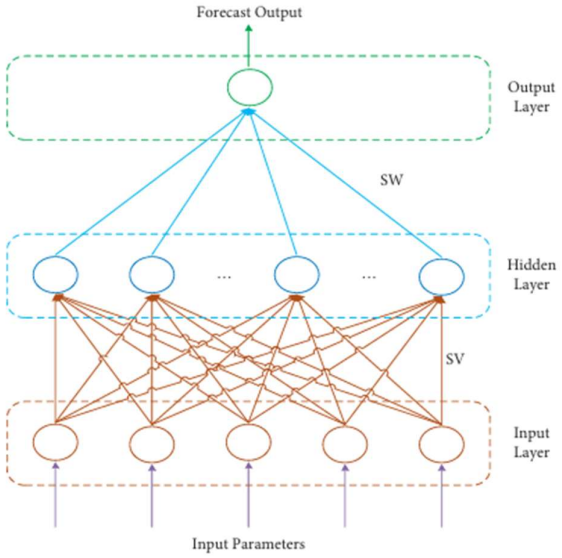
Machine learning merupakan salah satu kecerdasan buatan yang berfokus pada pengembangan model komputasi untuk mempelajari pola dari data. Model tersebut digunakan untuk menghasilkan prediksi atau keputusan secara otomatis. Pendekatan ini mengandalkan proses pelatihan algoritma menggunakan data sebagai dasar pembentukan model, sehingga sistem tidak lagi bergantung pada

aturan yang ditentukan secara manual, melainkan mengalami peningkatan performa melalui mekanisme pembelajaran berbasis pengalaman empiris (Kufel et al., 2023). Machine learning banyak dimanfaatkan untuk menyelesaikan tugas prediksi dan klasifikasi karena kemampuannya dalam merepresentasikan hubungan antarvariabel yang kompleks dan bersifat nonlinier. Salah satu metode yang sering digunakan adalah Artificial Neural Network (ANN), yang telah diimplementasikan pada berbagai bidang, seperti *classification, regression, prediction, smart grid, natural language processing, image processing,* hingga *medical diagnosis* (Madhiarasan & Louzazni, 2022). Selain itu, penggunaan machine learning dibidang medis menunjukkan keefektivitasnya dalam mengolah banyak fitur secara bersamaan untuk menghasilkan prediksi secara akurat (Yu et al., 2024).

Artificial Neural Network (ANN) merupakan salah satu metode dalam *machine learning* yang mengadaptasi prinsip kerja jaringan saraf biologis manusia. Pemrosesan informasi dilakukan melalui sejumlah neuron buatan yang saling terhubung. Proses pembelajaran pada model ini memanfaatkan data pelatihan melalui penyesuaian bobot secara bertahap dan berulang, sehingga sistem dirancang mampu mengidentifikasi pola serta menekan tingkat kesalahan prediksi. Berkat karakteristik tersebut, ANN telah banyak dimanfaatkan dalam berbagai permasalahan komputasi, termasuk tugas klasifikasi, regresi, dan prediksi (Madhiarasan & Louzazni, 2022).

Salah satu arsitektur yang dirancang untuk menyesuaikan karakteristik data serta tujuan pemodelan adalah Multilayer Perceptron (MLP). Multilayer Perceptron (MLP) merupakan arsitektur *feedforward artificial neural network* yang tersusun

atas tiga komponen utama, yaitu lapisan input, satu atau beberapa lapisan tersembunyi (*hidden layer*), serta lapisan output. Setiap lapisan berperan dalam mentransformasikan sinyal secara bertahap melalui proses komputasi berurutan, sehingga model mampu merepresentasikan keterkaitan antarvariabel secara lebih mendalam (Madhiarasan & Louzazni, 2022). Contoh arsitektur MLP dapat dilihat pada gambar berikut:



Gambar 2. 1 Contoh gambar arsitektur mlp

Sumber: (Madhiarasan & Louzazni, 2022)

1. *Input Layer*, Lapisan yang menerima fitur-fitur numerik sebagai representasi data masukan.
2. *Hidden Layer,* Lapisan tersembunyi yang melakukan proses komputasi berbobot terhadap input, kemudian memprosesnya melalui fungsi aktivasi non-linear sehingga jaringan mampu mengenali pola yang lebih kompleks.
3. *Output Layer,* Lapisan terakhir yang menghasilkan keluaran berupa nilai prediksi atau probabilitas kelas sebagai hasil klasifikasi.

Selain struktur arsitektur, MLP juga memiliki mekanisme pembelajaran untuk menyesuaikan bobot jaringan. Secara umum, proses pembelajaran memiliki dua tahap utama, yaitu:

1. *Feedforward*, Pada fase *feedforward*, data yang merepresentasikan fitur-fitur numerik dialirkan ke lapisan input jaringan. Informasi tersebut kemudian diteruskan ke lapisan tersembunyi (*hidden layer)*, di mana setiap neuron memproses seluruh nilai input melalui perkalian dengan bobot terkait dan penambahan bias untuk membentuk nilai agregat berbobot. Setelah nilai agregat berhasil ditransformasikan menggunakan fungsi aktivasi *nonlinier* seperti sigmoid atau ReLU, kemudian diteruskan ke lapisan output guna menghasilkan estimasi akhir, yang dapat berupa nilai kontinu maupun probabilitas keanggotaan kelas (Madhiarasan & Louzazni, 2022). Dengan demikian, tahap *feedforward* bertujuan untuk menghasilkan prediksi awal berdasarkan konfigurasi bobot jaringan pada saat proses berlangsung.
2. *Backpropagation*, Setelah tahap *feedforward* menghasilkan prediksi awal, *Multilayer perceptron* melanjutkan proses pembelajaran melalui mekanisme *backpropagation* untuk meningkatkan kinerja model. Pada fase ini, perbedaan antara hasil prediksi dan nilai target dihitung sebagai besaran kesalahan (*error*). Nilai kesalahan tersebut disalurkan kembali dari lapisan output ke lapisan tersembunyi hingga lapisan input untuk menyesuaikan bobot jaringan secara bertahap dan berulang. Penyesuaian bobot ini diarahkan untuk menurunkan tingkat kesalahan pada proses prediksi

berikutnya sehingga performa model mengalami peningkatan secara progresif (Madhiarasan & Louzazni, 2022).

Penelitian oleh (Madhiarasan & Louzazni, 2022) menunjukkan bahwa jaringan saraf dengan arsitektur multilayer memiliki kemampuan dalam menangani permasalahan dengan tingkat kompleksitas tinggi serta mampu merepresentasikan hubungan yang bersifat linear maupun nonlinier. Pada penelitian (Liu, 2024) yang berjudul *Multilayer perceptron-based literature reading preferences predict anxiety and depression in university students*, penerapan *Multilayer Perceptron* (MLP) menghasilkan tingkat akurasi klasifikasi yang tinggi pada data sosial dan psikologis. Menunjukkan bahwa MLP mampu mempelajari pola data yang kompleks, sehingga dinilai relevan untuk diterapkan dalam penelitian ini guna mengklasifikasikan kategori IQ anak berdasarkan variabel pendidikan orang tua.

Tahapan evaluasi performa model Multilayer Perceptron (MLP) dalam mengklasifikasikan kategori IQ anak berdasarkan dataset SB5 pada 5 kategori, *Moderate intellectual disability, Mild intellectual disability, Below average intelligence, Average intelligence, Above-average intelligence*. Evaluasi ini dilakukan dengan membandingkan label prediksi yang dihasilkan model terhadap label sebenarnya pada setiap kategori, dengan metrik evaluasi berupa Accuracy, F1-Score, dan AUC-ROC.

Evaluasi performa model klasifikasi diawali dengan penyusunan *confusion matrix*. Secara umum, hasil klasifikasi *confusion matrix* pada setiap kelas dterdiri dari empat komponen, yaitu: *True Positive* (TP)*, False Positive* (FP)*, False Negative* (FN), *True Negative* (TN).

a. *True Positive* (TP): jumlah data pada suatu kategori yang diprediksi benar sebagai kategori tersebut.
b. *False Positive* (FP): data dari kategori lain yang salah diprediksi masuk ke kategori tersebut.
c. *False Negative* (FN): data dari kategori tersebut tetapi diprediksi sebagai kategori lain.
d. *True Negative* (TN): data di luar kategori tersebut yang juga tidak diprediksi sebagai kategori tersebut.

Tabel 2. 3 Confusion matrix

Kelas Aktual	**Kelas Prediksi**				
	Moderate intellectual disability (P1)	***Mild intellectual disability (P2)***	***Below average intelligence (P3)***	***Average intelligence (P4)***	***Above-average intelligence (P5)***
Moderate intellectual disability (A1)	a_{11}	a_{12}	a_{13}	a_{14}	a_{15}
Mild intellectual disability (A2)	a_{21}	a_{22}	a_{23}	a_{24}	a_{25}
Below average intelligence (A3)	a_{31}	a_{32}	a_{33}	a_{34}	a_{35}
Average intelligence (A4)	a_{41}	a_{42}	a_{43}	a_{44}	a_{45}
Above-average intelligence (A5)	a_{51}	a_{52}	a_{53}	a_{54}	a_{55}

Tabel 2.3 confusion matrix, pada setiap baris merepresentasikan kelas sebenarnya (*aktual*), sedangkan setiap kolom menunjukkan kelas hasil prediksi model. Nilai pada diagonal utama (a_{ii}) menunjukkan jumlah data yang berhasil diklasifikasikan dengan benar pada masing-masing kategori IQ. Sebaliknya, nilai di luar diagonal menunjukkan kesalahan klasifikasi. Sebagai contoh, nilai a_{14} menyatakan jumlah anak dengan kategori *Moderate intellectual disability* yang

diprediksi sebagai *Average intelligence*, sedangkan nilai a_{53} menunjukkan jumlah anak dengan kategori *Above-average intelligence* yang diprediksi sebagai *Below average intelligence*.

Informasi yang diperoleh dari *confusion matrix* digunakan sebagai dasar dalam perhitungan metrik evaluasi performa model seperti accuracy, precision, recall, dan F1-score. Setiap metrik memberikan sudut pandang yang berbeda terhadap kemampuan klasifikasi.

1. *Accuracy* digunakan untuk mengukur tingkat ketepatan model secara keseluruhan dalam melakukan klasifikasi. *Accuracy* dihitung dengan,

$$\text{Accuracy} = \frac{\sum_{i=1}^{n} a_{ii}}{\sum_{i=1}^{n} \sum_{j=1}^{n} a_{ij}}$$

   dengan a_{ii}menyatakan jumlah data pada kelas ke-i yang diprediksi benar dan a_{ij}menyatakan jumlah seluruh data. Sebagai ilustrasi, pada klasifikasi lima kelas (A, B, C, D, dan E), nilai accuracy diperoleh dari penjumlahan seluruh prediksi benar pada diagonal matriks ($a_{11}+a_{22}+a_{33}+a_{44}+a_{55}$) dibandingkan dengan jumlah seluruh data.

2. *Precision* digunakan untuk mengukur ketepatan model dalam memberikan label pada suatu kelas, yang dihitung dengan,

$$\text{Precision} = \frac{TP}{TP + FP}$$

   Sebagai ilustrasi pada lima kelas (A, B, C, D, dan E), precision kelas A diperoleh dari perbandingan antara jumlah data yang benar diprediksi

sebagai A (a_{11}) dengan seluruh data yang diprediksi sebagai A ($a_{11}+a_{21}+a_{31}+a_{41}+a_{51}$).

3. *Recall* digunakan untuk mengukur kemampuan model dalam mengenali seluruh data yang benar-benar termasuk dalam suatu kelas.

$$\text{Recall} = \frac{TP}{TP + FN}$$

Sebagai ilustrasi pada lima kelas (A, B, C, D, dan E), recall kelas A diperoleh dari perbandingan antara jumlah data yang benar diprediksi sebagai A (a_{11}) dengan seluruh data yang sebenarnya termasuk kelas A ($a_{11}+a_{12}+a_{13}+a_{14}+a_{15}$).

4. *F1-score* digunakan untuk mengukur keseimbangan antara precision dan recall dalam satu nilai. Metrik ini diperlukan karena model dapat memiliki *precision* tinggi tetapi *recall* rendah, atau sebaliknya.

$$F1 = 2 \cdot \frac{Precision \cdot Recall}{Precision + Recall}$$

Sebagai ilustrasi pada lima kelas (A, B, C, D, dan E), nilai *F1-score* untuk kelas A diperoleh dari penggabungan nilai precision dan recall kelas A yang telah dihitung sebelumnya.

5. Selain accuracy dan F1-score, penelitian ini juga menggunakan metrik AUC-ROC. perhitungan AUC dilakukan dengan pendekatan *One-vs-Rest* (OvR), yang dimana setiap kelas dibandingkan dengan seluruh kelas lainnya. Nilai AUC dihitung berdasarkan skor *probabilitas* prediksi model,

kemudian dirata-ratakan menggunakan *macro-average* agar setiap kelas memiliki kontribusi yang sama terhadap nilai akhir.

Berdasarkan metrik evaluasi tersebut, setiap skenario menghasilkan konfigurasi performa yang berbeda-beda, kemudian digunakan untuk membandingkan kemampuan klasifikasi antar arsitektur model *Multilayer perceptron* (MLP).

2.4 Penelitian Terkait

Penelitian yang dilakukan (Olech et al., 2024) berjudul "*Intelligence assessment of Children & Youth Benefiting from Psychological-Educational Support System in Poland*" mempresentasikan dan memvalidasi dataset yang berisi 419.135 hasil asesmen inteligensi anak dan remaja di Polandia menggunakan SB5. Fokus penelitian ini bukan pada pengujian hipotesis, melainkan pada penjelasan proses pengumpulan, pengolahan, serta verifikasi kualitas data. Data dikumpulkan dari berbagai layanan psikologis dan Pendidikan. Tes SB5 dilakukan secara individual oleh psikolog terlatih, disertai pengumpulan data demografis dan konversi skor secara otomatis melalui aplikasi daring. Dataset yang dihasilkan mencakup 47 variabel yang terdiri atas data demografis, nilai mentah, skor standar, dan IQ. Karena sampel berasal dari populasi rujukan layanan psikologis, dataset ini merepresentasikan populasi klinis dab bukan populasi normatif. Validasi struktural dilakukan melalui CFA yang menunjukkan kecocokan model sangat baik, mengonfirmasi stabilitas dua faktor utama SB5 (verbal dan non-verbal).

Penelitian yang berjudul "*Heterogeneity of Cognitive Profiles in Children and Adolescents with Mild Intellectual Disability (MID)*" oleh (Sajewicz-Radtke et

al., 2022) bertujuan memahami keberagaman kemampuan kognitif pada anak dan remaja dengan kategori *Mild Intellectual Disability* (MID), karena kategori ini sering dianggap homogen berdasarkan rentang skor IQ 55-69. Data yang digunakan merupakan hasil asesmen resmi yang dilakukan oleh psikolog profesional di pusat layanan psikologi pedagogik Polandia pada periode 2019-2022, dengan total 16.411 peserta berusia 7-18 tahun. Peneliti menganalisis skor dari 10 subtes untuk mengidentifikasi pola kemampuan yang serupa. Setelah peserta dibagi dalam tiga kelompok usia, analisis klaster dilakukan dan hasil *NbClust* menunjukkan bahwa tiga klaster merupakan jumlah optimal untuk semua kelompok usia. Metode *k-means* kemudian digunakan untuk membentuk klaster berdasarkan kesamaan pola skor. Hasil analisis menunjukkan adanya tiga profil kognitif yang konsisten di setiap kelompok usia. Profil pertama ditandai kemampuan verbal dan non-verbal yang sangat rendah, mengindikasikan kemungkinan keterlambatan perkembangan bahasa. Profil kedua memiliki kemampuan non-verbal yang relatif lebih baik dibanding verbal, yang diinterpretasikan sebagai kurangnya stimulasi atau pengalaman belajar. Profil ketiga menunjukkan kekuatan relatif pada aspek verbal dibanding non-verbal, meskipun tetap berada dalam rentang kemampuan rendah, keterlambatan bahasa dan kurang stimulasi muncul secara konsisten dan stabil di seluruh rentang usia, menunjukkan adanya pola perkembangan yang stabil di populasi MID.

Penelitian yang dilakukan (Sajewicz-Radtke, Łada-Maśko, Olech, Leoniak, et al., 2025) berjudul "*Declarative memory profiles in children with non-specific intellectual disability: a cluster analysis approach*", mengidentifikasi pola atau

profil memori deklaratif pada anak dengan kategori *non-specific intellectual disability* (NSID) serta mencocokkan dengan skor kecerdasan dari tes SB5. Seluruh peserta mengikuti tes memori TOMAL-2, sebagian di antaranya juga menjalani asesmen IQ menggunakan SB5. Data ini diambil dari 114 anak berusia 10–17 tahun, yang kemudian dianalisis menggunakan metode *k-means clustering* untuk mengelompokkan berdasarkan skor memori, kemudian dilakukan analisis lanjutan pada subsampel 68 anak dengan memasukkan skor IQ verbal dan nonverbal. Penelitian menunjukkan adanya dua profil memori utama, satu kelompok dengan skor memori yang rendah pada seluruh aspek, dan kelompok lainnya dengan kemampuan memori yang lebih tinggi. Indeks yang paling membedakan kedua kelompok adalah *Learning Index*, yaitu ukuran efisiensi belajar, bukan *Verbal Delayed Recall*. Ketika skor IQ dimasukkan ke dalam analisis, hanya Nonverbal IQ yang memberikan perbedaan signifikan antarkelompok, sedangkan Verbal IQ tidak menunjukkan perbedaan yang berarti. Hasil ini mengindikasikan bahwa kemampuan belajar dan memori memberikan informasi yang lebih beragam dan lebih bermakna dibanding skor IQ saja dalam memahami profil kognitif anak dengan NSID.

Penelitian yang berjudul "*Measuring intelligence in Autism and ADHD: Measurement invariance of the-Binet 5th edition and impact of subtest scatter on abbreviated IQ accuracy*" oleh (Stephenson et al., 2023) menilai apakah tes SB5 bekerja secara konsisten pada anak-anak dengan ASD dan ADHD, serta menguji seberapa akurat ABIQ, dalam memprediksi skor IQ penuh (FSIQ). Data ini diambil dari 1679 anak usia 2-16 tahun yang menjalani tes SB5 dalam evaluasi klinis. Untuk

melihat apakah SB5 mengukur IQ secara konsisten pada kedua kelompok (ASD dan ADHD), peneliti menggunakan *multigroup confirmatory factor analysis* (CFA) dengan model bifaktor, dan hasilnya menunjukkan bahwa semua tingkat *measurement invariance* terpenuhi, sehingga SB5 dianggap tidak bias. Untuk menilai akurasi ABIQ, penelitian ini menggunakan model regresi yaitu regresi linear dan kuadratik. FSIQ sebagai variabel yang diprediksi dari ABIQ, nilai scatter, dan interaksi ABIQ dengan scatter. Model kuadratik memberikan hasil terbaik dan menunjukkan bahwa meskipun ABIQ cukup akurat secara umum, akurasinya berkurang ketika scatter meningkat, terutama karena scatter membuat ABIQ cenderung meng-overestimate FSIQ. Analisis tambahan dengan *multinomial logistic regression* juga menunjukkan bahwa scatter merupakan faktor utama yang memprediksi kesalahan estimasi tersebut.

Penelitian yang dilakukan (Reinaldi & Hidayat, 2021a) dengan judul "*Stanford-Binet Intelligence Scale Form L-M Predictive Power on Academic Achievement*" menilai apakah skor IQ dari Stanford-Binet Form L-M mampu memprediksi prestasi akademik siswa sekolah dasar. Studi ini dilakukan karena tes tersebut masih sering digunakan di Indonesia meskipun penelitian validitas terbarunya terbatas. Dengan menggunakan metode kuantitatif melalui analisis regresi sederhana, peneliti memanfaatkan data sekunder yaitu skor IQ siswa kelas 1-2 dari UKP UGM serta nilai Ujian Akhir siswa kelas 3-5 untuk mata pelajaran Matematika, Bahasa Indonesia, dan PKn. Sampel penelitian mencakup 102 siswa, dan skor IQ dianalisis sebagai prediktor nilai ujian. Hasil penelitian menunjukkan bahwa SB L-M memiliki kemampuan prediktif, tetapi pada tingkat yang rendah

hingga sedang. Skor IQ hanya mampu menjelaskan sekitar 4,3% sampai 25,4% variasi nilai akademik, sehingga sebagian besar prestasi dipengaruhi oleh faktor-faktor lain di luar kecerdasan.

Penelitian yang berjudul "*Association between parental education level and intelligence quotient of children referred to the mental healthcare system: a crosssectional study in Poland*" oleh (Sajewicz-Radtke, Łada-Maśko, Olech, Jurek, et al., 2025) bertujuan memahami keterkaitan antara tingkat pendidikan orang tua dan skor IQ pada anak-anak yang dirujuk ke layanan kesehatan mental di Polandia. Penelitian ini menggunakan desain *cross-sectional* dengan pendekatan analisis regresi, data yang digunakan berasal dari 80.303 anak berusia 3-18 tahun yang memiliki hasil tes kecerdasan lengkap beserta informasi mengenai pendidikan orang tua. Kemampuan intelektual diukur menggunakan Stanford–Binet 5 (SB5), yang memberikan gambaran komprehensif mengenai kapasitas kognitif anak. Melalui analisis regresi, peneliti menemukan bahwa semakin tinggi tingkat pendidikan orang tua, semakin tinggi skor IQ anak, dengan variabel pendidikan ibu memberikan pengaruh paling kuat. Faktor jenis kelamin tidak menunjukkan peran yang signifikan dalam hubungan tersebut. Namun, faktor usia anak memiliki dampak penting, pada keluarga dengan tingkat pendidikan rendah, skor IQ cenderung menurun seiring bertambahnya usia, sedangkan pada keluarga dengan pendidikan tinggi, skor IQ justru menunjukkan kecenderungan meningkat.

Tabel 2. 4 Penelitian terkait tes SB5

Peneliti	Judul Penelitian	Metode	Hasil	Perbedaan Penelitian
Olech et l. (2024)	*Intelligence Assessment of Children & Youth*	Analisis deskriptif, penyusunan & validasi dataset	Dataset 419.135 asesmen, 47 variabel, CFA mendukung struktur	Penelitian Olech et al. (2024) berfokus membuat dan memvalidasi dataset

Peneliti	Judul Penelitian	Metode	Hasil	Perbedaan Penelitian
	Benefiting from Psychological-Educational Support System in Poland	SB5 berskala besar, CFA untuk validasi struktur SB5.	dua-faktor SB5 (verbal & nonverbal). Dataset berasal dari populasi klinis.	SB5 terbesar, tidak membangun model atau klasifikasi. Sedangkan penelitian ini menggunakan dataset SB5 (hasil publikasi ini) untuk membangun model klasifikasi IQ otomatis dengan MLP berdasarkan pendidikan orang tua.
Sajewicz-Radtke et al. (2022)	*Heterogeneity of Cognitive Profiles in Children and Adolescents with MID*	*K-means cluster* analysis pada 16.411 anak MID.	Identifikasi 3 profil kognitif MID, perbedaan konsisten di seluruh usia (verbal rendah, nonverbal relatif baik, atau sebaliknya).	Penelitian Sajewicz-Radtke et al. (2022) menganalisis *profil kognitif* anak MID menggunakan 10 subtes SB5, berfokus pada keragaman kemampuan verbal & nonverbal. Sedangkan penelitian ini tidak menganalisis profil kognitif, tetapi mengklasifikasikan kategori IQ (Moderate intellectual disability, Mild intellectual disability, Below average intelligence, Average intelligence, Above-average intelligence) berdasarkan pendidikan orang tua menggunakan MLP.
Sajewicz-Radtke et al. (2025)	*Declarative Memory Profiles in Children with NSID*	*Cluster analysis* TOMAL-2 subsampel dianalisis dengan SB5 (verbal & nonverbal IQ).	Dua profil memori (lemah vs kuat), Learning Index menjadi pembeda utama, Nonverbal IQ lebih berpengaruh daripada Verbal IQ.	Penelitian j Sajewicz-Radtke et al. (2025) berfokus pada *profil memori deklaratif* dan hubungan dengan IQ verbal/nonverbal, tidak membangun model prediksi atau klasifikasi IQ. Penelitian ini menggunakan dataset SB5 untuk memprediksi kategori IQ berdasarkan variabel pendidikan

Peneliti	Judul Penelitian	Metode	Hasil	Perbedaan Penelitian
				orang tua melalui machine learning.
Stephenson et al. (2023)	*MeasuringintelligenceinAutismandADHD: Measurement invarianceoftheBinet5theditionandimpactofsubtestscatteron abbreviatedIQ accuracy*	Multigroup CFA, *regresi linear* & kuadratik untuk menguji akurasi ABIQ.	SB5 invariant untuk ASD & ADHD, ABIQ cukup akurat tetapi *scatter tinggi menyebabkan overestimasi IQ*.	Penelitian Stephenson et al. (2023) mengevaluasi keakuratan alat tes SB5, fokus pada kategori ASD/ADHD. Penelitian ini tidak menilai keakuratan alat tes, melainkan membangun model klasifikasi IQ menggunakan input pendidikan orang tua.
Reinaldi & Hidayat (2021)	*Stanford-Binet Intelligence Scale Form L-M Predictive Power on Academic Achievement*	Regresi sederhana, SB L-M sebagai prediktor nilai akademik.	IQ hanya menjelaskan 4.3%–25.4% prestasi akademik, validitas prediktif rendah sampai sedang.	Penelitian pada jurnal ini berfokus pada prediksi prestasi akademik dari IQ (Form L-M). Tidak terkait SB5 dan tidak memakai machine learning. Sedangakan penelitian ini berfokus klasifikasi kategori IQ dari pendidikan orang tua menggunakan algoritma MLP, bukan prestasi akademik.
Sajewicz-Radtke et al. (2025)	*Association between parental education level and intelligence quotient of children referred to the mental healthcare system: a cross sectional study in Poland*	Cross-sectional, analisis regresi pada 80.303 anak.	Pendidikan orang tua berpengaruh signifikan pada IQ, pendidikan ibu menjelaskan 18.23% variasi IQ	Penelitian Sajewicz-Radtke et al. (2025) menunjukkan hubungan linear antara pendidikan orang tua dan IQ (regresi). Pada penelitian ini mengembangkan model klasifikasi kategori IQ (Moderate intellectual disability, Mild intellectual disability, Below average intelligence, Average intelligence, Above-average intelligence) dari pendidikan orang tua

Peneliti	Judul Penelitian	Metode	Hasil	Perbedaan Penelitian
				menggunakan *machine learning* yang mampu menangkap hubungan non-linear.

Tabel 2.4 memaparkan ringkasan penelitian yang berkaitan dengan asesmen kecerdasan intelektual anak menggunakan Stanford-Binet Intelligence Scales (SB5) serta faktor-faktor yang memengaruhi variasi skor IQ anak. Tabel tersebut memuat informasi mengenai peneliti, judul penelitian, metode yang digunakan, hasil utama, serta perbedaan antara penelitian sebelumnya dan penelitian yang dilakukan. Dalam tabel ini terdapat penelitian yang menggunakan beberapa metode seperti analisis regresi, analisis klaster, confirmatory factor analysis (CFA), serta validasi struktur tes. Informasi pada tabel ini memberikan gambaran menyeluruh tentang bagaimana SB5 telah digunakan dalam berbagai konteks penelitian.

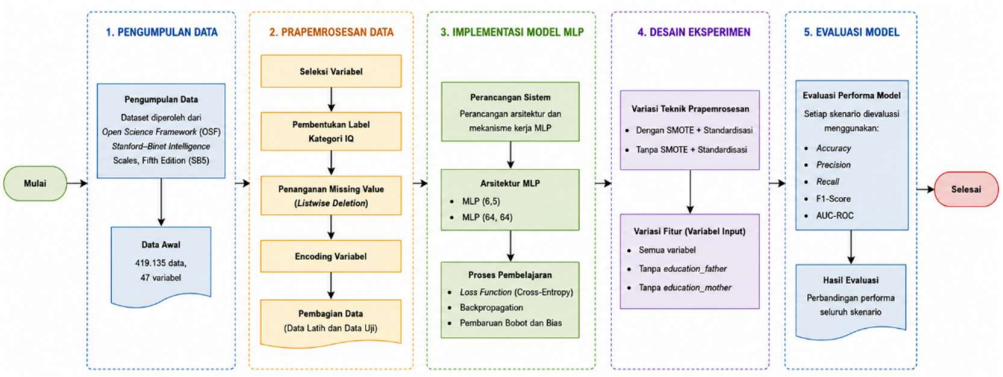
Berdasarkan perbandingan penelitian tersebut, sebagian besar studi masih menggunakan metode deskriptif dan analitis yang hanya menangkap hubungan linear, tanpa menggunakan model klasifikasi otomatis menggunakan *machine learning,* belum ada penelitian yang menggunakan variabel pendidikan orang tua sebagai input utama untuk mengklasifikasikan kategori IQ anak. Selain itu, penelitian sebelumnya tidak menerapkan model *machine learning* khususnya MLP untuk menangkap hubungan non-linear antara pendidikan orang tua dan IQ anak. Oleh karena itu, penelitian ini berupaya mengisi celah tersebut dengan menerapkan model *machine learning* berbasis *Multilayer Perceptron* (MLP) untuk menangkap hubungan *nonlinear* antara pendidikan orang tualanju dan IQ anak dalam mengklasifikasikan kategori IQ anak sesuai pada tabel 2.1.

BAB III

DESAIN DAN IMPLEMENTASI

3.1 Desain Penelitian

Penelitian ini menggunakan pendekatan kuantitatif dengan desain penelitian eksperimental berbasis komputasi. Desain ini dipilih untuk membangun, melatih, dan menguji model Jaringan Saraf Tiruan dengan arsitektur Multilayer Perceptron (MLP) dalam mengklasifikasikan kategori kemampuan intelektual anak (*Moderate intellectual disability, Mild intellectual disability, Below average intelligence, Average intelligence, Above-average intelligence*) berdasarkan tingkat pendidikan orang tua. Metode eksperimen ini diterapkan untuk mengevaluasi performa model MLP dalam melakukan klasifikasi serta mengukur tingkat akurasinya menggunakan metrik Accuracy, F1-Score, dan AUC-ROC. Secara sistematis, alur dan tahapan penelitian ini meliputi proses pengumpulan data, prapemrosesan data, implementasi model, perancangan skenario eksperimen, hingga evaluasi performa model. Rancangan penelitian secara keseluruhan ditunjukkan pada Gambar 3.1.



Gambar 3. 1 Alur Penelitian

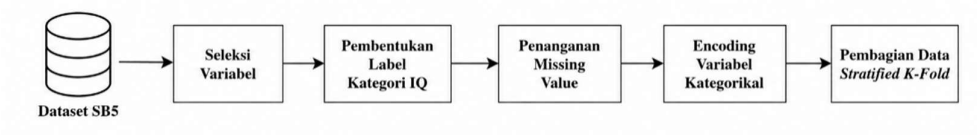
3.1.1 Pengumpulan Data

Studi ini menggunakan data yang bersifat kuantitatif sekunder yang diperoleh dari Open Science Framework (OSF), yang dapat diakses melalui tautan https://osf.io/qatrz/files?view_only=a9a4ed5c7dd74af0b338ca120b85b7bc. Data ini merupakan hasil tes Stanford–Binet Intelligence Scales, Fifth Edition (SB5) yang dikumpulkan dalam penelitian berjudul *"Association between parental education level and intelligence quotient of children referred to the mental healthcare system: a cross-sectional study in Poland"* oleh Sajewicz-Radtke et al. (2025). Dataset ini terdiri dari 419.135 entri dan 47 variabel yang mencakup pendidikan orang tua, data demografis anak, hasil subtes kognitif, serta skor IQ *verbal* dan *nonverbal*. Berdasarkan tujuan penelitian, digunakan lima variabel utama yang terdiri atas *education_mother, education_father, age_years, gender, dan ALL_SC_IQ*. Variabel *education_mother* dan *education_father* digunakan untuk merepresentasikan tingkat pendidikan orang tua, sedangkan *age_years* dan *gender* digunakan sebagai variabel demografis anak. Variabel *ALL_SC_IQ* digunakan sebagai dasar pembentukan kategori kemampuan intelektual anak yang menjadi target klasifikasi dalam penelitian ini.

3.1.2 Prapemrosesan Data

Tahapan prapemrosesan data dilakukan untuk memastikan dataset layak digunakan dalam proses klasifikasi menggunakan *Multilayer Perceptron* (MLP). Dataset awal yang diperoleh dari *Open Science Framework* (OSF) masih mengandung nilai kosong (*missing values*) pada beberapa variabel serta tipe data yang belum sesuai dengan kebutuhan model. Oleh karena itu, proses ini mencakup

seleksi variabel input, pembentukan label kategori IQ, penanganan *missing value*, *encoding* variabel kategorikal, serta pembagian data latih dan uji, seperti pada Gambar 3.2



Gambar 3. 2 Tahapan prapemrosesan

3.1.2.1 Seleksi Variabel

Tahap seleksi variabel dilakukan untuk menentukan fitur yang digunakan sebagai input dan target dalam model klasifikasi *Multilayer perceptron* (MLP). Dataset *Stanford–Binet Intelligence Scales Fifth Edition* (SB5) yang digunakan dalam penelitian ini terdiri atas 47 variabel yang dikelompokkan ke dalam beberapa kategori, yaitu: Variabel identitas dan administratif (*id, timestamp*), Variabel demografis (*gender, age_years, age_group, residence, diagnosis, education_mother, education_father, education_parents*), Skor mentah subtes verbal dan nonverbal (*FR_VE_raw – WM_VE_raw, FR_NV_raw – WM_NV_raw*), *Scaled score* format 119 per subtes (*FR_VE_119 – WM_VE_119 dan FR_NV_119 – WM_NV_119*), Skor komposit *scaled score* (*VERB_SC_119, NVER_SC_119, ALL_SC_119*), Skor IQ terstandarisasi (*VERB_SC_IQ, NVER_SC_IQ, ALL_SC_IQ*). Seperti pada Gambar 3.3.

NAMA VARIABEL	TIPE DATA	DESKRIPSI	KEGUNAAN DALAM PENELITIAN
id	Integer	Nomor identifikasi unik responden	-
timestamp	String (datetime)	Waktu pencatatan data	-
gender	Kategorikal	Jenis kelamin anak	*Inputan*
age_years	Float	Usia anak dalam satuan tahun	*Inputan*
age_group	Kategorikal (ordinal)	Kelompok usia anak berdasarkan interval umur	-
residence	Kategorikal	Tempat tinggal	-
covid	Kategorikal	Status keterkaitan dengan periode Covid	-
diagnosis	Kategorikal	Diagnosa anak, misalnya "intellectual disability", "autism", dll.	-
education_mother	Kategorikal (ordinal)	Tingkat pendidikan terakhir ibu	*Inputan*
education_father	Kategorikal (ordinal)	Tingkat pendidikan terakhir ayah	*Inputan*
education_parents	Kategorikal	Ringkasan tingkat pendidikan orang tua	-
FR_NV_raw - WM_NV_raw	Integer	Skor mentah subtes *Nonverbal*	-
FR_VE_raw - WM_VE_raw	Integer	Skor mentah subtes *Verbal*	-
FR_VE_119 - WM_VE_119	Integer	Skor subtes *verbal* yang dikonversi ke *scaled score* format (119) dengan norma usia	-
FR_NV_119 - WM_NV_119	Integer	Skor subtes *nonverbal* yang dikonversi ke *scaled score* format (119) dengan norma usia	-
FR_SC_119 - WM_SC_119	Integer	*Scaled score* komposit per faktor kognitif SB5 (gabungan *verbal* + *nonverbal*)	-
NVER_SC_119	Integer	Gabungan dari 5 s*caled score* subtes *nonverbal* (FR_NV_119 - WM_NV_119)	-
VERB_SC_119	Integer	Gabungan dari 5 s*caled score* subtes *verbal* (FR_VE_119 - WM_VE_119)	-
ALL_SC_119	Integer	Seluruh *scaled score* (VERB_SC_119 dan NVER_SC_119)	-
FR_SC_IQ - WM_SC_IQ	Integer	Skor konversi per faktor kognitif SB5 (gabungan *verbal* + *nonverbal*) ke IQ *standard score*	-
NVER_SC_IQ	Integer	Konversi nilai dari (NVER_SC_119) ke IQ *standard score*	-
VERB_SC_IQ	Integer	Konversi nilai dari (VERB_SC_119) ke IQ *standard score*	-
ALL_SC_IQ	Integer	Konversi nilai dari (ALL_SC_119) ke IQ *standard score* (Mean 100, SD 15)	*Output* (sebagai dasar pelabelan kategori)

Gambar 3. 3 Dataset *stanford binet intelligence scales fifth edition*

Setelah seluruh variabel dalam dataset SB5 diidentifikasi, tidak semua variabel langsung dimasukkan ke dalam model *multilayer perceptron* (MLP). Meskipun model MLP mampu menangkap hubungan non-linear dan interaksi kompleks antar variabel, penggunaan seluruh fitur tanpa seleksi awal berpotensi

meningkatkan risiko *overfitting*. Karena itu, diperlukan seleksi variabel dengan analisis korelasi terhadap skor IQ total dalam bentuk kontinu, yaitu ALL_SC_IQ. Variabel ini dipilih karena memiliki skor distribusi standar dengan (mean = 100, SD = 15) sebelum dikategorikan. Teknik korelasi yang digunakan disesuaikan dengan skala data masing-masing variabel. Untuk variabel *interval* seperti *age_years*, digunakan metode *Pearson* karena mengasumsikan hubungan *linear* antar dua variabel kontinu. Untuk variabel *ordinal* seperti tingkat pendidikan orang tua, metode *spearman* lebih tepat karena tidak menganggap jarak antar kategori setara (Suherman et al., n.d.). Sementara itu, untuk variabel *biner* seperti *gender*, digunakan metode *point-biserial* karena metode ini cocok digunakan untuk korelasi dari satu variabel *dikotomis* dengan satu variabel *kontinu* (LeBlanc & Cox, 2017). Untuk lebih ringkasnya bisa dilihat pada tabel 3.2.

Tabel 3. 1 Ringkasan variabel dan metode korelasi

Variabel	**Tipe Data**	**Metode**	**Penjelasan**
education_mother	*ordinal*	*Spearman correlation*	Dikarenakan data pendidikan ibu berskala ordinal (bertingkat) untuk mengetahui rangking dari variabelnya.
education_father	*ordinal*	*Spearman correlation*	Dikarenakan data Pendidikan ayah berskala ordinal (bertingkat) untuk mengetahui rangking dari variabelnya.
age_years	*kontinu*	*Pearson correlation*	Dikarenakan kedua variabel (ALL_SC_IQ dan age_years) berskala kontinu, dan untuk mengukur hubungan linearnya.
gender	*dikotomis* (2 kategori)	*Point-biserial correlation*	Dikarenakan metode ini dipakai khusus untuk perhitungan biner dan kontinu.

Setelah korelasi tersebut mendapatkan hasilnya, beberapa variabel yang cocok digunakan untuk klasifikasi pada penelitian ini, yang digunakan sebagai variabel *input* dan variabel *output* bisa dilihat pada tabel 3.3 dan 3.4.

Tabel 3. 2 Variabel *input*

Variabel	Tipe Data	Keterangan Kegunaan
education_mother	Kategorikal	Merepresentasikan tingkat pendidikan ibu sebagai indikator lingkungan pendidikan keluarga.
education_father	Kategorikal	Merepresentasikan tingkat pendidikan ayah sebagai indikator latar belakang pendidikan keluarga.
age_years	Numerik	Mengontrol variasi usia anak yang dapat memengaruhi distribusi skor IQ dalam populasi klinis.
gender	Kategorikal	Variabel demografis untuk menangkap kemungkinan perbedaan distribusi kategori IQ berdasarkan jenis kelamin.

Keempat variabel tersebut digunakan sebagai fitur input pada lapisan input (input layer) pada model MLP. Pemilihan variabel pendidikan orang tua didasarkan pada temuan penelitian sebelumnya yang menunjukkan adanya hubungan signifikan antara tingkat pendidikan orang tua dan skor IQ anak (Sajewicz-Radtke, Łada-Maśko, Olech, Jurek, et al., 2025). Variabel usia dan jenis kelamin disertakan sebagai variabel kontrol untuk menangkap kemungkinan variasi demografis yang memengaruhi distribusi skor IQ.

Tabel 3. 3 Variabel *output*

Variabel	Tipe Data	Keterangan Kegunaan
ALL_SC_IQ	Numerik	Variabel ini digunakan sebagai dasar pembentukan label kategori IQ.

Variabel ALL_SC_IQ merupakan skor IQ terstandarisasi total pada SB5 dengan distribusi norma populasi (*Mean* = 100, SD = 15), yang digunakan sebagai dasar pembentukan label kategori IQ. Skor ini tidak langsung digunakan sebagai output model dalam bentuk numerik, melainkan terlebih dahulu dikonversi menjadi lima kategori kelas IQ sesuai dengan rentang klasifikasi yang telah ditetapkan.

3.1.2.2 Pembentukan Label Kategori IQ

Pada dataset SB5, skor hasil IQ anak disajikan dalam bentuk skor IQ terstandarisasi (ALL_SC_IQ). Namun, penelitian ini tidak menggunakan skor IQ dalam bentuk numerik sebagai target, melainkan mengubahnya menjadi variabel kategorikal yang merepresentasikan tingkat kemampuan intelektual anak. Penentuan rentang tersebut mengacu pada penelitian (Sajewicz-Radtke, Łada-Maśko, Olech, Jurek, et al., 2025). Sebagaimana disajikan dalam tabel 2.1. Proses pembentukan label dilakukan dengan membuat variabel baru, yaitu Kategori_IQ, yang dihasilkan melalui pengelompokan nilai skor IQ terstandarisasi ke dalam lima rentang kategori tersebut.

3.1.2.3 Penanganan Missing Value

Setelah proses seleksi variabel dan pembentukan label kategori IQ, tahap selanjutnya adalah penanganan *missing value*, Tahapan ini bertujuan membersihkan dataset dari nilai yang hilang dan diterapkan pada seluruh variabel *input* seperti *education_mother, education_father, age_years, dan gender*, yang digunakan untuk meminimalisir kesalahan input pada tahap pelatihan model dan juga menghindari hasil yang bias.

Metode yang digunakan untuk penanganan *missing value* adalah *listwise deletion*, yaitu teknik penghapusan seluruh baris data jika terdapat minimal satu nilai dari variabel yang hilang (Zhou et al., 2024). Penggunaan metode ini dipilih agar data yang digunakan lengkap dan valid, yang kemudian dalam proses pemodelan algoritma *multilayer perceptron* menghasilkan klasifikasi yang akurat.

3.1.2.4 Encoding Variabel

Tahap berikutnya adalah melakukan encoding variabel kategorikal, Model *Multilayer Perceptron* tidak dapat memproses variabel kategorikal karena jaringan saraf bekerja dalam representasi numerik. Pada penelitian ini, variabel *education_mother* dan *education_father* diubah menjadi variabel numerik menggunakan teknik *ordinal encoding* dengan penentuan urutan kategori secara manual (Seveso et al., 2020). Urutan kategori ditetapkan berdasarkan tingkat pendidikan, yaitu *primary or lower secondary*, *vocational*, *secondary*, dan *higher*. Setiap kategori kemudian dikonversi menjadi bilangan bulat berurutan mulai dari 0 hingga 3 sesuai tingkatannya, seperti pada tabel 3.4

Tabel 3. 4 Encoding variabel *education_father* & *education_mother*

No	**Kategori**	***Encoding***
1	*primary or lower secondary*	0
2	*vocational*	1
3	*secondary*	2
4	*higher*	3

Variabel *gender* yang memiliki dua kategori diubah menggunakan teknik *label encoding* ke dalam bentuk numerik *biner* (0 dan 1) seperti pada tabel 3.7

Tabel 3. 5 Encoding variabel *gender*

No	**Kategori**	***Encoding***
1	*female*	0
2	*male*	1

3.1.2.5 Pembagian Data Latih dan Uji

Setelah seluruh variabel diubah dalam bentuk numerik, tahap selanjutnya adalah proses validasi model menggunakan metode *Stratified K-Fold Cross-Validation*. Dataset dibagi menjadi *k* bagian (*fold*) dengan ukuran yang relatif sama. Proses pelatihan dan pengujian dataset dilakukan sebanyak *k* iterasi, pada setiap iterasi (satu fold) digunakan sebagai data validasi, sedangkan *k-1* fold (seluruh fold)

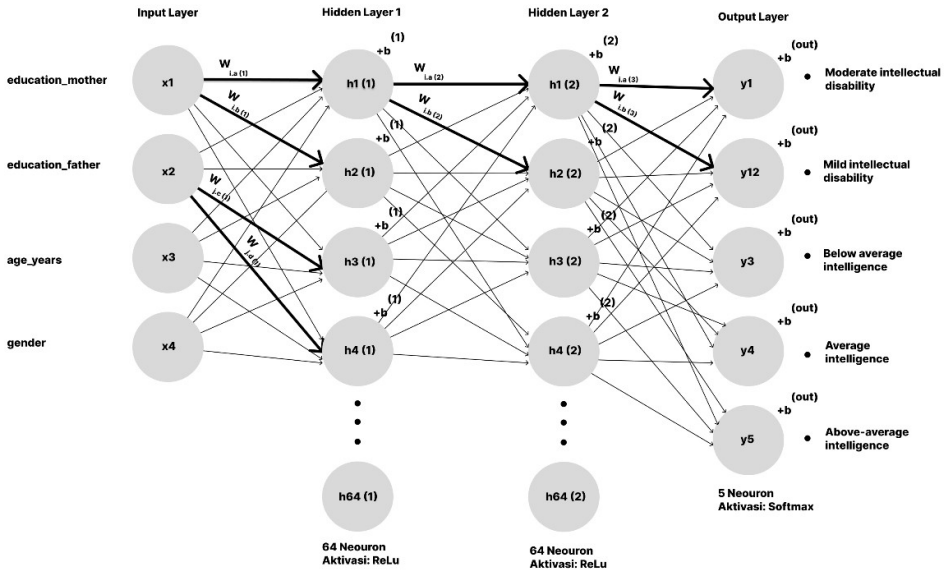
lainnya digunakan sebagai data pelatihan. Setiap data akan digunakan secara bergantian sebagai data latih dan data validasi, sehingga menghasilkan performa yang lebih optimal terhadap kemampuan generalisasi model (Abedin et al., 2026).

Pada setiap iterasi, dilakukan *feature scaling* menggunakan metode *standardization (Z-score)* untuk menyamakan skala antar fitur. Teknik ini penting karena perbedaan skala dapat memengaruhi kinerja model, terutama pada algoritma seperti *Multilayer Perceptron (MLP)* yang sensitif terhadap distribusi data. Selanjutnya, diterapkan metode *Synthetic Minority Over-sampling Technique (SMOTE)* pada data pelatihan. Metode ini menghasilkan data sintetis pada kelas minoritas melalui proses interpolasi antar sampel yang berdekatan, sehingga dapat memperbanyak distribusi data dan meningkatkan kemampuan model dalam mengenali pola pada kelas minoritas.

Proses *scaling* dan *oversampling* dilakukan secara terpisah pada setiap iterasi *k-fold* dan hanya diterapkan pada data pelatihan untuk menghindari *data leakage*. Dengan tahapan tersebut, model diharapkan mampu belajar dari data yang telah diseimbangkan dan distandarisasi secara optimal, sehingga menghasilkan performa yang lebih stabil dan dapat digeneralisasikan dengan baik. Hasil dari tahapan ini berupa nilai performa model pada setiap iterasi *k-fold cross-validation*, seperti *accuracy*, *precision*, *recall*, dan *F1-score*. Nilai-nilai tersebut kemudian dirata-ratakan untuk memperoleh estimasi performa akhir yang lebih stabil dan representatif terhadap kemampuan generalisasi model.

3.1.3 Arsitektur Model *Multilayer* Perceptron (MLP)

Model *Multilayer Perceptron* (MLP) yang digunakan dalam penelitian ini bekerja berdasarkan prinsip feedforward neural network dengan pembelajaran terarah. Struktur jaringan MLP terdiri dari tiga bagian utama, yaitu input layer, hidden layer, dan output layer (Madhiarasan & Louzazni, 2022). Dalam konfigurasi ini, setiap neuron pada suatu lapisan terhubung penuh dengan neuron pada lapisan berikutnya sehingga aliran informasi berlangsung searah dari input menuju output. Mekanisme komputasi terjadi melalui kombinasi bobot, bias, dan fungsi aktivasi nonlinier yang memungkinkan model merepresentasikan keterkaitan antarvariabel secara kompleks (Mfetoum et al., 2024).



Gambar 3. 4 Arsitektur model *multilayer perceptron*

Sumber: (Setiawan et al., n.d., p. 2025)

Gambar 3.4 menunjukkan arsitektur *MultiLayer Perceptron* MLP dengan penjelasan tiap - tiap layer sebagai berikut:

1. *Input layer*, terdiri dari 4 *neuron* yang masing-masing merepresentasikan variabel prediktor pada dataset. Setiap neuron menerima satu nilai fitur dari hasil prapemrosesan dan *encoding*, yaitu, *education_mother, education_father, age_years, gender*. Setiap *neuron* pada *input layer* berfungsi sebagai penerus sinyal yang langsung diteruskan ke *hidden layer* pertama sebagai vektor *input* jaringan. Sebagai contoh, satu data anak setelah proses *encoding* dapat direpresentasikan sebagai berikut:

$$x = [3,\ 2,\ 10,\ 1]$$

2. *Hidden layer* menerima masukan dari *input layer* dan mengolahnya melalui operasi linear berupa penjumlahan berbobot (*weighted sum*) yang disertai penambahan bias. Hasil perhitungan tersebut kemudian ditransformasikan menggunakan fungsi aktivasi nonlinier sehingga jaringan mampu merepresentasikan pola keterkaitan antara variabel pendidikan orang tua dan kategori IQ anak.

a. Pada *hidden layer* pertama, setiap neuron menerima seluruh nilai masukan dari empat fitur pada input layer, yaitu pendidikan ibu, pendidikan ayah, usia, dan jenis kelamin. Di mana setiap nilai input x_i dikalikan dengan bobot koneksinya $w_{ij}^{(1)}$, kemudian seluruh hasilnya dijumlahkan dan ditambahkan dengan bias $b_j^{(1)}$untuk menghasilkan nilai net input $z_j^{(1)}$. Secara matematis proses tersebut dirumuskan sebagai berikut:

$$z_j = \sum_i W_{ij} x_i + b_j \tag{3.1}$$

Keterangan:
z_j: total input ke neuron ke-j
W_{ij}: bobot koneksi dari neuron ke-i ke neuron ke-j
x_i: keluaran neuron ke-i dari lapisan sebelumnya
b_j: bias neuron ke-j

Nilai net input selanjutnya diproses menggunakan fungsi aktivasi *Rectified Linear Unit* (ReLU). Fungsi aktivasi ini hanya meneruskan nilai positif dan mengubah nilai negatif menjadi nol, sehingga hanya neuron yang merepresentasikan kombinasi fitur yang relevan yang akan aktif. Contoh sederhana perhitungan 2 neuron dan 4 fitur input, dengan nilai bobot dan bias sebagai berikut:

$$x = [x_1, x_2, x_3, x_4] = [3,2,10,1]$$

Bobot dan bias yang digunakan pada neuron pertama sebagai berikut:

$$w_1^{(1)} = [0.2, -0.4, 0.1, 0.6] \text{ dan } \quad b_1^{(1)} = -0.5$$

Maka perhitungan neuron pertama,

$$z_1^{(1)} = (0.2 \cdot 3) + (-0.4 \cdot 2) + (0.1 \cdot 10) + (0.6 \cdot 1) - 0.5$$

$$z_1^{(1)} = 0.6 - 0.8 + 1 + 0.6 - 0.5$$

$$z_1^{(1)} = 0.9$$

Kemudian proses aktivasi ReLU,

$$a_1^{(1)} = \max\left(0, z_1^{(1)}\right) \tag{3.2}$$

$$a_1^{(1)} = \max(0,0.9)$$

$$a_1^{(1)} = 0.9$$

Perhitungan neuron kedua, dengan bobot dan bias:

$$w_2^{(1)} = [-0.3, 0.2, -0.05, 0.1] \; dan \quad b_2^{(1)} = -0.2$$

$$z_2^{(1)} = (-0.3 \cdot 3) + (0.2 \cdot 2) + (-0.05 \cdot 10) + (0.1 \cdot 1) - 0.2$$

$$z_2^{(1)} = -0.9 + 0.4 - 0.5 + 0.1 - 0.2$$

$$z_2^{(1)} = -1.1$$

Hasil aktifasi ReLU dari neuron kedua,

$$a_2^{(1)} = \max(0, -1.1)$$

Sehingga *output* dari *hidden layer* 1 dengan 2 *neuron* 4 fitur,

$$a^{(1)} = [0.9,0]$$

b. Pada *hidden layer* kedua, menerima nilai masukan berupa hasil keluaran aktivasi dari *hidden layer* pertama dalam bentuk representasi fitur yang telah diproses oleh fungsi aktivasi ReLU. Pada tahap ini setiap neuron kembali melakukan operasi linier berupa penjumlahan berbobot terhadap keluaran hidden layer pertama $a^{(1)}$. Setiap nilai aktivasi $a_j^{(1)}$dikalikan dengan bobot koneksi antar layer $w_{jk}^{(2)}$, kemudian seluruh hasilnya dijumlahkan dan ditambahkan dengan bias $b_k^{(2)}$untuk menghasilkan nilai *net input* $z_k^{(2)}$. Dengan mekanisme ini, *hidden layer* kedua membentuk representasi yang lebih abstrak dari data, sehingga jaringan mampu mengenali hubungan yang lebih kompleks antar variabel. Contoh sederhana perhitungan *hidden layer* kedua dengan 2 neuron:

Output dari *hidden layer* pertama, bobot dan bias yang digunakan,

$$a^{(1)} = [0.9,0]$$

$$w_1^{(2)} = [0.5, -0.7]\ dan \quad b_1^{(2)} = 0.2$$

Perhitungan neuron pertama,

$$z_1^{(2)} = (0.5 \cdot 0.9) + (-0.7 \cdot 0) + 0.2$$

$$z_1^{(2)} = 0.45 + 0 + 0.2$$

$$z_1^{(2)} = 0.65$$

Fungsi aktivasi ReLU

$$a_1^{(2)} = \max(0,0.65) = 0.65$$

Perhitungan neuron kedua,

$$w_2^{(2)} = [-0.4,0.3]\ dan \quad b_2^{(2)} = -0.1$$

$$z_2^{(2)} = (-0.4 \cdot 0.9) + (0.3 \cdot 0) - 0.1$$

$$z_2^{(2)} = -0.36 + 0 - 0.1$$

$$z_2^{(2)} = -0.46$$

Fungsi aktivasi ReLU

$$a_2^{(2)} = \max(0, -0.46)$$

$$a_2^{(2)} = 0$$

Sehingga keluaran dari *hidden layer* kedua adalah,

$$a^{(2)} = [0.65, 0]$$

c. *Output layer* merupakan lapisan terakhir pada jaringan Multilayer Perceptron yang menghasilkan prediksi akhir berupa kategori IQ anak. Pada tahap ini setiap neuron pada *output layer* menerima seluruh nilai keluaran dari hidden layer kedua, kemudian menghitung kombinasi linier antara bobot dan bias. Jumlah *neuron* pada *output layer* ditentukan oleh jumlah kelas target. Pada penelitian ini terdapat lima kategori IQ, yaitu *Moderate intellectual disability, Mild intellectual disability, Below average intelligence, Average intelligence, dan Above-average intelligence,*

sehingga *output layer* memiliki K = 5 *neuron*. Nilai masukan pada *neuron* ke-k dihitung dengan:

$$z_k^{(3)} = \sum_{j=1}^{n} w_{jk}^{(3)} a_j^{(2)} + b_k^{(out)} \quad (3.3)$$

Keterangan:
$a_j^{(2)}$ = keluaran neuron ke-j dari hidden layer kedua
$w_{jk}^{(3)}$= bobot dari neuron hidden ke neuron output
$b_k^{(out)}$= bias neuron output

Kemudian nilai tersebut diproses dengan fungsi aktivasi *softmax,* yang digunakan untuk memperoleh nilai probabilitas pada masing-masing kelas sebagai berikut:

$$\widehat{y_k} = \frac{e^{z_k^{(3)}}}{\sum_{i=1}^{K} e^{z_i^{(3)}}} \quad (3.4)$$

$\hat{y}_k$: probabilitas prediksi untuk kelas ke-k
$z_k^{(3)}$: nilai (hasil kombinasi linier bobot dan bias) pada neuron output ke-k
e: bilangan eksponensial (≈ 2,71828)
K: jumlah total kelas (pada penelitian ini $K = 5$)
$\sum_{i=1}^{K} e^{z_i^{(3)}}$: total seluruh nilai eksponensial dari semua neuron output

Kemudian hasil keluaran pada fungsi aktivasi *softmax* dipilih dengan nilai probabilitas paling besar sebagai klasifikasi akhir:

$$\text{Prediksi}=\max(\widehat{y_k})$$

Dengan demikian output layer berfungsi mengubah representasi nilai fitur abstrak dari *hidden layer* menjadi keputusan kategorikal yang menunjukkan kategori IQ anak. Untuk perhitungan manual dari output layer sebagai berikut:

Menghitung nilai (hasil kombinasi linier bobot dan bias) dari masing – masing neuron,

Neuron 1

$$z_1 = (0.8 \cdot 0.65) + (-0.2 \cdot 0) + 0.1$$

$$z_1 = 0.52 + 0 + 0.1$$

$$z_1 = 0.62$$

Neuron 2

$$z_2 = (-0.4 \cdot 0.65) + (0.6 \cdot 0) + 0.2$$

$$z_2 = -0.26 + 0 + 0.2$$

$$z_2 = -0.06$$

Neuron 3

$$z_3 = (0.3 \cdot 0.65) + (0.1 \cdot 0) - 0.3$$

$$z_3 = 0.195 - 0.3$$

$$z_3 = -0.105$$

Neuron 4

$$z_4 = (-0.7 \cdot 0.65) + (0.4 \cdot 0) + 0.05$$

$$z_4 = -0.455 + 0.05$$

$$z_4 = -0.405$$

Neuron 5

$$z_5 = (0.2 \cdot 0.65) + (-0.5 \cdot 0) + 0$$

$$z_5 = 0.13$$

Sehingga hasil akhirnya seperti ini,

$$z = [0.62, -0.06, -0.105, -0.405, 0.13]$$

Fungsi aktivasi *softmax*,

$$e^{0.62} \approx 1.859$$

$$e^{-0.06} \approx 0.942$$

$$e^{-0.105} \approx 0.900$$

$$e^{-0.405} \approx 0.667$$

$$e^{0.13} \approx 1.139$$

$$\text{Total} = \sum e^{z} = 1.859 + 0.942 + 0.900 + 0.667 + 1.139 = 5.507$$

Sehingga hasil *output layer* dari perhitungan 2 *neuron* 4 fitur,

$$P_1 = \frac{1.859}{5.507} = 0.338$$

$$P_2 = \frac{0.942}{5.507} = 0.171$$

$$P_3 = \frac{0.900}{5.507} = 0.163$$

$$P_4 = \frac{0.667}{5.507} = 0.121$$

$$P_5 = \frac{1.139}{5.507} = 0.207$$

Berdasarkan hasil perhitungan dari fungsi aktivasi softmax, diperoleh nilai probabilitas masing-masing kelas sebesar $P_1 = 0.338$, $P_2 = 0.171$, $P_3 = 0.163$, $P_4 = 0.121$, dan $P_5 = 0.207$. Karena nilai probabilitas terbesar terdapat pada kelas ke-1 (0.338), maka model memprediksi data tersebut sebagai kelas ke-1, yaitu Moderate intellectual disability. Namun, prediksi tersebut belum tentu sesuai dengan label sebenarnya pada data. Diperlukan perhitungan (loss function) untuk mengukur seberapa besar kesalahan prediksi model terhadap label sebenarnya.

3.1.4 Loss Function

Setelah proses *feedforward* menghasilkan nilai aktivasi akhir tiap kelas melalui fungsi *softmax* pada output layer, langkah berikutnya adalah menghitung kesalahan prediksi (*error*) dengan membandingkan hasil prediksi dengan label sebenarnya. Fungsi yang digunakan dalam *loss function* adalah *categorical cross-entropy*.

$$L = -\sum_{k=1}^{K} y_k \log(\widehat{y_k}) \tag{3.5}$$

Keterangan:
$\hat{y}_k$= probabilitas hasil prediksi untuk kelas ke-k (dari softmax)
y_k= label sebenarnya dalam bentuk *one-hot encoding*
K= jumlah kelas (5 kategori IQ)

Karena label hanya memiliki satu komponen bernilai 1 dan lainnya 0, sehingga persamaan *loss fungtion* dapat disederhanakan menjadi:

$$L = -\log(\widehat{y_{\text{true}}}) \tag{3.6}$$

Keterangan:
y_{true} = probabilitas $\widehat{\text{pada}}$ kelas yang benar.

Perhitungan sederhana untuk *loss function* sebagai berikut:

Nilai hasil *output layer*,

$$\hat{y} = (0.338, 0.171, 0.163, 0.121, 0.207)$$

Berdasarkan nilai probabilitas terbesar, model memprediksi kelas ke-1. Namun label sebenarnya pada data adalah kelas ke-4, sehingga dihitung *loss* menggunakan probabilitas pada kelas ke-4 (*Average intelligence*), maka vektor target:

$$y = (0,0,0,1,0)$$

Sehingga hasil loss function,

$$L = -\log(\widehat{y_4})$$

$$L = -\log(0.121)$$

$$L \approx 2.11$$

Nilai *loss* yang diperoleh sebesar 2.11, yang menunjukkan bahwa nilai probabilitas yang diberikan model terhadap kelas yang benar masih rendah. Kinerja model dianggap semakin baik ketika nilai *loss* mendekali nol, dikarenakan hasil nilai *loss function* semakin mendekati satu, menandakan model belum mampu mengenali pola data dengan baik sehingga bobot jaringan perlu diperbarui melalui proses backpropagation.

3.1.5 Backpropagation

Setelah nilai *loss* diperoleh dari fungsi *cross-entropy*, jaringan melakukan proses *backpropagation* untuk menyesuaikan bobot agar kesalahan prediksi semakin kecil pada proses berikutnya. Proses ini dilakukan dengan menyebarkan kesalahan prediksi (*error*) dari lapisan *output* ke lapisan sebelumnya secara bertahap hingga mencapai hidden layer pertama untuk mengetahui nilai kontribusi masing-masing neuron terhadap kesalahan model, yang dapat dihitung dengan,

$$\delta_k^{(out)} = \widehat{y_k} - y_k \tag{3.7}$$

Keterangan:
$\delta_k^{(3)}$= error neuron output ke-k
$\hat{y}_k$= probabilitas hasil softmax
y_k= label sebenarnya (*one-hot*)

Perhitungan hasil *error output layer,*

$$\delta^{(3)} = [0.338, 0.171, 0.163, 0.121 - 1, 0.207]$$

$$\delta^{(3)} = [0.338, 0.171, 0.163, -0.879, 0.207]$$

Kemudian gradien bobot dihitung dari hasil *error output* dengan aktivasi *neuron* pada hidden layer kedua:

$$\frac{\partial L}{\partial w_{j,k}} = \delta_k^{(out)} \cdot a_j^{(2)}$$

Hasil dari *gradient* bobot antara neuron pertama *hidden layer* kedua dan neuron output ke-4,

$$\frac{\partial L}{\partial w_{1,4}} = (-0.879) \cdot (0.65) = -0.571$$

Berdasarkan hasil nilai *gradient* bobot pada setiap neuron tersebut menunjukkan tingkat kontribusi neuron terhadap kesalahan yang besar pada prediksi model dan menjadi dasar dalam perbaikan pada tahap selanjutnya.

3.1.6 Pembaruan Bobot dan Bias

Setelah *gradien error* diperoleh melalui proses *backpropagation*, tahap berikutnya adalah memperbarui parameter jaringan, yaitu bobot dan bias. Tujuan pembaruan ini adalah mengurangi nilai *loss* sehingga prediksi model semakin mendekati label sebenarnya pada proses selanjutnya. Pembaruan parameter jaringan tidak menggunakan gradient descent sederhana, melainkan *optimizer Adam* (*Adaptive Moment Estimation*). Yang dimana setiap bobot diperbarui dengan langkah yang berbeda sesuai karakteristik gradiennya masing-maisng. Pada Adam Optimizer proses pembaruan parameter dilakukan melalui empat tahapan, yaitu

perhitungan gradien, estimasi momen pertama (*first moment*), estimasi momen kedua (*second moment*), koreksi bias (*bias correction*), dan pembaruan parameter. Gradien terhadap suatu bobot pada iterasi ke-t:

$$g_t = \frac{\partial L}{\partial w_t} \tag{3.8}$$

Estimasi Rata-rata Gradien (*Momentum* Pertama),

$$m_t = \beta_1 m_{t-1} + (1 - \beta_1) g_t \tag{3.9}$$

Estimasi Variansi Gradien (Momentum Kedua),

$$v_t = \beta_2 v_{t-1} + (1 - \beta_2) g_t^2 \tag{3.10}$$

Koreksi Bias,

$$\hat{m}_t = \frac{m_t}{1 - \beta_1^t} \tag{3.11}$$

$$\hat{v}_t = \frac{v_t}{1 - \beta_2^t} \tag{3.12}$$

Pembaruan Bobot

$$w_{t+1} = w_t - \eta \frac{\hat{m}_t}{\sqrt{\hat{v}_t} + \varepsilon} \tag{3.13}$$

Pembaruan Bias

$$b_{t+1} = b_t - \eta \frac{\hat{m}_t^{(b)}}{\sqrt{\hat{v}_t^{(b)}} + \varepsilon} \tag{3.14}$$

Keterangan:
w_t: bobot pada iterasi ke-t
b_t: bias pada iterasi ke-t
g_t: gradien error
η: learning rate
β_1: koefisien momentum gradien (≈ 0.9)
β_2: koefisien momentum kuadrat gradien (≈ 0.999)
ε: konstanta kecil untuk mencegah pembagian nol

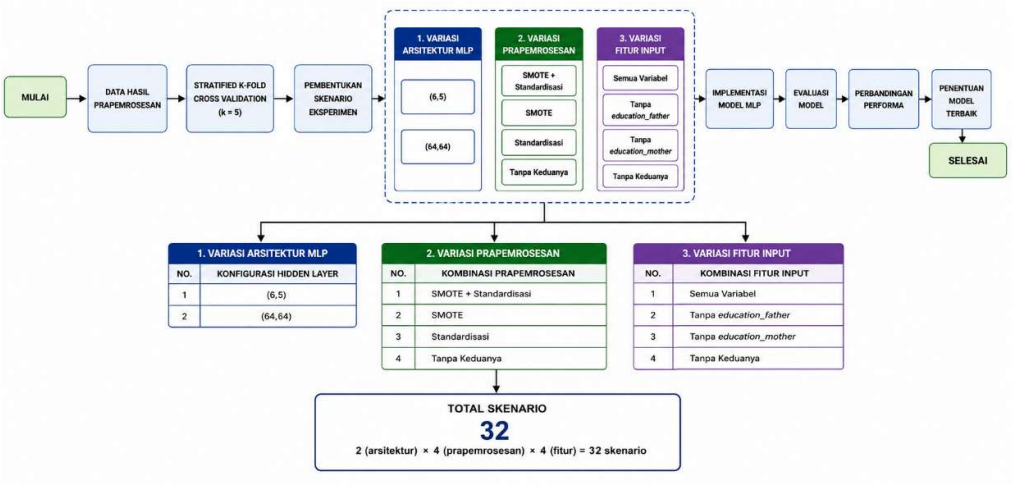
Dengan dilakukannya pembaruan untuk seluruh bobot dan bias pada setiap neuron, diharapkan dapat mengoptimalkan parameter untuk meminimalisir nilai *loss* selama proses pelatihan berlangsung. Selain itu, penelitian ini menerapkan mekanisme *Early Stopping* untuk mengurangi risiko *overfitting* pada proses pelatihan jaringan saraf. Menurut (Prechelt, 1998), *Early Stopping* memanfaatkan data validasi untuk mendeteksi saat performa model mulai mengalami penurunan akibat *overfitting*, sehingga proses pelatihan dapat dihentikan sebelum kemampuan generalisasi model menurun.

3.1.7 Skenario Eksperimen

Skenario eksperimen pada penelitian ini dirancang untuk mengevaluasi performa model *Multilayer Perceptron* (MLP) dalam mengklasifikasikan kategori kemampuan intelektual anak berdasarkan tingkat pendidikan orang tua. Pengujian dilakukan dengan mengeksplorasi beberapa komponen yang berpotensi memengaruhi performa model, yaitu arsitektur jaringan, teknik prapemrosesan data, dan kombinasi fitur yang digunakan sebagai masukan model. Proses eksperimen diawali dengan data hasil prapemrosesan yang selanjutnya divalidasi menggunakan metode *Stratified K-Fold Cross Validation* dengan nilai k = 5. Pada setiap iterasi, data dibagi menjadi data pelatihan dan data validasi dengan

mempertahankan proporsi kelas pada setiap *fold*. Kombinasi seluruh variasi tersebut menghasilkan tiga puluh dua (32) skenario eksperimen yang dianalisis pengaruhnya terhadap performa klasifikasi.

Variasi pertama dilakukan pada arsitektur model MLP dengan menggunakan dua konfigurasi *hidden layer*, yaitu (6, 5) dan (64, 64). Variasi kedua dilakukan pada teknik prapemrosesan data, yang terdiri atas empat kombinasi, yaitu: (1) penerapan SMOTE dan standardisasi secara bersamaan, (2) penerapan SMOTE saja tanpa standardisasi, (3) penerapan standardisasi saja tanpa SMOTE, dan (4) tanpa penerapan keduanya. Variasi ketiga dilakukan pada fitur input yang digunakan dalam model, yaitu: seluruh variabel input (*education_father*, *education_mother*, *age_years*, dan *gender*), tanpa variabel *education_father* (hanya *education_mother*, *age_years*, dan *gender*), tanpa variabel *education_mother* (hanya *education_father*, *age_years*, dan *gender*), serta tanpa keduanya (hanya *age_years* dan *gender*). Setiap kombinasi skenario selanjutnya digunakan untuk melatih dan menguji model MLP. Hasil prediksi dari setiap skenario dievaluasi menggunakan metrik *Accuracy*, *Precision*, *Recall*, *Macro F1-Score*, dan AUC-ROC. Seluruh hasil evaluasi kemudian dibandingkan untuk menentukan konfigurasi model yang memberikan performa terbaik. Alur skenario eksperimen yang digunakan dalam penelitian ini ditunjukkan pada Gambar 3.5.



Gambar 3. 5 Skenario pengujian

Berdasarkan alur pada Gambar 3.5, kombinasi variasi arsitektur, teknik prapemrosesan, dan fitur input menghasilkan beberapa skenario pengujian yang digunakan untuk mengevaluasi performa model menggunakan metrik Accuracy, F1-Score, dan AUC-ROC. Pada penelitian ini terdapat tiga puluh dua skenario uji coba yang ditunjukkan pada Tabel 3.6 berikut.

Tabel 3. 6 Skenario uji coba

Model	**Arsitektur *hidden layer***	***Smote***	***Standarisasi***	***Father's education***	***Mother's educaction***
1	(6, 5)	✓	✓	✓	✓
2		✓	X	✓	✓
3		X	✓	✓	✓
4		X	X	✓	✓
5		✓	✓	X	✓
6		✓	X	X	✓
7		X	✓	X	✓
8		X	X	X	✓
9		✓	✓	✓	X
10		✓	X	✓	X
11		X	✓	✓	X
12		X	X	✓	X
13		✓	✓	X	X
14		✓	X	X	X
15		X	✓	X	X
16		X	X	X	X

Model	Arsitektur *hidden layer*	*Smote*	*Standarisasi*	*Father's education*	*Mother's educaction*
17	(64, 64)	✓	✓	✓	✓
18		✓	✗	✓	✓
19		✗	✓	✓	✓
20		✗	✗	✓	✓
21		✓	✓	✗	✓
22		✓	✗	✗	✓
23		✗	✓	✗	✓
24		✗	✗	✗	✓
25		✓	✓	✓	✗
26		✓	✗	✓	✗
27		✗	✓	✓	✗
28		✗	✗	✓	✗
29		✓	✓	✗	✗
30		✓	✗	✗	✗
31		✗	✓	✗	✗
32		✗	✗	✗	✗

3.2 Implementasi

Tahap Implementasi merupakan tahap penerapan model *Multilayer Perceptron (MLP)* yang telah di *buid* ke dalam sebuah GUI yang dapat digunakan oleh *user*. Pada penelitian ini, sistem dikembangkan menggunakan framework *Streamlit* berbasis Python yang berfungsi sebagai wadah antara pengguna dan model prediksi kategori IQ. Melalui GUI ini, pengguna dapat melakukan prediksi kategori IQ baik secara individu maupun secara massal menggunakan dataset.

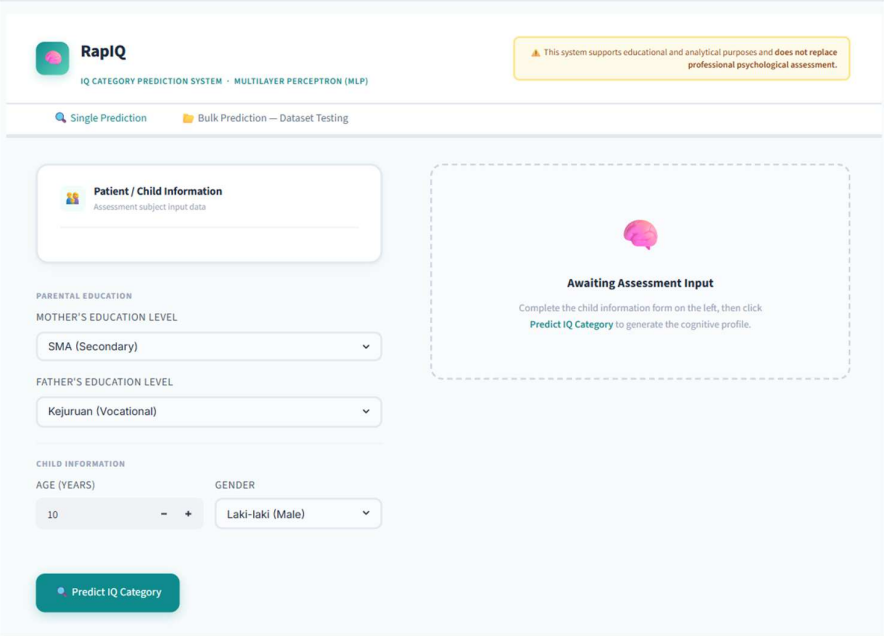
Sistem terdiri atas dua halaman utama, yaitu Single Prediction dan Bulk Prediction. Halaman Single Prediction digunakan untuk melakukan prediksi berdasarkan data individu, sedangkan halaman Bulk Prediction digunakan untuk melakukan prediksi terhadap banyak data sekaligus dalam format CSV. Selain itu, sistem juga mengintegrasikan model MLP dan proses standardisasi data sehingga hasil prediksi dapat ditampilkan secara langsung kepada pengguna.

3.2.1 Implementasi User Interface (UI)

User Interface pada penelitian ini dibuat menggunakan Streamlit, yaitu framework Python *open-source* yang memungkinkan pembuatan aplikasi web interaktif secara cepat tanpa memerlukan keahlian pengembangan web khusus. Streamlit dipilih karena kompatibilitasnya yang tinggi dibidang *data science* Python, kemampuannya dalam menampilkan hasil prediksi model machine learning secara langsung, serta kemudahan integrasi dengan pustaka ilmiah seperti NumPy, Pandas, dan scikit-learn.

a. Halaman *Single Prediction*

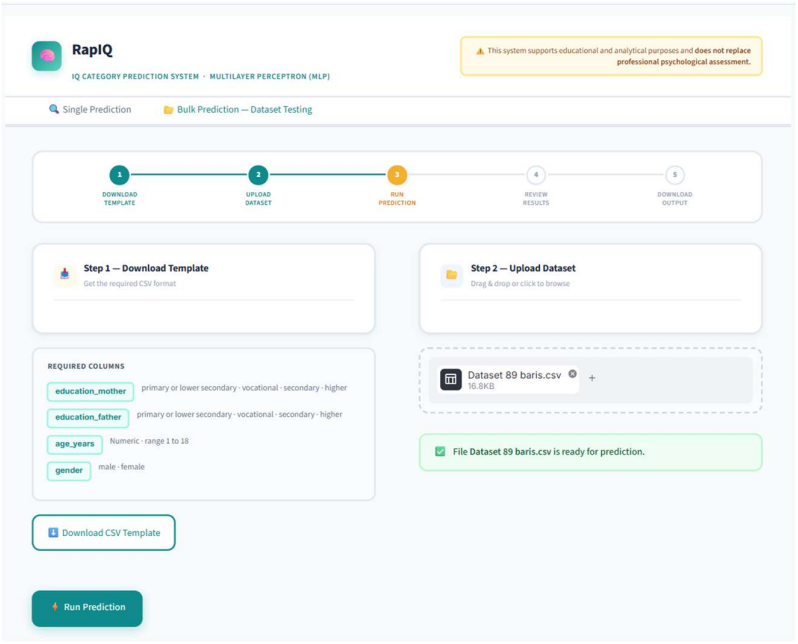
Halaman *Single Prediction* digunakan untuk melakukan prediksi kategori IQ berdasarkan data individu. Pengguna diminta memasukkan tingkat pendidikan ibu, tingkat pendidikan ayah, usia anak, dan jenis kelamin sebagai variabel masukan model. Setelah tombol prediksi ditekan, data akan dikirimkan ke model MLP untuk dilakukan proses prediksi dan menghasilkan kategori IQ yang diperkirakan, seperti pada Gambar 3.6.



Gambar 3. 6 Contoh halaman *single prediction*

b. Halaman *Bulk Prediction*

Halaman *Bulk Prediction* dirancang untuk memproses prediksi secara massal dengan cara mengunggah file dataset dalam format CSV. Halaman ini memungkinkan pengguna untuk melakukan prediksi terhadap banyak data anak sekaligus, sehingga lebih efisien digunakan dalam konteks asesmen skala besar. Sistem akan membaca dataset yang diunggah, melakukan prapemrosesan secara otomatis, menjalankan model prediksi, dan menampilkan hasil prediksi untuk setiap baris data seperti Gambar 3.7.



Gambar 3. 7 Contoh halaman *bulk prediction*

Implementasi komponen input pada halaman *Single Prediction* menggunakan elemen Streamlit yang sesuai dengan tipe data masing-masing variabel. Variabel tingkat pendidikan orang tua diimplementasikan menggunakan komponen selectbox karena memiliki nilai kategorikal yang terbatas, sedangkan variabel usia diimplementasikan menggunakan komponen number_input dengan pembatasan rentang nilai yang valid. Potongan kode pada gambar 3.8 menunjukkan implementasi komponen utama pada halaman *Single Prediction*.

```
st.markdown('<div class="rq-section-label">Parental Education</div>', unsafe_allow_html=True)
edu_mother_lbl = st.selectbox("Mother's Education Level",  list(EDU_MAP_UI), key="s_edu_m")
edu_father_lbl = st.selectbox("Father's Education Level",  list(EDU_MAP_UI), key="s_edu_f")

st.markdown('<div class="rq-hr"></div><div class="rq-section-label">Child Information</div>',
unsafe_allow_html=True)
c3, c4 = st.columns(2)
with c3:
    age = st.number_input("Age (years)", min_value=1, max_value=18, value=10, step=1, key="s_age")
with c4:
    gender_lbl = st.selectbox("Gender", list(GENDER_MAP_UI), key="s_gender")

st.markdown("", unsafe_allow_html=True)
st.markdown("<br>", unsafe_allow_html=True)
predict_clicked = st.button("🔍  Predict IQ Category", key="btn_single")
```

Gambar 3. 8 *Sourcecode input* data pada halaman *single prediction*

Kode di atas menunjukkan penggunaan st.selectbox() untuk memilih tingkat pendidikan ibu dari daftar opsi yang telah dipetakan melalui kamus EDU_MAP_UI, serta penggunaan st.number_input() untuk memasukkan usia anak dalam satuan tahun dengan pembatasan nilai minimum 1 dan maksimum 18. Prinsip yang sama diterapkan untuk input variabel pendidikan ayah dan jenis kelamin anak.

Pada halaman Bulk Prediction, komponen utama yang digunakan adalah file_uploader, yang memungkinkan pengguna mengunggah file CSV langsung dari UI web. Pada gambar 3.9 menunjukkan implementasi komponen unggah dataset.

```
uploaded = st.file_uploader(
    "Upload CSV dataset",
    type=["csv"],
    key="bulk_upload",
    label_visibility="collapsed",
)
```

Gambar 3. 9 *Sourcecode input* data pada halaman *bulk prediction*

Komponen st.file_uploader() pada kode di atas dikonfigurasi untuk hanya menerima file dengan ekstensi .csv. Ketika pengguna mengunggah file, Streamlit secara otomatis menyediakan objek file yang dapat diproses menggunakan pustaka Pandas untuk pembacaan dan prapemrosesan data lebih lanjut sebelum diteruskan ke model prediksi.

Tabel 3. 7 Komponen utama ui streamlit

Komponen UI	**Tipe Komponen**	**Fungsi**
Pilihan tingkat pendidikan ibu	*st.selectbox()*	Memilih kategori pendidikan ibu dari daftar opsi (primary/vocational/secondary/higher)
Pilihan tingkat pendidikan ayah	*st.selectbox()*	Memilih kategori pendidikan ayah dari daftar opsi yang sama

Komponen UI	Tipe Komponen	Fungsi
Input usia anak	*st.number_input()*	Memasukkan nilai numerik usia anak (1-18 tahun)
Pilihan jenis kelamin	*st.selectbox()*	Memilih jenis kelamin anak (Laki-laki/Perempuan)
Unggah dataset CSV	*st.file_uploader()*	Mengunggah file CSV berisi data banyak anak untuk prediksi massal

3.2.2 Implementasi Model *Multilayer Perceptron*

Subbab ini menjelaskan cara model *Multilayer Perceptron* (MLP) dari hasil penelitian diintegrasikan ke dalam aplikasi Streamlit. Model yang telah dilatih pada tahap eksperimen disimpan dalam format file .pkl menggunakan pustaka joblib, sehingga dapat dimuat kembali tanpa perlu melatih ulang model setiap kali aplikasi dijalankan. Pendekatan ini memastikan efisiensi komputasi sekaligus konsistensi hasil prediksi.

a. Pemanggilan Model

Model MLP dan objek scaler yang telah disimpan dimuat ke dalam aplikasi menggunakan fungsi joblib.load(). Pemanggilan dilakukan satu kali saat aplikasi diinisialisasi agar tidak terjadi pembacaan berulang yang dapat memperlambat performa sistem.

```
model, scaler, errors = load_artifacts(
    MODEL_PATH,
    SCALER_PATH
)

def load_artifacts(model_path, scaler_path):
    model = joblib.load(model_path)
    scaler = joblib.load(scaler_path)
    return model, scaler
```

Gambar 3. 10 Kode program pemanggilan model

Variabel MODEL_PATH dan SCALER_PATH mendefinisikan lokasi file model dan scaler yang tersimpan di direktori aplikasi. Fungsi joblib.load() memuat objek Python yang telah diserialisasi dari file .pkl ke dalam memori, sehingga model dan scaler siap digunakan untuk proses prediksi tanpa perlu melakukan pelatihan ulang.

b. Implementasi Prediksi

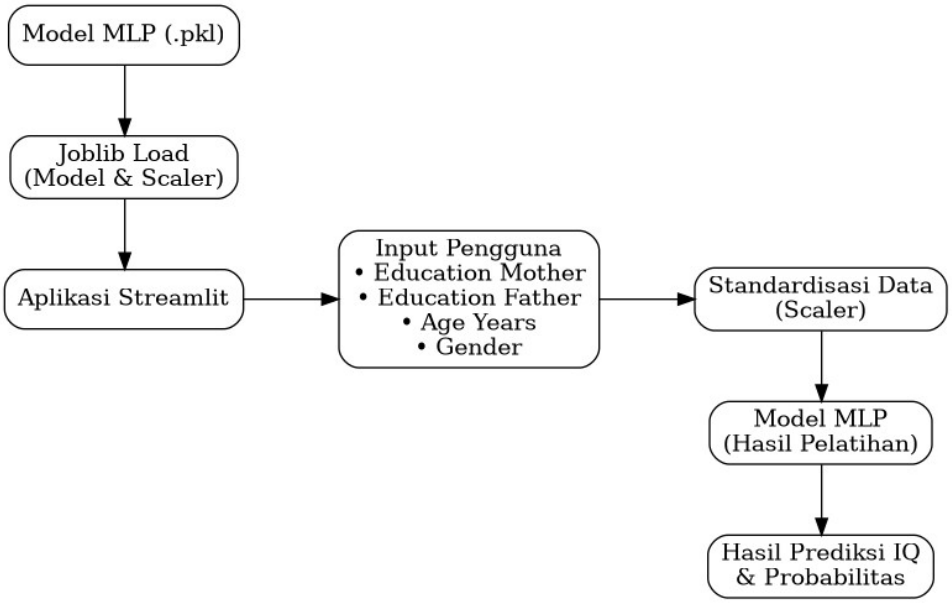
Setelah model dan scaler berhasil dimuat, proses prediksi dilakukan dengan tiga langkah utama, yaitu standardisasi data input menggunakan scaler, pengambilan kelas prediksi dari model, dan pengambilan nilai probabilitas untuk setiap kelas. Implementasi prediksi ditunjukkan pada kode berikut.

```
X_scaled   = scaler.transform(X)
pred_class = int(model.predict(X_scaled)[0])
proba      = model.predict_proba(X_scaled)[0]
```

Gambar 3. 11 Kode program prediksi menggunakan mlp

Baris pertama melakukan transformasi data input X menggunakan scaler yang sama dengan yang digunakan saat pelatihan, sehingga distribusi nilai fitur konsisten antara fase pelatihan dan inferensi. Baris kedua menghasilkan kelas prediksi berupa label kategori IQ dengan probabilitas tertinggi, sedangkan baris ketiga menghasilkan array probabilitas untuk setiap kelas yang dapat ditampilkan kepada pengguna sebagai informasi tambahan mengenai tingkat keyakinan model.

Gambar 3.12 mengilustrasikan alur integrasi model MLP dengan aplikasi Streamlit, mulai dari pemuatan model hingga pengiriman hasil prediksi ke antarmuka pengguna.



Gambar 3. 12 . Integrasi model mlp dengan *streamlit*

3.2.3 Alur Implementasi Sistem

Alur implementasi sistem menggambarkan rangkaian proses yang terjadi dari sisi pengguna hingga dihasilkannya prediksi kategori IQ anak. Secara keseluruhan, sistem bekerja melalui lima langkah utama yang saling berkaitan dan berlangsung secara berurutan seperti pada Gambar 3.13.

Gambar 3. 13 Alur implementasi sistem

Langkah pertama, pengguna memasukkan data secara manual melalui formulir pada halaman *Single Prediction* atau mengunggah dataset dalam format CSV melalui halaman *Bulk Prediction*. Langkah kedua, sistem melakukan validasi dan prapemrosesan data untuk memastikan seluruh kolom yang dibutuhkan tersedia, tidak terdapat nilai kosong, serta tipe data sesuai dengan kebutuhan model. Pada tahap ini, variabel kategorikal seperti tingkat pendidikan orang tua dan jenis kelamin dikonversi ke bentuk numerik melalui proses *encoding*. Langkah ketiga, data yang telah dipraproses distandardisasi menggunakan *scaler* yang disimpan bersama model agar distribusi nilai fitur tetap konsisten dengan data yang digunakan saat pelatihan. Langkah keempat, data yang telah distandardisasi diproses oleh model ANN berbasis *Multilayer Perceptron* (MLP) untuk menghasilkan prediksi kategori IQ beserta probabilitas masing-masing kelas. Langkah kelima, hasil prediksi ditampilkan melalui antarmuka Streamlit. Pada halaman *Single Prediction*, hasil ditampilkan dalam bentuk prediksi individu yang mudah dibaca, sedangkan pada halaman *Bulk Prediction*, hasil prediksi seluruh data ditampilkan dalam bentuk tabel dan dapat diunduh dalam format CSV untuk keperluan analisis lebih lanjut.

BAB IV

HASIL DAN PEMBAHASAN

4.1 Hasil Eksperimen

Hasil ini menyajikan hasil eksperimen yang dilakukan untuk mengevaluasi performa model *Multilayer Perceptron* (MLP) dalam mengklasifikasikan kategori kemampuan intelektual anak berdasarkan tingkat pendidikan orang tua. Secara keseluruhan, penelitian ini melibatkan tiga puluh dua (32) skenario eksperimen yang dibangun dari kombinasi dua variasi arsitektur *hidden layer*, dua pilihan teknik prapemrosesan (SMOTE dan standardisasi), serta empat variasi kombinasi variabel input pendidikan orang tua. Evaluasi performa dilakukan menggunakan metrik *accuracy*, *precision*, *recall*, *Macro F1-Score*, dan AUC-ROC, dengan metode validasi *Stratified 5-Fold Cross Validation*. Seluruh skenario eksperimen dikelompokkan ke dalam dua arsitektur utama, yaitu arsitektur *hidden layer* (6, 5) yang mencakup Model 1 hingga Model 16, dan arsitektur *hidden layer* (64, 64) yang mencakup Model 17 hingga Model 32. Pembahasan pada masing-masing kelompok arsitektur difokuskan pada dua aspek utama, yaitu pengaruh teknik prapemrosesan data (SMOTE dan standardisasi) serta pengaruh penggunaan variabel pendidikan orang tua terhadap performa klasifikasi model.

4.1.1 Arsitektur Hidden Layer (6, 5)

Kelompok pengujian pertama menggunakan arsitektur *Multilayer Perceptron* (MLP) dengan dua *hidden layer* masing-masing berisi 6 neuron dan 5 neuron, mencakup Model 1 hingga Model 16 sesuai dengan tabel 3.6. Pengujian ini

dirancang untuk mengevaluasi pengaruh teknik prapemrosesan data serta kontribusi variabel pendidikan orang tua terhadap performa klasifikasi pada arsitektur berkapasitas kecil tersebut. Berdasarkan keseluruhan evaluasi menggunakan metrik *accuracy*, *precision*, *recall*, *Macro F1-Score*, dan AUC-ROC, Model 1 ditetapkan sebagai model terbaik pada kelompok arsitektur ini. Model 1 merupakan konfigurasi dengan SMOTE, standardisasi, serta seluruh variabel input (*education_father*, *education_mother*, *age_years*, dan *gender*), menghasilkan *accuracy* sebesar 27,90%, *Macro F1-Score* sebesar 25,80%, *recall* sebesar 37,60%, dan nilai AUC-ROC pada rentang 0,63-0,78. Model ini menunjukkan performa yang paling seimbang dalam mengenali seluruh kategori kemampuan intelektual dibandingkan dengan model lainnya dalam kelompok arsitektur (6, 5).

4.1.1.1 Pengaruh *Smote* dan *Standarisasi*

Berdasarkan hasil evaluasi pada Model 1, Model 5, Model 9, dan Model 13 yang menerapkan kombinasi teknik SMOTE dan standardisasi, diperoleh rata-rata akurasi sebesar 27,26% serta rata-rata F1-Score sebesar 23,98%. Temuan ini menunjukkan bahwa penerapan kedua metode *preprocessing* tersebut menghasilkan tingkat akurasi yang masih relatif rendah apabila dibandingkan dengan model yang tidak menggunakan SMOTE. Namun demikian, nilai F1-Score yang dihasilkan cenderung lebih tinggi dibandingkan beberapa kombinasi *preprocessing* lainnya. Kondisi ini menunjukkan bahwa model memiliki kemampuan yang lebih baik dalam mengenali dan mengklasifikasikan setiap kelas secara lebih merata, sehingga performa klasifikasi antar kelas dapat ditingkatkan.

Tabel 4. 1 Hasil uji coba model dengan full preprocessing

Model	SMOTE	Standardisasi	Father's education	Mother's education	Akurasi	F1-Score	Waktu Training (detik)
1	✓	✓	✓	✓	27,90	25,80	14,72
5	✓	✓	✗	✓	25,33	24,13	31,58
9	**✓**	**✓**	**✓**	**✗**	**28,20**	**26,15**	**15,87**
13	✓	✓	✗	✗	27,62	17,87	61,35

Berdasarkan Tabel 4.1, Model 9 memperoleh kinerja terbaik dengan nilai akurasi sebesar 28,20% dan *F1-Score* sebesar 26,15%. Hasil analisis *confusion matrix* menunjukkan bahwa penerapan SMOTE yang dikombinasikan dengan standardisasi mampu mengurangi kecenderungan model dalam memprediksi kelas mayoritas secara dominan. Kondisi tersebut meningkatkan kemampuan model dalam mengidentifikasi kelas-kelas minoritas secara lebih merata. Meskipun demikian, peningkatan keseimbangan klasifikasi tersebut belum diikuti oleh peningkatan performa prediksi secara keseluruhan, yang ditunjukkan oleh nilai akurasi yang masih berada pada tingkat yang relatif rendah.

Kelompok model yang menerapkan SMOTE tanpa standardisasi terdiri atas Model 2, Model 6, Model 10, dan Model 14. Berdasarkan hasil pengujian, kelompok model tersebut menghasilkan rata-rata akurasi sebesar 28,56% dan rata-rata *F1-Score* sebesar 22,89%. Apabila dibandingkan dengan kelompok model yang menggunakan kombinasi SMOTE dan standardisasi, terdapat peningkatan kecil pada nilai akurasi, namun nilai *F1-Score* mengalami penurunan. Temuan ini mengindikasikan bahwa penghapusan proses standardisasi tidak memberikan peningkatan kinerja yang berarti ketika SMOTE diterapkan.

Tabel 4. 2 Hasil uji coba model tanpa standarisasi

Model	SMOTE	Standardisasi	Father's education	Mother's education	Akurasi	F1-Score	Waktu Training (detik)
2	✓	✗	✓	✓	27,14	25,42	15,52
6	✓	✗	✗	✓	25,13	24,00	270,45
10	✓	✗	✓	✗	28,06	23,23	14,57
14	**✓**	**✗**	**✗**	**✗**	**33,98**	**18,92**	**42,08**

Berdasarkan tabel 4.2, Model 14 mencatat nilai akurasi tertinggi sebesar 33,98%, sedangkan Model 2 memperoleh nilai *F1-Score* tertinggi sebesar 25,40%. Berdasarkan hasil analisis *confusion matrix* dan *classification report*, model dalam kelompok ini masih mengalami tingkat kesalahan klasifikasi yang cukup tinggi antar kelas, khususnya pada kelas yang memiliki karakteristik serupa. Oleh karena itu, meskipun terjadi peningkatan pada nilai akurasi, kemampuan klasifikasi secara keseluruhan masih tergolong rendah karena model belum mampu membedakan setiap kelas secara optimal. Kelompok model yang menerapkan standardisasi tanpa SMOTE terdiri atas Model 3, Model 7, Model 11, dan Model 15. Berdasarkan hasil pengujian, kelompok ini memperoleh rata-rata akurasi sebesar 52,70% dan rata-rata *F1-Score* sebesar 19,90%. Nilai akurasi tersebut menunjukkan peningkatan yang cukup signifikan dibandingkan kelompok model yang menggunakan SMOTE.

Tabel 4. 3 Hasil uji coba model tanpa smote

Model	SMOTE	Standardisasi	Father's education	Mother's education	Akurasi	F1-Score	Waktu Training (detik)
3	✗	✓	✓	✓	53,52	20,81	5,85
7	✗	✓	✗	✓	51,34	23,63	3,20
11	✗	✓	✓	✗	52,62	21,26	5,01
15	✗	✓	✗	✗	53,37	13,92	9,25

Berdasarkan tabel 4.3, Model 3 mencatat kinerja terbaik dengan nilai akurasi sebesar 53,50% dan *F1-Score* sebesar 20,80%. Namun, peningkatan akurasi tersebut tidak diikuti oleh kenaikan *F1-Score* yang proporsional. Berdasarkan hasil analisis *confusion matrix*, model dalam kelompok ini cenderung memusatkan prediksi pada kelas *Average Intelligence*, sehingga jumlah prediksi yang benar meningkat dan menghasilkan nilai akurasi yang lebih tinggi. Akan tetapi, kemampuan model dalam mengidentifikasi kelas minoritas masih terbatas, sebagaimana terlihat dari nilai *F1-Score* yang relatif rendah. Temuan ini mengindikasikan bahwa model lebih efektif dalam mengenali kelas mayoritas dibandingkan dalam melakukan klasifikasi seluruh kelas secara seimbang.

Kelompok model yang tidak menerapkan SMOTE maupun standardisasi terdiri atas Model 4, Model 8, Model 12, dan Model 16. Berdasarkan hasil pengujian, kelompok ini menghasilkan rata-rata akurasi tertinggi, yaitu sebesar 53,70%, dengan rata-rata *F1-Score* sebesar 19,09%. Temuan tersebut mengindikasikan bahwa penghapusan kedua teknik *preprocessing* mampu menghasilkan kinerja terbaik apabila ditinjau dari aspek akurasi.

Tabel 4. 4 Hasil uji coba model tanpa smote dan standarisasi

Model	SMOTE	Standardisasi	Father's education	Mother's education	Akurasi	F1-Score	Waktu Training (detik)
4	✗	✗	✓	✓	53,23	20,39	5,49
8	**✗**	**✗**	**✗**	**✓**	**53,20**	**20,40**	**72,04**
12	✗	✗	✓	✗	52,42	18,28	3,47
16	✗	✗	✗	✗	53,37	14,35	17,14

Berdasarkan tabel 4.4, Model 8 memperoleh hasil terbaik dengan nilai akurasi sebesar 55,80% dan *F1-Score* sebesar 23,34%. Berdasarkan analisis

confusion matrix, model-model dalam kelompok ini cenderung memusatkan prediksi pada kelas *Average Intelligence* yang merupakan kelas mayoritas dalam dataset. Kecenderungan tersebut menyebabkan nilai akurasi meningkat karena sebagian besar data berhasil diklasifikasikan ke dalam kelas mayoritas. Namun demikian, kemampuan model dalam mengidentifikasi kelas-kelas minoritas masih relatif rendah sehingga peningkatan nilai *F1-Score* tidak sebanding dengan peningkatan akurasi yang diperoleh. Oleh karena itu, meskipun kelompok model ini menghasilkan akurasi tertinggi, kemampuannya dalam mengenali seluruh kelas secara merata masih belum optimal.

Analisis pengaruh teknik prapemrosesan pada arsitektur (6, 5) dilakukan dengan membandingkan empat konfigurasi utama yang menggunakan variabel input lengkap (*education_father*, *education_mother*, *age_years*, dan *gender*), yaitu Model 1 (SMOTE + Standardisasi), Model 2 (SMOTE saja, tanpa Standardisasi), Model 3 (Standardisasi saja, tanpa SMOTE), dan Model 4 (tanpa keduanya). Ringkasan perbandingan keempat konfigurasi disajikan pada Tabel 4.5 berikut.

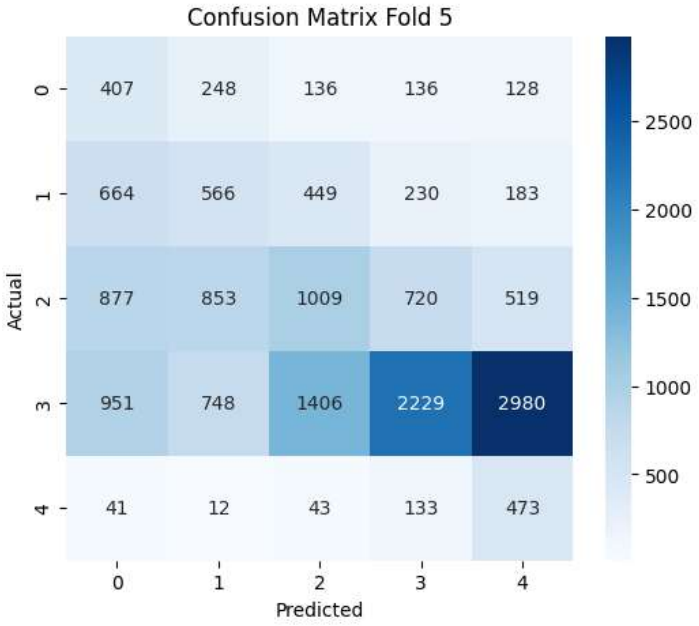
Tabel 4. 5 Perbandingan pengaruh smote dan standarisasi pada arsitektur (6, 5)

Konfigurasi	Model	SMOTE	Standardisasi	Akurasi	Precision	Recall	F1-Score
Full Preprocessing	**1**	✓	✓	**27.90%**	**29.20%**	**37.60%**	**25.80%**
SMOTE saja	2	✓	✗	27.10%	29.10%	37.60%	25.40%
Standardisasi saja	3	✗	✓	53.50%	27.40%	23.80%	20.80%
Tanpa Keduanya	4	✗	✗	53.20%	19.40%	23.80%	20.40%

Berdasarkan Tabel 4.5, terdapat perbedaan yang jelas antara kelompok model dengan SMOTE dan tanpa SMOTE. Model 3 (standardisasi saja) dan Model

4 (tanpa *preprocessing*) menghasilkan akurasi yang tinggi, masing-masing sebesar 53,50% dan 53,20%, tetapi memiliki *Macro F1-Score* yang rendah, yaitu 20,80% dan 20,40%. Hal ini menunjukkan bahwa model cenderung berfokus pada kelas mayoritas sehingga kurang mampu mengenali kelas minoritas. Sebaliknya, Model 1 dan Model 2 yang menggunakan SMOTE menghasilkan akurasi yang lebih rendah (27,90% dan 27,10%), tetapi memperoleh *Macro F1-Score* yang lebih tinggi (25,80% dan 25,40%) serta *recall* sebesar 37,60%, yang menunjukkan kemampuan klasifikasi antar kelas yang lebih seimbang. Sementara itu, pengaruh standardisasi relatif kecil karena perbedaan *Macro F1-Score* antara Model 1 dan Model 2 hanya 0,40%, antara Model 3 dan Model 4 sebesar 0,41%. Dengan demikian, SMOTE menjadi faktor utama yang meningkatkan kualitas klasifikasi pada arsitektur MLP (6,5). Berdasarkan hasil tersebut, Model 1 ditetapkan sebagai model terbaik dengan *Macro F1-Score* sebesar 25,80% dan nilai AUC-ROC pada rentang 0,63-0,78.

Berdasarkan keseluruhan hasil pengujian pada arsitektur MLP (6,5), Model 1 dipilih sebagai model terbaik karena mampu memberikan keseimbangan kinerja klasifikasi yang lebih baik dibandingkan sebagian besar model lainnya. Model ini menerapkan kombinasi SMOTE dan standardisasi dengan menggunakan variabel *education_father* dan *education_mother*, serta menghasilkan akurasi sebesar 27,90% dan *Macro F1-Score* sebesar 25,80%. Meskipun nilai akurasinya lebih rendah dibandingkan beberapa model yang tidak menggunakan SMOTE, Model 1 menunjukkan kemampuan yang lebih baik dalam mengidentifikasi seluruh kategori IQ secara lebih merata.



Gambar 4. 1 Confusion matrix (model terbaik)

Confusion matrix pada Gambar 4.1 diambil dari fold ke-5. Kelas *Average Intelligence* memiliki jumlah prediksi benar tertinggi, yaitu sebanyak 2.229 data, yang menunjukkan bahwa model mampu mengenali pola pada kelas tersebut dengan cukup baik. Namun, masih terdapat kesalahan klasifikasi yang cukup tinggi, terutama ketika 2.980 data pada kelas *Average Intelligence* diprediksi sebagai *Above Average Intelligence*. Selain itu, sebanyak 25,02% data pada kelas *Moderate Intellectual Disability* dan 19,74% data pada kelas *Mild Intellectual Disability* diprediksi sebagai kategori *Average Intelligence* atau *Above Average Intelligence*. Sebaliknya, hanya 7,55% data pada kelas *Above Average Intelligence* yang diprediksi sebagai kategori *Moderate Intellectual Disability* atau *Mild Intellectual Disability*. Hasil ini menunjukkan bahwa model cenderung memprediksi kelas dengan tingkat IQ lebih rendah ke kategori yang lebih tinggi. Meskipun demikian, model mampu menghasilkan prediksi pada seluruh kelas target, sehingga distribusi

prediksi yang dihasilkan lebih merata dibandingkan model yang hanya berfokus pada kelas mayoritas.

4.1.1.2 Pengaruh Kombinasi Variabel Orang Tua

Selain pengaruh teknik preprocessing, penelitian ini juga menganalisis pengaruh kombinasi variabel pendidikan orang tua terhadap performa model klasifikasi kategori IQ anak. Analisis dilakukan untuk mengetahui kontribusi variabel *education_father* dan *education_mother* dalam proses klasifikasi. Untuk memperoleh perbandingan yang lebih objektif, evaluasi dilakukan pada model dengan konfigurasi preprocessing yang sama, yaitu menggunakan SMOTE dan standardisasi. Dengan demikian, perubahan performa yang terjadi dapat lebih merepresentasikan pengaruh variabel pendidikan orang tua dibandingkan pengaruh teknik preprocessing yang digunakan.

Analisis pengaruh variabel pendidikan orang tua dilakukan dengan membandingkan empat konfigurasi yang seluruhnya menggunakan SMOTE dan standardisasi untuk memastikan konsistensi kondisi prapemrosesan, yaitu: Model 1 (variabel lengkap: *education_father* + *education_mother*), Model 5 (tanpa *education_father*, hanya *education_mother*), Model 9 (tanpa *education_mother*, hanya *education_father*), dan Model 13 (tanpa keduanya, hanya *age_years* dan *gender*). Ringkasan perbandingan disajikan pada Tabel 4.6.

Tabel 4. 6 Perbandingan pengaruh variabel pendidikan orang tua pada arsitektur (6, 5)

Konfigurasi Variabel	Model	Father's education	Mother's education	Akurasi	Precision	Recall	F1-Score
Lengkap (Father + Mother)	**1**	✓	✓	**27.90%**	**29.20%**	**37.60%**	**25.80%**

Konfigurasi Variabel	Model	Father's education	Mother's education	Akurasi	Precision	Recall	F1-Score
Tanpa Mother's Education	9	✓	✗	28.20%	29.25%	37.13%	26.15%
Tanpa Father's Education	5	✗	✓	25.30%	28.80%	36.90%	24.10%
Tanpa Keduanya	13	✗	✗	27.62%	22.08%	26.97%	17.87%

Berdasarkan Berdasarkan Tabel 4.6, beberapa temuan penting dapat diidentifikasi. Model 1 yang menggunakan kedua variabel pendidikan orang tua (*education_father* dan *education_mother*) menghasilkan *Macro F1-Score* sebesar 25,80%, yang menunjukkan kinerja klasifikasi yang relatif baik dan seimbang. Sementara itu, Model 9 yang hanya menggunakan variabel *education_father* memperoleh *Macro F1-Score* sebesar 26,15%, sedangkan Model 5 yang hanya menggunakan variabel *education_mother* menghasilkan *Macro F1-Score* sebesar 24,13%. Perbedaan 2,02% tersebut menunjukkan bahwa model yang menggunakan variabel *education_father* memiliki kemampuan klasifikasi yang lebih baik dibandingkan model yang hanya menggunakan variabel *education_mother*.

Selain itu, penghapusan variabel *education_father* menyebabkan penurunan performa yang lebih besar dibandingkan penghapusan variabel *education_mother*. Temuan ini mengindikasikan bahwa variabel *education_father* memiliki peran yang lebih kuat dalam mendukung proses klasifikasi pada arsitektur MLP (6,5). Di sisi lain, ketika kedua variabel pendidikan orang tua dihilangkan sekaligus (Model 13), nilai *Macro F1-Score* turun menjadi 17,87%, jauh lebih rendah dibandingkan model yang masih menggunakan setidaknya satu variabel pendidikan orang tua.

Hasil tersebut menunjukkan bahwa variabel *education_father* dan *education_mother* sama-sama berperan dalam meningkatkan kemampuan model untuk membedakan kategori IQ anak. Model 1 ditetapkan sebagai model terbaik dalam evaluasi pengaruh variabel pendidikan orang tua pada arsitektur MLP (6,5) karena memanfaatkan kedua variabel secara bersamaan dan menghasilkan kinerja klasifikasi yang lebih seimbang pada seluruh kategori IQ.

Tabel 4. 7 Ringkasan hasil skenario arsitektur (6, 5)

Model	**SMOTE**	**Standardisasi**	**Father's education**	**Mother's education**	**Akurasi**	**F1-Score**	**Waktu Training (detik)**
1	✓	✓	✓	✓	**27,90**	**25,80**	**14,72**
2	✓	✗	✓	✓	27,14	25,42	15,52
3	✗	✓	✓	✓	53,52	20,81	5,85
4	✗	✗	✓	✓	53,23	20,39	5,49
5	✓	✓	✗	✓	25,33	24,13	31,58
6	✓	✗	✗	✓	25,13	24,00	270,45
7	✗	✓	✗	✓	51,34	23,63	3,20
8	✗	✗	✗	✓	53,20	20,40	72,04
9	✓	✓	✓	✗	28,20	26,15	15,87
10	✓	✗	✓	✗	28,06	23,23	14,57
11	✗	✓	✓	✗	52,62	21,26	5,01
12	✗	✗	✓	✗	52,42	18,28	3,47
13	✓	✓	✗	✗	27,62	17,87	61,35
14	✓	✗	✗	✗	33,98	18,92	42,08
15	✗	✓	✗	✗	53,37	13,92	9,25
16	✗	✗	✗	✗	53,37	14,35	17,14

4.1.2 Arsitektur Hidden Layer (64, 64)

Kelompok pengujian kedua menggunakan arsitektur *Multilayer Perceptron* (MLP) dengan dua *hidden layer* masing-masing berisi 64 neuron, mencakup Model 17 hingga Model 32. Arsitektur ini memiliki kapasitas representasi yang jauh lebih besar dibandingkan arsitektur (6, 5), sehingga secara teoritis mampu menangkap pola data yang lebih kompleks. Berdasarkan evaluasi menyeluruh menggunakan metrik *accuracy*, *precision*, *recall*, *Macro F1-Score*, dan AUC-ROC, Model 17 ditetapkan sebagai model terbaik pada kelompok arsitektur ini. Model 17 merupakan konfigurasi dengan SMOTE, standardisasi, serta seluruh variabel input (*education_father*, *education_mother*, *age_years*, dan *gender*), menghasilkan *accuracy* sebesar 28,66%, *Macro F1-Score* sebesar 26,30%, *recall* sebesar 37,90%, dan nilai AUC-ROC pada rentang 0,63–0,79. Model ini menunjukkan performa terbaik dan paling seimbang dalam mengklasifikasikan seluruh kategori kemampuan intelektual pada kelompok arsitektur (64, 64).

4.1.2.1 Pengaruh *Smote* dan *Standarisasi*

Pengaruh teknik prapemrosesan pada arsitektur (64, 64) dilakukan dengan membandingkan empat konfigurasi yang menggunakan variabel input lengkap, yaitu Model 17 (SMOTE + Standardisasi), Model 18 (SMOTE saja, tanpa Standardisasi), Model 19 (Standardisasi saja, tanpa SMOTE), dan Model 20 (tanpa keduanya). Ringkasan perbandingan disajikan pada Tabel 4.8.

Tabel 4. 8 Perbandingan pengaruh smote dan standarisasi pada arsitektur (64, 64)

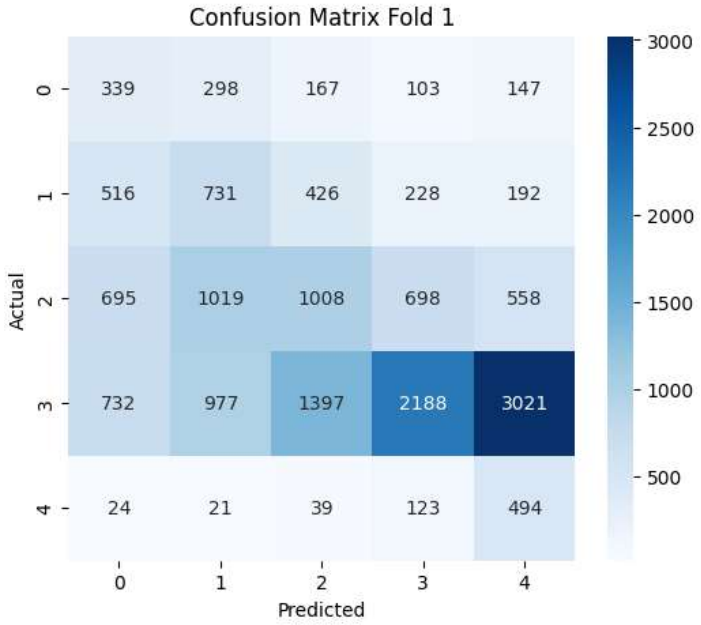
Konfigurasi Variabel	Model	Smote	Standarisa	Akurasi	Precision	Recall	F1-Score
Full Preprocessing	**17**	✓	✓	**28.70%**	**29.50%**	**37.90%**	**26.30%**

Konfigurasi Variabel	Model	Smote	Standarisa	Akurasi	Precision	Recall	F1-Score
SMOTE saja	18	✓	✗	29.10%	29.60%	38.00%	26.30%
Standardisasi saja	19	✗	✓	53.72%	34.73%	25.51%	24.06%
Tanpa Keduanya	20	✗	✗	53.63%	37.19%	24.63%	22.28%

Berdasarkan Tabel 4.8, terlihat pola hasil yang serupa dengan arsitektur MLP (6,5). Model 19 yang hanya menggunakan standardisasi dan Model 20 yang tidak menerapkan *preprocessing* menghasilkan akurasi yang tinggi, masing-masing sebesar 53,72% dan 53,63%. Namun, nilai *Macro F1-Score* yang diperoleh relatif rendah, yaitu 24,06% dan 22,28%. Perbedaan yang cukup besar antara akurasi dan *Macro F1-Score* tersebut menunjukkan bahwa model cenderung berfokus pada kelas mayoritas, yaitu *Average Intelligence*, sehingga kemampuan dalam mengenali kelas-kelas minoritas masih terbatas.

Sebaliknya, Model 17 (SMOTE + standardisasi) dan Model 18 (SMOTE tanpa standardisasi) menghasilkan *Macro F1-Score* yang sama, yaitu 26,30%, dengan nilai *recall* masing-masing sebesar 37,90% dan 38,00%. Hasil ini menunjukkan bahwa penerapan SMOTE mampu meningkatkan kemampuan model dalam mengenali kelas minoritas, yang tercermin dari peningkatan *recall* lebih dari 13 poin persentase dibandingkan model tanpa SMOTE. Sementara itu, pengaruh standardisasi relatif kecil karena perbedaan akurasi antara Model 17 dan Model 18 hanya sebesar 0,40%, dengan nilai *Macro F1-Score* yang identik. Temuan ini mengindikasikan bahwa SMOTE merupakan faktor utama yang meningkatkan keseimbangan klasifikasi pada arsitektur MLP (64,64), sedangkan kontribusi standardisasi terhadap peningkatan performa model tidak terlalu signifikan.

Meskipun metrik yang dihasilkan Model 17 dan Model 18 hampir sama, Model 17 dipilih sebagai model terbaik dalam analisis pengaruh *preprocessing* karena tetap mempertahankan standardisasi, memiliki nilai AUC-ROC yang setara (0,63-0,79), menunjukkan polapenurunan *loss* yang lebih konsisten selama proses pelatihan.



Gambar 4. 2 *Confusion matrix* (model 17)

Berdasarkan *confusion matrix* pada Gambar 4.2, terlihat bahwa model masih mengalami kesalahan klasifikasi antar kategori tingkat kecerdasan. Pada kategori *Moderate Intellectual Disability*, kesalahan prediksi terbesar mengarah ke kategori *Mild Intellectual Disability* sebesar 41,68%, diikuti *Below Average Intelligence* sebesar 23,36%, *Above-Average Intelligence* sebesar 20,56%, dan *Average Intelligence* sebesar 14,41% dari total kesalahan pada kelas tersebut. Pada kategori *Mild Intellectual Disability*, kesalahan prediksi paling banyak terjadi ke kategori *Moderate Intellectual Disability* sebesar 37,38%, diikuti *Below Average Intelligence* sebesar 30,89%, *Average Intelligence* sebesar 16,52%, dan *Above-*

Average Intelligence sebesar 13,91%. Pada kategori *Below Average Intelligence*, kesalahan klasifikasi terbesar mengarah ke kategori *Mild Intellectual Disability* sebesar 51,39%, diikuti *Average Intelligence* sebesar 35,21% dan *Above-Average Intelligence* sebesar 28,17%. Pola yang sama pada kategori *Average Intelligence*, di mana sebagian besar kesalahan prediksi diklasifikasikan sebagai *Above-Average Intelligence*, yaitu sebesar 49,41% dari total kesalahan pada kelas tersebut.

4.1.2.2 Pengaruh Kombinasi Variabel Pendidikan Orang Tua

Analisis pengaruh variabel pendidikan orang tua pada arsitektur (64, 64) dilakukan dengan membandingkan empat konfigurasi yang seluruhnya menggunakan SMOTE dan standardisasi, yaitu: Model 17 (variabel lengkap), Model 21 (tanpa *education_father*, hanya *education_mother*), Model 25 (tanpa *education_mother*, hanya *education_father*), dan Model 29 (tanpa keduanya, hanya *age_years* dan *gender*). Ringkasan perbandingan disajikan pada Tabel 4.9.

Tabel 4. 9 Perbandingan pengaruh variabel pendidikan orang tua pada arsitektur (64, 64)

Konfigurasi Variabel	Model	Father's education	Mother's education	Akurasi	Precision	Recall	F1-Score
Lengkap (Father + Mother)	**17**	✓	✓	**28.70%**	**29.50%**	**37.90%**	**26.30%**
Tanpa Education Mother	25	✓	✗	27.83%	29.25%	37.25%	25.84%
Tanpa Education Father	21	✗	✓	25.13%	28.88%	37.24%	24.03%
Tanpa Keduanya	29	✗	✗	29.28%	24.44%	27.84%	18.95%

Tabel 4.9, temuan yang konsisten dengan arsitektur (6, 5) kembali teridentifikasi. Model 17 (variabel lengkap) menghasilkan *Macro F1-Score*

tertinggi sebesar 26,30%, yang membuktikan bahwa penggunaan kedua variabel pendidikan orang tua secara bersamaan memberikan representasi informasi yang paling komprehensif bagi model.

Model 25 (tanpa *education_mother*, hanya menggunakan *education_father*) menghasilkan *Macro F1-Score* sebesar 25,84%, sedangkan Model 21 (tanpa *education_father*, hanya menggunakan *education_mother*) hanya menghasilkan *Macro F1-Score* sebesar 24,03%. Selisih *Macro F1-Score* antara konfigurasi tanpa *education_mother* (Model 25, F1=25,84%) dan tanpa *education_father* (Model 21, F1=24,03%) adalah sebesar 1,81 poin persentase. Apabila dihitung secara relatif terhadap nilai Model 25, penurunan performa akibat penghapusan *education_father* mencapai sekitar 7,00% (selisih 1,81 poin dari baseline 25,84%). Temuan ini mengkonfirmasi bahwa education_father memberikan kontribusi yang lebih besar terhadap performa model dibandingkan *education_mother* pada arsitektur (64, 64). Menghapus variabel *education_father* menyebabkan penurunan *Macro F1-Score* yang lebih signifikan (2,27 poin dari Model 17) dibandingkan penghapusan *education_mother* yang hanya menyebabkan penurunan 0,46 poin.

Penghapusan kedua variabel pendidikan orang tua (Model 29) menyebabkan penurunan *Macro F1-Score* yang sangat drastis menjadi 18,95%, menandakan bahwa tanpa informasi pendidikan orang tua, model sangat kesulitan dalam membedakan kategori kemampuan intelektual secara akurat. Model 17 (variabel lengkap, SMOTE + Standardisasi) ditetapkan sebagai model terbaik dalam analisis pengaruh variabel pendidikan orang tua pada arsitektur (64, 64).

Tabel 4. 10 Ringkasan hasil skenario arsitektur (64, 64)

Model	SMOTE	Standardisasi	Education father	Education mother	Akurasi	F1-Score
17	✓	✓	✓	✓	**28,66**	**26,30**
18	✓	✗	✓	✓	29,10	26,28
19	✗	✓	✓	✓	53,72	24,06
20	✗	✗	✓	✓	53,63	22,28
21	✓	✓	✗	✓	25,13	24,03
22	✓	✗	✗	✓	25,13	24,00
23	✗	✓	✗	✓	55,82	23,73
24	✗	✗	✗	✓	55,80	23,34
25	✓	✓	✓	✗	27,83	25,84
26	✓	✗	✓	✗	28,18	25,93
27	✗	✓	✓	✗	52,81	23,31
28	✗	✗	✓	✗	52,52	21,42
29	✓	✓	✗	✗	29,28	18,95
30	✓	✗	✗	✗	33,22	20,00
31	✗	✓	✗	✗	53,38	14,18
32	✗	✗	✗	✗	53,37	14,16

4.1.3 Analisis Perbandingan Model Arsitektur

Setelah ditetapkan model terbaik dari masing-masing kelompok arsitektur pada subab 4.1.1 dan 4.1.2, yaitu Model 1 (arsitektur (6, 5)) dan Model 17 (arsitektur (64, 64)), dilakukan analisis komparatif akhir (*head-to-head*) untuk menentukan model terbaik secara keseluruhan dari seluruh penelitian ini. Perbandingan dilakukan berdasarkan seluruh dimensi evaluasi yang digunakan, meliputi *accuracy*, *precision*, *recall*, *Macro F1-Score*, AUC-ROC, kestabilan performa antar *fold*, pola penurunan *loss*, *confusion matrix*, serta *classification report*. Ringkasan perbandingan kedua model terbaik disajikan pada Tabel 4..

Tabel 4. 11 Perbandingan model terbaik (6, 5 dan 64, 64)

Metrik Evaluasi	Model 1 Arsitektur (6, 5)	Model 17 Arsitektur (64, 64)	Selisih
Arsitektur	(6, 5)	(64, 64)	-
SMOTE	✓	✓	-
Standardisasi	✓	✓	-
Variabel Input	Lengkap (4 variabel)	Lengkap (4 variabel)	-
Accuracy (%)	27,90	28,66	+0,76
Precision (%)	29,20	29,50	+0,30
Recall (%)	37,60	37,90	+0,30
F1-Score (%)	25,80	26,30	+0,50
AUC-ROC (rentang)	0,63 – 0,78	0,63 – 0,79	+0,01
Stabilitas K-Fold	Stabil	Stabil	-
Indikasi Overfitting	Tidak ada	Tidak ada	-
Kesalahan Klasifikasi *extream*	12,13%	13,95%	+1,82%

Berdasarkan Tabel 4.11, perbandingan langsung antara Model 1 dan Model 17 menunjukkan bahwa Model 17 memperoleh kinerja yang sedikit lebih baik pada seluruh metrik evaluasi. Model 17 menghasilkan *accuracy* sebesar 28,66%, lebih tinggi 0,76% dibandingkan Model 1 (27,90%). Selain itu, nilai *Macro F1-Score* Model 17 mencapai 26,30%, lebih tinggi 0,50% dibandingkan Model 1 (25,80%), sedangkan nilai *recall* meningkat dari 37,60% menjadi 37,90%. Rentang AUC-ROC Model 17 juga sedikit lebih tinggi, yaitu 0,63-0,79 dibandingkan 0,63-0,78 pada Model 1. Berdasarkan analisis *confusion matrix*, sebagian besar kesalahan klasifikasi pada Model 17 terjadi antar kelas yang memiliki karakteristik berdekatan, sementara kesalahan pada kelas yang berjauhan relatif lebih rendah. Selain itu, kedua model menunjukkan pola penurunan *loss* yang stabil selama proses pelatihan tanpa indikasi *overfitting* yang kuat.

Meskipun arsitektur Model 17 lebih kompleks dibandingkan Model 1, peningkatan performa yang diperoleh relatif kecil. Hasil ini menunjukkan bahwa keterbatasan performa model dipengaruhi oleh informasi data yang tersedia pada variabel input daripada jumlah kapasitas jaringan yang digunakan. Dengan demikian, peningkatan kompleksitas arsitektur belum mampu memberikan peningkatan performa yang signifikan. Berdasarkan keseluruhan hasil evaluasi, Model 17 ditetapkan sebagai model terbaik dalam penelitian ini. Model tersebut menggunakan arsitektur MLP (64,64), menerapkan SMOTE dan standardisasi, serta memanfaatkan seluruh variabel input (*education_father*, *education_mother*, *age_years*, dan *gender*). Model 17 menghasilkan *accuracy* sebesar 28,66%, *precision* sebesar 29,50%, *recall* sebesar 37,90%, *Macro F1-Score* sebesar 26,30%, dan AUC-ROC pada rentang 0,63-0,79. Hasil tersebut menunjukkan bahwa Model 17 memiliki kinerja klasifikasi yang paling baik dan paling seimbang dalam memprediksi kategori IQ anak berdasarkan tingkat pendidikan orang tua.

Tabel 4. 12 Ringkasan hasil seluruh model

Model	**SMOTE**	**Standardisasi**	**Education father**	**Education mother**	**Akurasi**	**F1-Score**
1	✓	✓	✓	✓	27,90	25,80
2	✓	✗	✓	✓	27,14	25,42
3	✗	✓	✓	✓	53,52	20,81
4	✗	✗	✓	✓	53,23	20,39
5	✓	✓	✗	✓	25,33	24,13
6	✓	✗	✗	✓	25,13	24,00
7	✗	✓	✗	✓	51,34	23,63
8	✗	✗	✗	✓	53,20	20,40
9	✓	✓	✓	✗	28,20	26,15

Model	SMOTE	Standardisasi	Education father	Education mother	Akurasi	F1-Score
10	✓	✗	✓	✗	28,06	23,23
11	✗	✓	✓	✗	52,62	21,26
12	✗	✗	✓	✗	52,42	18,28
13	✓	✓	✗	✗	27,62	17,87
14	✓	✗	✗	✗	33,98	18,92
15	✗	✓	✗	✗	53,37	13,92
16	✗	✗	✗	✗	53,37	14,35
17	✓	✓	✓	✓	28,66	26,30
18	✓	✗	✓	✓	29,10	26,28
19	✗	✓	✓	✓	53,72	24,06
20	✗	✗	✓	✓	53,63	22,28
21	✓	✓	✗	✓	25,13	24,03
22	✓	✗	✗	✓	25,13	24,00
23	✗	✓	✗	✓	55,82	23,73
24	✗	✗	✗	✓	55,80	23,34
25	✓	✓	✓	✗	27,83	25,84
26	✓	✗	✓	✗	28,18	25,93
27	✗	✓	✓	✗	52,81	23,31
28	✗	✗	✓	✗	52,52	21,42
29	✓	✓	✗	✗	29,28	18,95
30	✓	✗	✗	✗	33,22	20,00
31	✗	✓	✗	✗	53,38	14,18
32	✗	✗	✗	✗	53,37	14,16

4.3 Impelemtasi GUI

Model Model yang digunakan pada implementasi antarmuka grafis (GUI) adalah model Multilayer Perceptron (MLP) dengan arsitektur hidden layer (64,64)

yang menerapkan teknik SMOTE dan standardisasi data, serta menggunakan seluruh variabel input, yaitu *education_mother*, *education_father*, *age_years*, dan *gender*. Pemilihan model ini didasarkan pada hasil perbandingan seluruh skenario eksperimen yang telah dibahas pada Subbab 4.1.3. Berdasarkan hasil evaluasi menggunakan accuracy, precision, recall, Macro F1-Score, AUC-ROC, kestabilan performa antar fold pada Stratified 5-Fold Cross Validation, pola penurunan loss, confusion matrix, dan classification report, Model 13 menunjukkan performa yang paling baik dan paling seimbang dibandingkan seluruh model yang diuji.

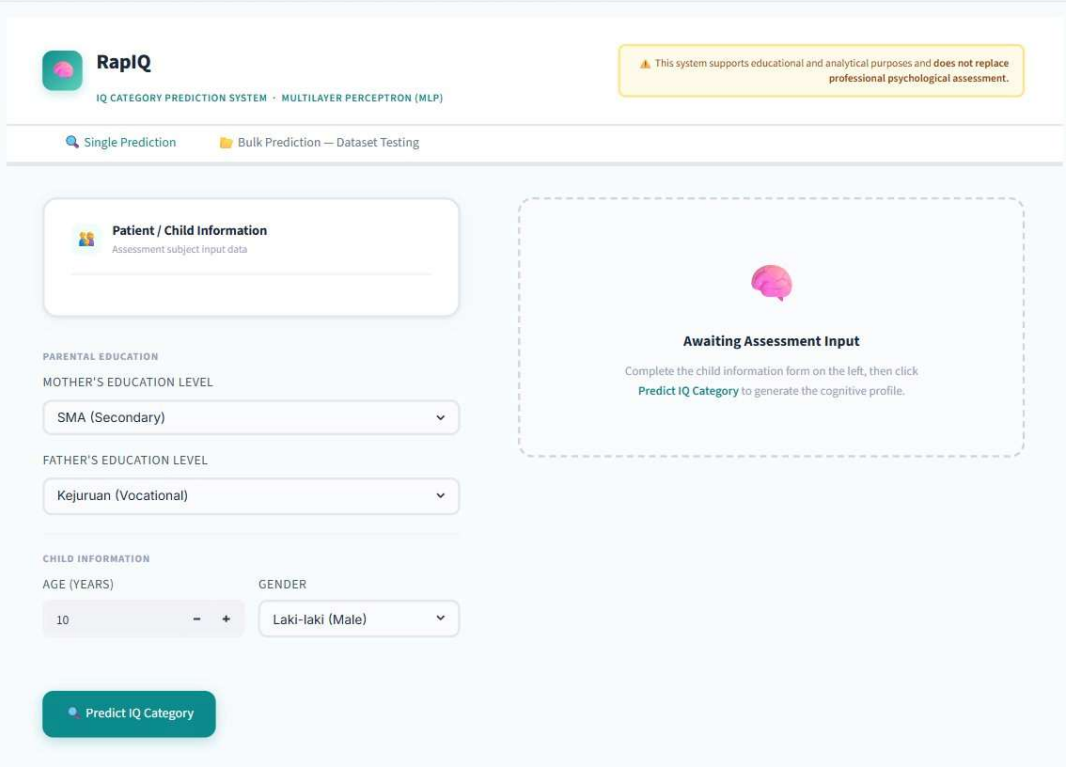
Model 17 menghasilkan accuracy sebesar 28,66%, precision sebesar 29,50%, recall sebesar 37,90%, dan Macro F1-Score sebesar 26,30%, yang merupakan performa terbaik secara keseluruhan pada penelitian ini. Selain itu, hasil confusion matrix menunjukkan bahwa sebagian besar kesalahan klasifikasi terjadi pada kelas-kelas yang berdekatan, sedangkan kesalahan klasifikasi antar kategori yang memiliki perbedaan karakteristik yang jauh relatif lebih rendah. Model ini juga menunjukkan kemampuan generalisasi yang baik melalui performa yang stabil pada seluruh fold serta tidak menunjukkan indikasi overfitting selama proses pelatihan. Berdasarkan pertimbangan tersebut, Model 17 dipilih sebagai model yang digunakan pada implementasi GUI karena mampu memberikan performa klasifikasi yang paling baik, paling seimbang, dan paling representatif untuk memprediksi kategori tingkat kecerdasan anak berdasarkan variabel yang digunakan dalam penelitian.

Antarmuka GUI yang dibangun diberi nama RapIQ (Rapid IQ Category Prediction System), sebuah sistem prediksi kategori IQ anak berbasis web yang

menggunakan model MLP sebagai mesin inferensi. Sistem ini dirancang untuk mendukung keperluan edukasi dan analisis, dan secara eksplisit tidak dimaksudkan sebagai pengganti asesmen psikologis profesional, sebagaimana tertera dalam peringatan pada halaman RapIQ. RapIQ menyediakan dua mode operasi utama, yaitu *Single Prediction* untuk prediksi satu individu, dan *Bulk Prediction* untuk prediksi massal berbasis dataset CSV.

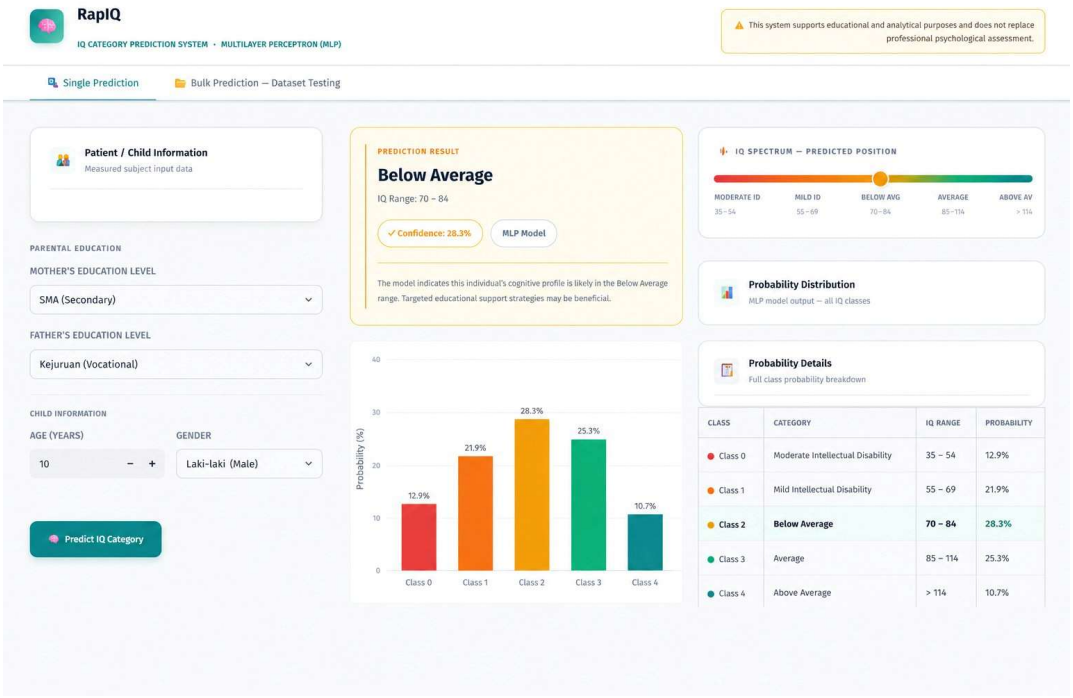
A. Single Prediction

Single Prediction mengharuskan pengguna memasukkan data seorang anak secara individual untuk memperoleh prediksi kategori IQ. Pengisian input terdiri atas empat variabel, yaitu tingkat pendidikan ibu (*education_mother*), tingkat pendidikan ayah (*education_father*), usia anak dalam tahun (*age_years*), dan jenis kelamin (*gender*). Pada contoh yang ditampilkan, inputan yang dimasukkan adalah pendidikan ibu SMA (Secondary), pendidikan ayah Kejuruan (Vocational), usia anak 10 tahun, dan jenis kelamin Laki-laki. Tampilan pengisian input dapat dilihat pada Gambar 4.17.



Gambar 4. 3 Tampilan formulir *input single prediction*

Setelah pengguna menekan tombol *Predict IQ Category*, sistem akan memproses data input melalui model MLP yang telah dilatih dan menampilkan hasil prediksi secara lengkap. Hasil prediksi mencakup kategori IQ yang diprediksi beserta rentang IQ yang bersesuaian, nilai kepercayaan (confidence score), distribusi probabilitas ke seluruh kelas, serta posisi kategori pada spektrum IQ. Tampilan hasil prediksi dapat dilihat pada Gambar 4.98.



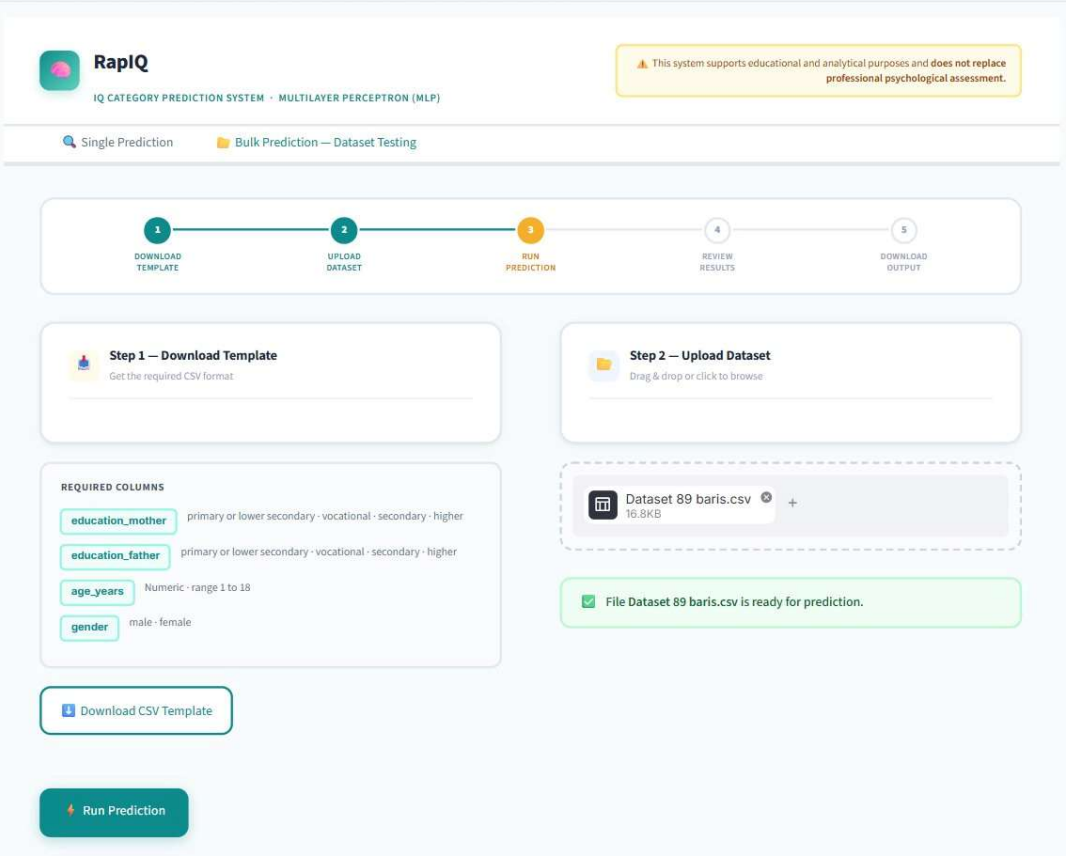
Gambar 4. 4 Tampilan hasil prediksi *single prediction*

Berdasarkan Gambar 4.98, sistem memprediksi kategori IQ anak dengan inputan tersebut sebagai Below Average dengan rentang IQ 70-84 dan *confidence score* sebesar 28,3%. Nilai *confidence score* yang relatif rendah ini konsisten dengan karakteristik model yang telah dibahas sebelumnya, dimana distribusi probabilitas tersebar cukup merata ke seluruh kelas dengan penerapan SMOTE yang menyeimbangkan data latih. Pada panel distribusi probabilitas (*Probability Details*), terlihat bahwa model memberikan probabilitas tertinggi pada Class 2 (*Below Average*) sebesar 28,3%, diikuti Class 3 (*Average*) sebesar 25,3%, Class 1 (*Mild Intellectual Disability*) sebesar 21,9%, Class 0 (*Moderate Intellectual Disability*) sebesar 12,9%, dan Class 4 (*Above Average*) sebesar 10,7%. Distribusi probabilitas yang tidak jauh berbeda antar kelas ini, mencerminkan kemampuan model dalam mengenali seluruh kategori secara lebih seimbang, sebagai bukti dari

penerapan SMOTE pada data latih. Hasil prediksi Below Average dapat dipahami karena kombinasi pendidikan orang tua yang berada pada tingkat menengah ke bawah (SMA dan Kejuruan) serta usia anak yang masih berada pada rentang 10 tahun, sehingga model memetakan profil tersebut ke kategori di bawah rata-rata berdasarkan pola yang dipelajari dari data SB5. Panel IQ Spectrum pada bagian kanan atas menampilkan posisi prediksi secara visual di sepanjang spektrum kategori IQ, dengan penanda yang berada pada rentang Below Average (70-84).

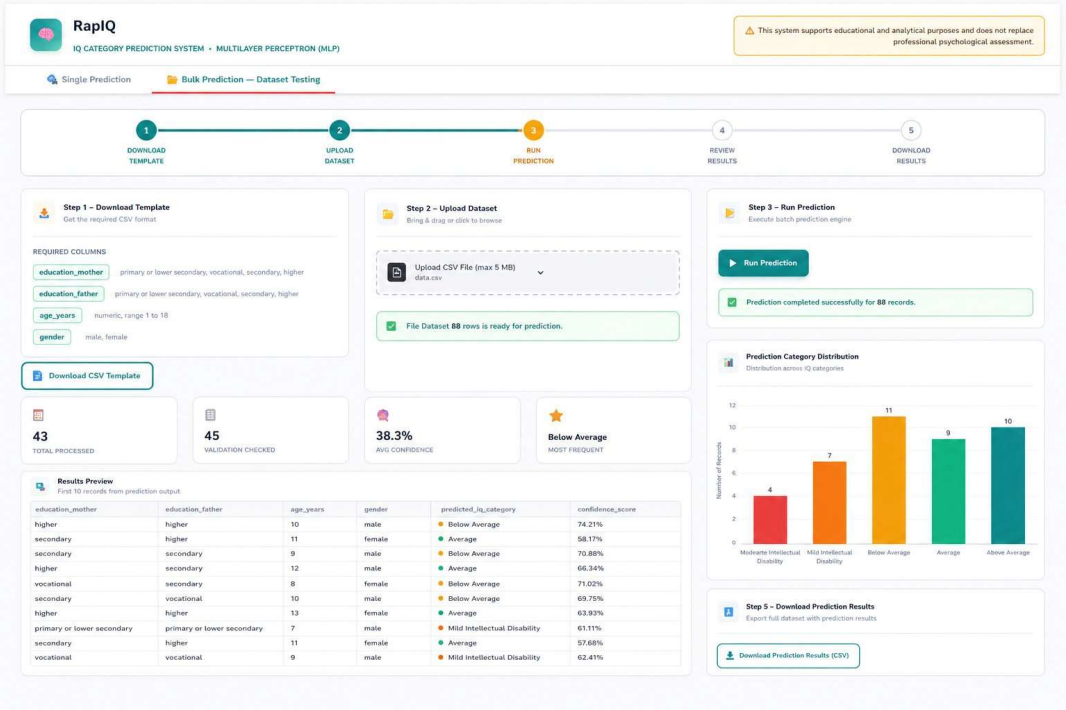
B. Bulk Prediction

Bulk Prediction dirancang untuk memproses prediksi secara massal terhadap banyak data sekaligus melalui unggahan berkas CSV. Mode ini mengikuti alur kerja lima langkah yang ditampilkan secara visual, yaitu: (1) *Download Template*, (2) *Upload Dataset*, (3) *Run Prediction*, (4) *Review Results*, dan (5) *Download Output*. Pengguna terlebih dahulu mengunduh template CSV yang memuat format kolom yang diperlukan, yaitu *education_mother, education_father, age_years*, dan *gender*. Selanjutnya berkas CSV yang telah diisi diunggah ke sistem. Tampilan antarmuka input Bulk Prediction dapat dilihat pada Gambar 4.99.



Gambar 4. 5 Tampilan antarmuka *input bulk prediction*

Pada contoh pengujian, berkas *Dataset 99 baris.csv* berukuran 16,8 KB berhasil diunggah dan divalidasi oleh sistem, yang ditandai dengan pesan konfirmasi bahwa berkas siap untuk diproses. Setelah pengguna menekan tombol *Run Prediction*, sistem memproses seluruh baris data melalui model MLP dan menampilkan hasil secara komprehensif. Tampilan hasil Bulk Prediction dapat dilihat pada Gambar 4.100.



Gambar 4. 6 Tampilan hasil *bulk prediction*

Berdasarkan Gambar 4.100, dari 88 baris data yang berhasil diproses, sistem berhasil menyelesaikan prediksi untuk 43 rekord dengan 45 rekord yang telah melalui validasi. Rata-rata *confidence score* yang dihasilkan adalah 38,3%, dengan kategori yang paling sering diprediksi adalah *Below Average*. Panel pratinjau hasil menampilkan 10 rekord pertama lengkap dengan kolom variabel input, kategori IQ yang diprediksi, dan nilai kepercayaan masing-masing rekord. Sebagai contoh, anak dengan pendidikan ibu dan ayah keduanya higher dan usia 10 tahun (laki-laki) diprediksi sebagai *Below Average* dengan prediksi sekitar 74,21%, sedangkan anak dengan pendidikan ibu secondary, ayah higher, usia 11 tahun (perempuan) diprediksi sebagai *Average* dengan kepercayaan 58,17%. Perbedaan hasil ini mencerminkan pola yang dipelajari model dari data latih, dimana kombinasi variabel menghasilkan pemetaan ke kategori yang berbeda-beda.

Grafik distribusi kategori prediksi (*Prediction Category Distribution*) yang ditampilkan pada panel antarmuka bersumber dari akumulasi hasil prediksi seluruh rekord dalam dataset yang diunggah. Grafik batang tersebut menunjukkan bahwa kategori *Below Average* memiliki frekuensi prediksi tertinggi dengan 11 rekord, diikuti *Above Average* (10 rekord), *Average* (9 rekord), *Mild Intellectual Disability* (7 rekord), dan *Moderate Intellectual Disability* (4 rekord). Distribusi prediksi yang menyebar ke seluruh kategori ini merupakan dampak langsung dari penerapan SMOTE pada model MLP (64, 64), dimana penyeimbangan data latih mendorong model untuk tidak hanya memprediksi satu kelas dominan, tetapi mampu mengenali seluruh kategori IQ secara lebih proporsional. Hal ini konsisten dengan temuan eksperimen model 17 pada Subbab 4.1.2.1 yang menunjukkan bahwa model dengan SMOTE menghasilkan distribusi prediksi yang lebih merata dibandingkan model tanpa SMOTE. Pengguna dapat mengunduh hasil prediksi lengkap dalam format CSV melalui tombol *Download Prediction Results* yang tersedia pada antarmuka.

4.4 Pembahasan dan Integrasi Islam

Secara keseluruhan, hasil eksperimen penelitian ini menunjukkan bahwa model *Multilayer Perceptron* (MLP) mampu diimplementasikan untuk mengklasifikasikan lima kategori IQ anak berdasarkan tingkat pendidikan orang tua, usia, dan jenis kelamin, meskipun performa yang dicapai masih termasuk *moderat* atau rendah. Skenario terbaik dalam penelitian ini adalah model 17 dengan konfigurasi yang menerapkan teknik *Synthetic Minority Over-sampling Technique* (SMOTE) dikombinasikan dengan standardisasi serta menggunakan seluruh variabel, yang menghasilkan akurasi akhir sebesar 28,66% dengan nilai F1-*Score*

sebesar 26,30% dan AUC-ROC sebesar 0,70. Nilai akurasi yang terlihat rendah ini tidak dapat dipastikan sebagai kegagalan model, melainkan sebagai cerminan karakteristik data dan kompleksitas masalah yang dihadapi.

Alasan mendasar mengapa skenario dengan SMOTE justru menunjukkan penurunan akurasi dibandingkan skenario tanpa SMOTE (53,23%), yang terletak pada perbedaan cara model mengambil keputusan. Tanpa SMOTE, model cenderung mendominasi prediksi pada kelas mayoritas, yaitu *Average intelligence*, sehingga nilai akurasi tampak tinggi namun gagal mengenali kelas-kelas minoritas seperti *Moderate intellectual disability* dan *Above-average intelligence*. Hal ini terbukti dari nilai AUC kelas *above_average* yang bernilai 0,00 pada skenario tanpa SMOTE, menandakan model sama sekali tidak mampu membedakan kelas tersebut dari kelas lainnya. Setelah penerapan SMOTE, model dapat mempelajari pola dari seluruh kelas dengan lebih proporsional, sehingga distribusi prediksi menjadi lebih merata meskipun nilai akurasi turun. Kondisi ini menunjukkan bahwa model dengan SMOTE lebih andal dalam konteks penggunaan nyata, sebab mampu memberikan prediksi yang bermakna untuk semua kategori IQ, bukan hanya kategori yang paling sering muncul dalam data.

Keterbatasan performa model secara umum juga dapat dijelaskan melalui karakteristik variabel input yang digunakan. Tingkat pendidikan orang tua, usia, dan jenis kelamin merupakan prediktor yang memiliki hubungan terhadap skor IQ, namun kemampuannya dalam membedakan kategori IQ secara individual masih sangat terbatas. Penelitian (Sajewicz-Radtke, Łada-Maśko, Olech, Jurek, et al., 2025) menunjukkan adanya keterkaitan antara pendidikan orang tua dan nilai IQ

anak, Namun, temuan tersebut tidak dapat diinterpretasikan sebagai hubungan sebab-akibat secara langsung. Kecerdasan anak dapat dipengaruhi oleh berbagai faktor lain, termasuk faktor genetik, lingkungan rumah, dan stimulasi kognitif, yang tidak diukur secara langsung dalam model analisis penelitian ini. Oleh sebab itu, model MLP dalam penelitian ini sesungguhnya bukan gagal mengklasifikasi IQ anak berdasakan pendidikan orang tua, melainkan mencerminkan batas atas informasi yang dapat diambil dari fitur yang tersedia. Dari perbandingan kedua arsitektur, menunjukkan bahwa peningkatan kapasitas jaringan dari (6, 5) ke (64, 64) tidak menghasilkan peningkatan performa yang signifikan, karena variabel input yang digunakan belum menyediakan informasi yang cukup untuk membedakan setiap kelas IQ secara lebih akurat, peningkatan kompleksitas jaringan tidak menghasilkan peningkatan performa model yang signifikan.

Hasil ini menunjukkan bahwa proses klasifikasi menggunakan *machine learning* belum sepenuhnya mampu merepresentasikan kemampuan intelektual manusia hanya berdasarkan latar belakang pendidikan orang tua. Model hanya mempelajari pola dari data yang tersedia, sehingga hasil klasifikasi yang dihasilkan bersifat terbatas dan tidak dapat menggambarkan keseluruhan potensi serta nilai seorang manusia. Dalam perspektif Islam, hal ini memiliki relevansi mendalam dengan firman Allah SWT dalam QS. Az-Zukhruf ayat 32:

أَهُمْ يَقْسِمُوْنَ رَحْمَتَ رَبِّكَ؟ نَحْنُ قَسَمْنَا بَيْنَهُمْ مَعِيشَتَهُمْ فِى الْحَيٰوةِ الدُّنْيَا؛ وَرَفَعْنَا بَعْضَهُمْ فَوْقَ بَعْضٍ دَرَجٰتٍ لِّيَتَّخِذَ بَعْضُهُمْ بَعْضًا سُخْرِيًّا؟ وَرَحْمَتُ رَبِّكَ خَيْرٌ مِّمَّا يَجْمَعُوْنَ ٣٢

“Apakah mereka yang membagi-bagikan rahmat Tuhanmu? Kami telah menentukan penghidupan mereka dalam kehidupan dunia, dan Kami telah meninggikan sebagian mereka atas sebagian yang lain beberapa derajat, agar

sebagian mereka dapat memanfaatkan sebagian yang lain. Dan rahmat Tuhanmu lebih baik dari apa yang mereka kumpulkan." (QS. Az-Zukhruf [43]: 32).

Ayat ini menegaskan bahwa perbedaan kapasitas dan kemampuan di antara manusia merupakan bagian dari ketetapan Allah SWT, yang bertujuan agar setiap individu dapat saling melengkapi dalam kehidupan bermasyarakat. Ibnu Katsir dalam tafsirnya menjelaskan bahwa Allah-lah yang membagi penghidupan manusia di dunia dan mengangkat salah satu manusia dariapada manusia yang lain dalam hal derajat, bukan agar yang satu lebih mulia secara mutlak dari yang lain, melainkan agar manusia saling membutuhkan satu sama lain, sehingga terbentuk keadaaan sosial yang saling bergantung dan melengkapi. Ibnu Katsir menegaskan bahwa rahmat Allah jauh lebih baik dari segala sesuatu yang dikumpulkan manusia di dunia, sehingga perbedaan derajat duniawi tersebut tidak dapat dijadikan tolok ukur kemuliaan sejati seseorang (Ibnu Katsir, *Tafsir al-Qur'an al-'Adzim*, Jilid 7, hlm. 225–226). Dengan demikian, hasil klasifikasi IQ yang dihasilkan model dalam penelitian ini tidak dapat dijadikan penilaian mutlak atas nilai atau derajat seorang anak, karena potensi manusia bersifat multidimensional dan tidak dapat diprediksi hanya melalui variabel statistik yang terbatas.

Dengan demikian, integrasi nilai keislaman dalam penelitian ini menegaskan bahwa teknologi *machine learning*, termasuk MLP, hanyalah alat bantu yang bersifat probabilistik dan tidak dapat dijadikan ukuran mutlak kemampuan intelektual seseorang. Oleh karena itu, pemanfaatannya harus dilakukan secara bijaksana dengan memperhatikan keterbatasan model, serta menyadari bahwa potensi yang dianugerahkan Allah kepada setiap manusia jauh lebih luas daripada yang dapat diukur oleh sistem klasifikasi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Penelitian ini berhasil mengimplementasikan metode *Multilayer Perceptron* (MLP) untuk mengklasifikasikan kemampuan intelektual anak ke dalam lima kategori, yaitu *Moderate Intellectual Disability*, *Mild Intellectual Disability*, *Below Average Intelligence*, *Average Intelligence*, dan *Above-average Intelligence*, berdasarkan variabel *education_father*, *education_mother*, *age_years*, dan *gender* dari data *Stanford-Binet Intelligence Scales Fifth Edition* (SB5). Penelitian melibatkan 32 skenario eksperimen yang dibentuk dari kombinasi dua arsitektur *hidden layer*, dua teknik prapemrosesan, dan empat kombinasi fitur pendidikan orang tua, kemudian dievaluasi menggunakan *Stratified 5-Fold Cross Validation* dengan metrik *accuracy*, *precision*, *recall*, *Macro F1-Score*, dan *AUC-ROC*.

Hasil analisis menunjukkan bahwa penerapan *Synthetic Minority Oversampling Technique* (SMOTE) memberikan pengaruh paling besar terhadap peningkatan performa model. Model yang menggunakan SMOTE secara konsisten menghasilkan nilai *recall* dan *Macro F1-Score* yang lebih tinggi dibandingkan model tanpa SMOTE karena distribusi data latih menjadi lebih seimbang dan bias terhadap kelas mayoritas (*Average Intelligence*) berkurang. Sebaliknya, pengaruh standardisasi relatif kecil, yang ditunjukkan oleh perbedaan *Macro F1-Score* antara model dengan dan tanpa standardisasi yang tidak melebihi 0,50 poin persentase.

Analisis pengaruh variabel pendidikan orang tua menunjukkan bahwa variabel *education_father* memberikan kontribusi yang lebih besar dibandingkan

education_mother. Pada arsitektur (6, 5), model yang hanya menggunakan *education_father* (Model 9) memperoleh *Macro F1-Score* 26,15%, lebih tinggi 2,02 poin persentase dibandingkan model yang hanya menggunakan *education_mother* (Model 5) sebesar 24,13%. Pola serupa juga ditemukan pada arsitektur (64, 64), dengan selisih 1,81 poin persentase antara Model 25 (25,84%) dan Model 21 (24,03%). Namun, penggunaan kedua variabel pendidikan orang tua tetap menghasilkan performa yang paling optimal dan stabil, sedangkan penghapusan keduanya menyebabkan penurunan *Macro F1-Score* yang signifikan.

Berdasarkan perbandingan antararsitektur, Model 17 ditetapkan sebagai model terbaik dalam penelitian ini. Model tersebut menggunakan arsitektur MLP *hidden layer* (64, 64), menerapkan SMOTE dan standardisasi, serta memanfaatkan seluruh variabel *input*. Model 17 menghasilkan *accuracy* 28,66%, *precision* 29,50%, *recall* 37,90%, *Macro F1-Score* 26,30%, dan *AUC-ROC* pada rentang 0,63-0,79. Dibandingkan model terbaik arsitektur (6, 5), yaitu Model 1, peningkatan performa yang diperoleh relatif kecil dengan selisih *Macro F1-Score* hanya 0,50%. Hasil ini menunjukkan bahwa keterbatasan performa model lebih dipengaruhi oleh kualitas dan kelengkapan fitur daripada kapasitas arsitektur.

Secara keseluruhan, penelitian ini berhasil mencapai tujuan yang telah ditetapkan dan membuktikan bahwa metode MLP dapat digunakan untuk mengklasifikasikan kategori kemampuan intelektual anak. Meskipun demikian, performa yang dihasilkan masih tergolong rendah, menunjukkan bahwa variabel pendidikan orang tua, usia, dan jenis kelamin belum cukup untuk merepresentasikan kompleksitas kemampuan intelektual secara menyeluruh.

5.2 Saran

Berdasarkan hasil penelitian dan keterbatasan yang ditemukan, terdapat beberapa saran yang dapat dipertimbangkan untuk penelitian selanjutnya:

1. Penambahan variabel prediktor: menambahkan variabel lain yang memiliki korelasi lebih kuat terhadap kemampuan intelektual anak, seperti status sosial ekonomi keluarga, lingkungan tempat tinggal, pola asuh, prestasi akademik, hingga faktor psikologis, guna meningkatkan daya prediksi model.
2. Eksperimen algoritma pembanding: melakukan komparasi performa dengan metode *machine learning* atau *deep learning* lainnya, seperti *Random Forest, XGBoost,* atau *Support Vector Machine* (SVM), untuk menemukan arsitektur yang paling efektif dalam menangani karakteristik data kategori IQ.
3. Eksplorasi penanganan *data imbalance*: mencoba teknik penanganan data tidak seimbang alternatif selain SMOTE, seperti ADASYN, *Undersampling*, atau pendekatan *Cost-Sensitive Learning* untuk melihat pengaruhnya terhadap peningkatan nilai *F1-Score* pada kelas minoritas.
4. Optimasi *hyperparameter* yang lebih luas: melakukan pencarian parameter (*tuning*) secara lebih mendalam pada komponen *learning rate, batch size,* fungsi aktivasi, dan jumlah *epoch* untuk memaksimalisasi konvergensi model MLP.
5. Analisis interprestabilitas model: mengintegrasikan metode interpretabilitas seperti SHAP atau LIME untuk memberikan penjelasan yang lebih

mendalam mengenai bagaimana setiap variabel input (seperti pendidikan ayah atau ibu) secara spesifik memengaruhi keputusan klasifikasi model.

6. Perluasan cakupan dataset: menggunakan dataset yang lebih besar dan bervariasi dari berbagai populasi untuk meningkatkan kemampuan generalisasi model sehingga hasil penelitian tidak terbatas pada satu kelompok sampel tertentu saja.

Penelitian ini diharapkan dapat menjadi referensi dalam pengembangan *machine learning* di bidang psikologi dan pendidikan, dengan tetap mengedepankan aspek etika dalam penggunaan hasil klasifikasi intelektual anak.

DAFTAR PUSTAKA

Abedin, T., Xu, H., & Uddin, S. (2026). The impact of K selection in K-fold cross-validation on bias and variance in supervised learning models. *Scientific Reports*, *16*(1), 6084. https://doi.org/10.1038/s41598-026-37247-x

Cermakova, P., Chlapečka, A., Csajbók, Z., Andrýsková, L., Brázdil, M., & Marečková, K. (2023). Parental education, cognition and functional connectivity of the salience network. *Scientific Reports*, *13*(1), 2761. https://doi.org/10.1038/s41598-023-29508-w

Chase, D., & Chute, D. L. (2005). *Underlying factor structures of the Stanford-Binet Intelligence Scales—Fifth Edition* [Doctor of Philosophy, Drexel University]. https://doi.org/10.17918/etd-863

Freire, P., Manuylovich, E., Prilepsky, J. E., & Turitsyn, S. K. (2023). Artificial neural networks for photonic applications—from algorithms to implementation: Tutorial. *Advances in Optics and Photonics*, *15*(3), 739. https://doi.org/10.1364/AOP.484119

Ibnu Katsir, I. (n.d.). *Tafsir al-Qur'an al-'Adzim* (Jilid 7). Dar Thaybah lil-Nashr wa al-Tawzi'. https://www.muslim-library.com/arabic/.

Klein, M., & Kühhirt, M. (2023). *Parental education and children's cognitive development: A prospective approach*. PsyArXiv. https://doi.org/10.31234/osf.io/s5eyg

Kufel, J., Bargieł-Łączek, K., Kocot, S., Koźlik, M., Bartnikowska, W., Janik, M., Czogalik, Ł., Dudek, P., Magiera, M., Lis, A., Paszkiewicz, I., Nawrat, Z., Cebula, M., & Gruszczyńska, K. (2023). What Is Machine Learning, Artificial Neural Networks and Deep Learning?—Examples of Practical Applications in Medicine. *Diagnostics*, *13*(15), 2582. https://doi.org/10.3390/diagnostics13152582

LeBlanc, V., & Cox, M. A. A. (2017). Interpretation of the point-biserial correlation coefficient in the context of a school examination. *The Quantitative Methods for Psychology*, *13*(1), 46–56. https://doi.org/10.20982/tqmp.13.1.p046

Liu, Y. (2024). Multilayer perceptron-based literature reading preferences predict anxiety and depression in university students. *Frontiers in Psychology*, *15*, 1425471. https://doi.org/10.3389/fpsyg.2024.1425471

Madhiarasan, M., & Louzazni, M. (2022). Analysis of Artificial Neural Network: Architecture, Types, and Forecasting Applications. *Journal of Electrical*

and Computer Engineering, *2022*, 1–23. https://doi.org/10.1155/2022/5416722

McKinney, W. S., Nelson, M., Shaffer, R. C., Dominick, K. C., Erickson, C. A., & Schmitt, L. M. (2025). Brief Report: Differences Between Stanford-Binet Abbreviated and Full-Scale Estimates of IQ in Fragile X Syndrome Vary Across Development. *Journal of Autism and Developmental Disorders*. https://doi.org/10.1007/s10803-025-07062-w

Mfetoum, I. M., Ngoh, S. K., Molu, R. J. J., Nde Kenfack, B. F., Onguene, R., Naoussi, S. R. D., Tamba, J. G., Bajaj, M., & Berhanu, M. (2024). A multilayer perceptron neural network approach for optimizing solar irradiance forecasting in Central Africa with meteorological insights. *Scientific Reports*, *14*(1), 3572. https://doi.org/10.1038/s41598-024-54181-y

Mohammad, F., & Al Mansoor, K. M. (2024). SLA-MLP: Enhancing Sleep Stage Analysis from EEG Signals Using Multilayer Perceptron Networks. *Diagnostics*, *14*(23), 2657. https://doi.org/10.3390/diagnostics14232657

Olech, M., Jurek, P., Radtke, B. M., Sajewicz-Radtke, U., & Łada-Maśko, A. (2024). Intelligence Assessment of Children & Youth Benefiting from Psychological-Educational Support System in Poland. *Scientific Data*, *11*(1), 826. https://doi.org/10.1038/s41597-024-03663-9

Prechelt, L. (1998). Automatic early stopping using cross validation: Quantifying the criteria. *Neural Networks*, *11*(4), 761–767. https://doi.org/10.1016/S0893-6080(98)00010-0

Reinaldi, E. T., & Hidayat, R. (2021a). Stanford-Binet Intelligence Scale Form L-M Predictive Power on Academic Achievement. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, *10*(2), 133–141. https://doi.org/10.15408/jp3i.v10i2.20009

Reinaldi, E. T., & Hidayat, R. (2021b). Stanford-Binet Intelligence Scale Form L-M Predictive Power on Academic Achievement. *Jurnal Pengukuran Psikologi Dan Pendidikan Indonesia (JP3I)*, *10*(2), 133–141. https://doi.org/10.15408/jp3i.v10i2.20009

Sajewicz-Radtke, U., Jurek, P., Olech, M., Łada-Maśko, A. B., Jankowska, A. M., & Radtke, B. M. (2022). Heterogeneity of Cognitive Profiles in Children and Adolescents with Mild Intellectual Disability (MID). *International Journal of Environmental Research and Public Health*, *19*(12), 7230. https://doi.org/10.3390/ijerph19127230

Sajewicz-Radtke, U., Łada-Maśko, A., Olech, M., Jurek, P., Bieleninik, Ł., & Radtke, B. M. (2025). Association between parental education level and

intelligence quotient of children referred to the mental healthcare system: A cross-sectional study in Poland. *Scientific Reports*, *15*(1), 4142. https://doi.org/10.1038/s41598-025-88591-3

Sajewicz-Radtke, U., Łada-Maśko, A., Olech, M., Leoniak, K. J., & Radtke, B. M. (2025). Declarative memory profiles in children with non-specific intellectual disability: A cluster analysis approach. *Frontiers in Psychiatry*, *16*, 1581144. https://doi.org/10.3389/fpsyt.2025.1581144

Setiawan, D. W., Lapatta, N. T., Nugraha, D. W., & Lamasitudju, C. A. (n.d.). *Performance Comparison of Multilayer Perceptron (MLP) and Random Forest for Early Detection of Cardiovascular Disease*. *9*(6).

Seveso, A., Campagner, A., Ciucci, D., & Cabitza, F. (2020). Ordinal labels in machine learning: A user-centered approach to improve data validity in medical settings. *BMC Medical Informatics and Decision Making*, *20*(S5), 142. https://doi.org/10.1186/s12911-020-01152-8

Shatskaya, A., Tarasova, K., & Surilova, I. (2025). Bridging the gap: The impact of parental education and child leisure activities on cognitive and socioemotional skills in preschoolers. *Frontiers in Psychology*, *16*, 1568020. https://doi.org/10.3389/fpsyg.2025.1568020

Stephenson, K. G., Levine, A., Russell, N. C. C., Horack, J., & Butter, E. M. (2023). Measuring intelligence in Autism and ADHD: Measurement invariance of the -BINET 5TH EDITION and impact of subtest scatter on abbreviated IQ accuracy. *Autism Research*, *16*(12), 2350–2363. https://doi.org/10.1002/aur.3034

Suherman, A. F., Lisnaeni, P. P., Izqiatullailiyah, S. A., & Herlinawati, T. (n.d.). A Comparative Analysis of Spearman and Pearson Correlation Using SPSS. *A. F.*

Tafsir Ibn Kathir. (n.d.).

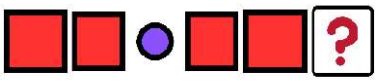
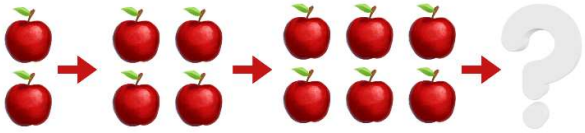
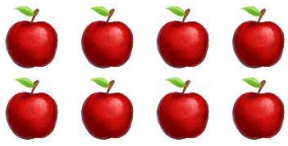
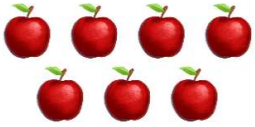
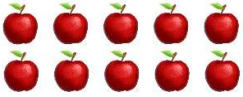
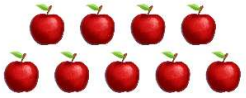
Wang, J. (2025). Training a convolutional neural network for exoplanet classification with transit photometry data. *Scientific Reports*, *15*(1), 15408. https://doi.org/10.1038/s41598-025-98935-8

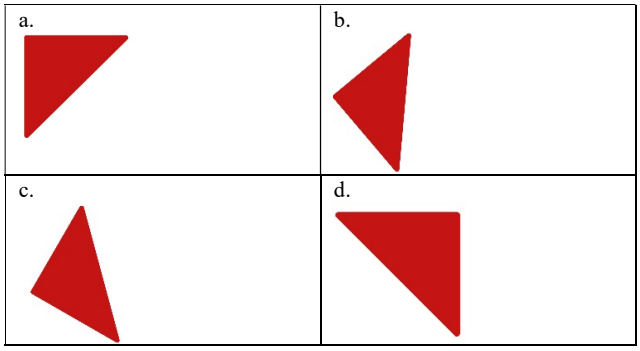
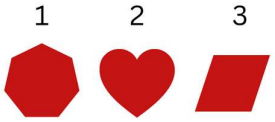
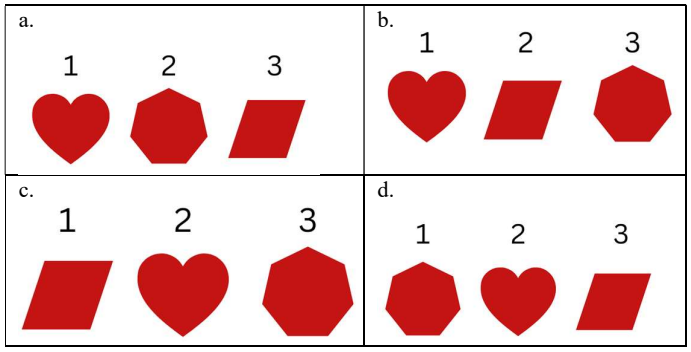
Yu, M., Wang, S., He, K., Teng, F., Deng, J., Guo, S., Yin, X., Lu, Q., & Gu, W. (2024). Predicting the complexity and mortality of polytrauma patients with machine learning models. *Scientific Reports*, *14*(1), 8302. https://doi.org/10.1038/s41598-024-58830-0

Zhou, Y., Aryal, S., & Bouadjenek, M. R. (2024). *Review for Handling Missing Data with special missing mechanism* (arXiv:2404.04905). arXiv. https://doi.org/10.48550/arXiv.2404.04905

LAMPIRAN-LAMPIRAN

Lampiran Contoh Soal Tes IQ Berdasarkan Kategorinya

Kategori Tes	Non Verbal	Verbal
Fluid reasoning (FR)	Perhatikan urutan gambar berikut !  Gambar manakah yang paling tepat untuk mengisi kotak tanda tanya tersebut?	"Burung berhubungan dengan terbang sebagaimana ikan berhubungan dengan …" a) Air b) Sirip c) Berenang d) Laut
Knowledge (KN)	Perhatikan gambar berikut !  Bagian sepeda yang digunakan untuk duduk ditunjukkan oleh nomor: a. 1 b. 2 c. 3 d. 4 e. 5	Apa arti kata "transparan" ?
Quantitative Reasoning (QR)	 Gambar manakah yang paling tepat untuk melanjutkan pola jumlah tersebut? a.  b.  c.  d. 	Jika 3 + 5 = 8, maka 8 + 4 = …?

<table>
<tr>
<td>Visual-Spatial Processing (VS)</td>
<td>Perhatikan segitiga di bawah ini,
 
Jika segitiga diputar 90°. Bentuk manakah yang menunjukkan hasil putaran tersebut?
 
a.
b.
c.
d.</td>
<td>“Ada sebuah bola berada di atas meja. Meja berada di samping kursi”,
Di manakah posisi bola terhadap kursi?</td>
</tr>
<tr>
<td>Working Memory (WM)</td>
<td>Peserta diperlihatkan urutan gambar berikut selama 5 detik:
 
1 2 3
Setelah gambar hilang, peserta diminta menyebutkan kembali urutan gambar tersebut.
 
a.
1 2 3
b.
1 2 3
c.
1 2 3
d.
1 2 3</td>
<td>Dengarkan deretan angka berikut:
4 – 7 – 2 – 9
Sekarang sebutkan angka tersebut dari belakang ke depan.</td>
</tr>
</table>