

**EKSPLORASI ALGORITMA *TREE BASED MODEL*
UNTUK KASUS TIPE KEPERIBADIAN DENGAN
*MYERS BRIGGS TYPE INDICATOR (MBTI)***



Disusun Oleh:

N a m a : Rani Asriya Effendy
NIM : 20523191

**PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA
FAKULTAS TEKNOLOGI INDUSTRI
UNIVERSITAS ISLAM INDONESIA**

2025

HALAMAN PENGESAHAN DOSEN PEMBIMBING


**EKSPLORASI ALGORITMA *TREE BASED MODEL*
UNTUK KASUS TIPE KEPRIBADIAN DENGAN
*MYERS BRIGGS TYPE INDICATOR (MBTI)***

TUGAS AKHIR



الجمعة الاستد الاندو

Yogyakarta, 14 Januari 2025
Pembimbing,



(Irving Vitra Papurungan, S.T., M.Sc., Ph.D.)

HALAMAN PENGESAHAN DOSEN PENGUJI

**EKSPLORASI ALGORITMA *TREE BASED MODEL*
UNTUK KASUS TIPE KEPERIBADIAN DENGAN
*MYERS BRIGGS TYPE INDICATOR (MBTI)***


TUGAS AKHIR

Telah dipertahankan di depan sidang penguji sebagai salah satu syarat untuk memperoleh gelar Sarjana Komputer dari Program Studi Informatika – Program Sarjana di Fakultas Teknologi Industri Universitas Islam Indonesia
Yogyakarta, 14 Januari 2025

Tim Penguji

Ketua

Irving Vitra Papatungan, S.T., M.Sc.,
Ph.D.


14/02/2025

Anggota 1

Erika Ramadhani, S.T., M.Eng.

Anggota 2

Kurniawan Dwi Irianto, S.T., M.Sc.

Mengetahui,

Ketua Program Studi Informatika – Program Sarjana

Fakultas Teknologi Industri

Universitas Islam Indonesia



(Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D.)

HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR

Yang bertanda tangan di bawah ini:

Nama : Rani Asriya Effendy

NIM : 20523191

Tugas akhir dengan judul:

**EKSPLORASI ALGORITMA *TREE BASED MODEL*
UNTUK KASUS TIPE KEPRIBADIAN DENGAN
*MYERS BRIGGS TYPE INDICATOR (MBTI)***

Menyatakan bahwa seluruh komponen dan isi dalam tugas akhir ini adalah hasil karya saya sendiri. Apabila di kemudian hari terbukti ada beberapa bagian dari karya ini adalah bukan hasil karya sendiri, tugas akhir yang diajukan sebagai hasil karya sendiri ini siap ditarik kembali dan siap menanggung risiko dan konsekuensi apapun.

Demikian surat pernyataan ini dibuat, semoga dapat dipergunakan sebagaimana mestinya.

Yogyakarta, 14 Januari 2025

A handwritten signature in black ink is written over a square QR code. The QR code has a pink and white pattern. The signature is written in a cursive style.

(Rani Asriya Effendy)

HALAMAN PERSEMBAHAN

Alhamdulillah Robbil 'Alamin puji serta syukur atas segala nikmat dan karunia yang telah Allah SWT berikan kepada saya, keluarga, saudara-saudara, dan sahabat-sahabat. Shalawat beserta salam kepada Nabi Muhammad SAW sebagai pemberi syafaat kepada seluruh umat manusia.

Tugas akhir ini saya persembahkan untuk kedua orang tua saya Bapak Ir Ahmad Farid Effendy, S.T., M.M dan Ibu Eni Srihastuty S.T., yang selalu mendoakan dan mendukung saya dalam keadaan apapun. Terima kasih telah mendidik dan selalu mengingatkan saya untuk menjadi pribadi yang lebih baik setiap harinya. Terimakasih untuk Bapak Irving Vitra Paputungan, S.T., M.Sc., Ph.D. selaku dosen pembimbing yang telah meluangkan waktu untuk membimbing dan memberi dukungan kepada saya hingga akhir, agar saya dapat menyelesaikan tugas akhir ini hingga selesai. Selama pengerjaan tugas akhir ini tentu terdapat kesalahan dari perkataan saya maupun perbuatan. Saya memohon maaf jika ada kesalahan dari perkataan atau perbuatan saya yang tidak berkenan di hati bapak. Semoga Allah SWT. membalas semua kerja keras dan kebaikan bapak. *Aamiin Ya Rabbal Alamin.*

HALAMAN MOTO

إِن سَلَكَ طَرِيقًا يَبْتَغِي فِيهِ عِلْمًا سَهَّلَ اللَّهُ لَهُ طَرِيقًا إِلَى الْجَنَّةِ

"Barangsiapa menempuh jalan untuk mencari ilmu, maka Allah akan memudahkan baginya jalan ke surga" (HR. Muslim)

"The only way to do great work is to love what you do"

- Steve Jobs

KATA PENGANTAR

Bismillahirrahmanirrahim,

Assalamualaikum warahmatullahi wabarakatuh.

Dengan nama Allah yang Maha Pengasih lagi Maha Penyayang, saya ucapkan puji dan syukur atas kehadiran Allah SWT. yang telah memberikan penulis kesempatan dan kekuatan untuk menyelesaikan laporan tugas akhir yang berjudul “Eksplorasi Algoritma *Tree Based Model* untuk Kasus tipe kepribadian dengan *Myers-Briggs Type Indicator (MBTI)*.” Tugas akhir ini merupakan salah satu syarat yang harus dipenuhi dalam menyelesaikan pendidikan dan meraih gelar pada jenjang sarjana Jurusan Informatika, Fakultas Teknologi Industri, Universitas Islam Indonesia. Pembuatan tugas akhir ini tidak akan berjalan dengan baik tanpa adanya dukungan dan bimbingan dari berbagai pihak. Maka dari itu saya ingin mengucapkan terima kasih sebesar-besarnya kepada:

1. Allah SWT. yang telah memberikan Rahmat dan karunia kepada penulis agar dapat menyelesaikan laporan tugas akhir ini,
2. Bapak Dr. Raden Teduh Dirgahayu, S.T., M.Sc. selaku Ketua Jurusan Informatika Fakultas Teknologi Industri Universitas Islam Indonesia,
3. Bapak Dhomas Hatta Fudholi, S.T., M.Eng., Ph.D. selaku Ketua Program Studi Informatika Program Sarjana Fakultas Teknologi Industri Universitas Islam Indonesia,
4. Bapak Irving Vitra Papatungan, S.T., M.Sc., Ph.D. selaku dosen pembimbing yang telah memberi motivasi dan bimbingan dengan sabar dalam penyelesaian tugas akhir ini hingga terselesaikan,
5. Seluruh Dosen Pengajar Jurusan Informatika atas ilmu pengetahuan yang diberikan kepada penulis selama menjadi mahasiswa di Universitas Islam Indonesia,
6. Kedua Orang Tua penulis, Bapak Ir Ahmad Farid Effendy, S.T., M.M dan Ibu Eni Srihastuty S.T., atas dukungan, doa, dan bimbingan yang selalu diberikan kepada penulis,
7. Seluruh teman dan sahabat dari penulis yang selalu memberikan dukungan agar

penulis dapat menyelesaikan tugas akhir ini dengan baik.

Yogyakarta, 14 Januari 2025

A handwritten signature in black ink, appearing to read 'Rani Asriya Effendy', with a long horizontal stroke extending to the right.

(Rani Asriya Effendy)

SARI

“Eksplorasi Algoritma *Tree Based Model* untuk Kasus tipe kepribadian dengan *Myers-Briggs Type Indicator* (MBTI).” Penelitian ini dilakukan karena pentingnya memahami tipe kepribadian individu dalam berbagai aspek kehidupan, seperti pendidikan, pekerjaan, dan hubungan sosial.

Myers-Briggs Type Indicator (MBTI) merupakan salah satu alat ukur yang paling umum digunakan untuk mengklasifikasikan tipe kepribadian individu. Namun, penggunaan MBTI masih memiliki beberapa keterbatasan, seperti kurangnya akurasi dan kesulitan dalam memilih parameter yang tepat. Dalam beberapa tahun terakhir, penggunaan algoritma *machine learning* telah meningkat pesat dalam berbagai bidang, termasuk psikologi. Algoritma *Tree Based Model* merupakan salah satu jenis algoritma yang paling umum digunakan dalam klasifikasi data. Oleh karena itu, penelitian ini bertujuan untuk menganalisis dan mengeksplorasi performa berbagai model berbasis pohon keputusan dalam mengklasifikasikan tipe kepribadian MBTI. Penelitian ini menggunakan dataset yang dikumpulkan menggunakan *google form* yang diisi bersamaan dengan pengisian tes kepribadian pada website *16Personalities*, responden akan mengisi jawaban yang sama persis dengan tes kepribadian yang sudah dilakukan. Dan sampel yang terkumpul berjumlah 174 sampel. Model yang akan digunakan untuk eksplorasi dalam penelitian ini meliputi *Decision Tree*, *Random Forest*, *Extra Trees*, *Gradient Boosting*, *XGBoost*, dan *Ada Boost*. Tahapan pada penelitian ini mencakup *Pre-Processing Data*, Analisis data, *Modelling* dan *Testing*. Hasil penelitian Model terbaik berdasarkan prediksi yang dilakukan dengan target semua dimensi MBTI dan nilai perbandingan antara *train set* dan *test set* sebesar 8:2 adalah model *Extra Trees* dengan nilai akurasi 0.728, kemudian model tertinggi setelah *Extra Trees* ada *Random Forest* dengan nilai akurasi 0.7178, kemudian model *XGBoost* dengan nilai akurasi 0.7142, kemudian model *AdaBoost* dengan nilai akurasi 0.703, dilanjut dengan model *Gradient Boosting* dengan nilai akurasi 0.70, dan terakhir model yang memiliki nilai paling rendah ada *Decision Tree* dengan nilai 0.603. Penelitian ini diharapkan dapat memberikan kontribusi pada pengembangan algoritma pembelajaran mesin dalam psikologi, serta dapat membantu dalam pengembangan sistem yang lebih akurat dan efektif dalam mengklasifikasikan tipe kepribadian individu. Selain itu, penelitian ini juga dapat membantu dalam memahami lebih lanjut tentang tipe kepribadian individu dan bagaimana algoritma pembelajaran mesin dapat digunakan untuk meningkatkan pemahaman tersebut.

Kata kunci: MBTI, *Tree Based Model*, klasifikasi kepribadian, *machine learning*, *16Personalities*

DAFTAR ISI

PROGRAM STUDI INFORMATIKA – PROGRAM SARJANA	1
HALAMAN PENGESAHAN DOSEN PEMBIMBING	ii
HALAMAN PENGESAHAN DOSEN PENGUJI	iii
HALAMAN PERNYATAAN KEASLIAN TUGAS AKHIR	iv
HALAMAN PERSEMBAHAN	v
HALAMAN MOTO	vi
KATA PENGANTAR	vii
SARI	ix
DAFTAR ISI	x
DAFTAR TABEL	xii
DAFTAR GAMBAR	xiii
BAB I PENDAHULUAN	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah dan Pertanyaan Penelitian	3
1.3 Batasan Masalah dan Lingkup Penelitian	4
1.4 Tujuan Penelitian dan Manfaat penelitian	4
1.5 Metodologi	4
1.6 Sistematika	5
BAB II LANDASAN TEORI	6
2.1 Kepribadian	6
2.1.1 Definisi Kepribadian	6
2.2 Faktor Kepribadian	6
2.3 Teori Kepribadian	7
2.4 Alat ukur Kepribadian	8
2.4.1 Hippocrates	8
2.4.2 The Big Five	9
2.4.3 Tipe Enneagram	10
2.4.4 The Myers-Briggs Type Indicator (MBTI)	11
2.5 Aplikasi 16 Personalities	14
2.5.1 Koefisiensi Internal	15
2.5.2 Keandalan Uji-Ulang	15
2.5.3 Validitas Diskriminan	16
2.6 Kajian Pustaka	17
2.6.1 Algoritma Tree Based Model	17
2.6.2 Penelitian Terdahulu	21
BAB III METODE PENELITIAN	25
3.1 Desain Penelitian	25
3.2 Variabel Penelitian	25
3.3 Subjek Penelitian	27
3.4 Pengumpulan Data	27
3.4.1 Pengisian Kuesioner	27
3.5 Pre-Processing Data	27
3.5.1 Penanganan Missing Value	27
3.5.2 Split Label	28

3.5.3	Analisis Statistic Summary.....	28
3.5.4	Response Distribution of Each Answers.....	29
3.5.5	Personality Distribution.....	29
3.5.6	Correlation values Of Each Question Respones to Personalities.....	30
3.5.7	Another Correlation Analysis	36
3.5.8	Modelling and Testing Modelling.....	37
3.5.9	Testing.....	38
3.5.10	Explore More	39
BAB IV	HASIL DAN PEMBAHASAN.....	42
4.1	Analisis Pengumpulan Data.....	42
4.1.1	Analisis Informasi Data.....	46
4.2	Analisis Pre-Processing Data.....	47
4.2.1	Analisis Hasil Penanganan <i>Missing Value</i>	47
4.3	Analisis Hasil <i>Split Label</i>	48
4.4	Analisis Hasil Statistic Summary	50
4.5	Analisis hasil Response Distribution	51
4.6	Analisis Hasil Personality Distribution.....	51
4.7	Analisis hasil Correlation Value.....	54
4.7.1	Correlation Value dimensi E-I.....	55
4.7.2	Correlation Value dimensi S-N.....	56
4.7.3	Correlation Value dimensi T-F.....	58
4.7.4	Correlation Value dimensi J-P.....	60
4.8	Analisis Hasil Another Correlation.....	61
4.9	Analisis Hasil <i>Modelling</i>	63
4.9.1	Analisis Hasil <i>Modelling</i> dimensi E-I.....	64
4.9.2	Analisis Hasil dimensi S-N.....	67
4.9.3	Analisis Hasil dimensi T-F	70
4.9.4	Analisis Hasil dimensi J-P.....	73
4.10	Analisis Hasil <i>Testing</i>	75
4.10.1	Analisis hasil High Score Correlation.....	77
BAB V	KESIMPULAN DAN SARAN	79
5.1	Kesimpulan	79
5.2	Saran	79
	DAFTAR PUSTAKA	81

DAFTAR TABEL

Tabel 2. 1 Nilai Cronbach Alpha	15
Tabel 2. 2 Keandalan Uji Ulang.....	16
Tabel 2. 3 Validitas dan Diskriminan.....	16
Tabel 2. 4 Penelitian Terdahulu	21
Tabel 3. 1 Split Label.....	28
Tabel 3. 2 Statistic Summary	28
Tabel 3. 3 Response Distribution of Each Answers	29
Tabel 3. 4 Tampilan tabel dataset teratas	31
Tabel 3. 5 Hasil Correlation values.....	31
Tabel 3. 6 Blue Print MBTI	34
Tabel 3. 7 Jumlah korelasi Pertanyaan untuk setiap dimensi.....	36
Tabel 3. 8 Jumlah korelasi pertanyaan untuk setiap dimensi	37
Tabel 3. 9 Dimensi E-1	37
Tabel 3. 10 Dimensi S-N.....	38
Tabel 3. 11 Dimensi T-F.....	38
Tabel 3. 12 Dimensi J-P.....	38
Tabel 3. 13 Dimensi E-I.....	40
Tabel 3. 14 Dimensi S-N.....	40
Tabel 3. 15 Dimensi T-F.....	40
Tabel 3. 16 Dimensi J-P.....	40
Tabel 4. 1 Jawaban pertanyaan	46
Tabel 4. 2 Tabel Hasil Informasi Data	46
Tabel 4. 3 Data Lima Teratas dari Dataset.....	50
Tabel 4. 4 Hasil Statistic Summary	50
Tabel 4. 5 Nilai Pertanyaan yang Memiliki Korelasi Tinggi	62
Tabel 4. 6 Nilai Pertanyaan yang Memiliki Sedikit Korelasi.....	63
Tabel 4. 7 Hasil <i>Modelling</i> Dimensi E-I.....	64
Tabel 4. 8 Hasil <i>Modelling</i> Dimensi S-N.....	67
Tabel 4. 9 Hasil <i>Modelling</i> Dimensi T-F	70
Tabel 4. 10 Hasil <i>Modelling</i> Dimensi J-P.....	73

DAFTAR GAMBAR

Gambar 3. 1 Diagram Penelitian	26
Gambar 3. 2 (A) E-I (B) S-N.....	30
Gambar 3. 3 (A) T-F (B) J-P.....	30
Gambar 3. 4 Perubahan kepribadian menjadi boolean.....	31
Gambar 3. 5 Visualisasi kepribadian E-I.....	32
Gambar 3. 6 Visualisasi kepribadian S-N	32
Gambar 3. 7 Visualisasi kepribadian T-F.....	33
Gambar 3. 8 Visualisasi kepribadian J-P.....	33
Gambar 3. 9 Rumus Custom Accuracy Score	39
Gambar 3. 10 Hasil Testing untuk semua Dimensi.....	39
Gambar 3. 11 Pertanyaan yang memiliki korelasi tinggi untuk setiap dimensi	39
Gambar 3. 12 Hasil dari Testing.....	41
Gambar 4. 1 Tampilan pertanyaan pada Website 16Personalities.....	42
Gambar 4. 2 Tampilan pertanyaan teratas pada google form.....	42
Gambar 4. 3 Jumlah Responden.....	47
Gambar 4. 4 Jumlah Angka Null.....	47
Gambar 4. 5 Kode untuk mengisi nilai null dengan nilai modus	48
Gambar 4. 6 Jumlah Angka Null.....	48
Gambar 4. 7 Kode perintah untuk memisah komponen pertama dan kedua	48
Gambar 4. 8 Hasil Split Label.....	49
Gambar 4. 9 Kode perintah untuk memisahkan label menjadi 4 bagian	49
Gambar 4. 10 <i>Histogram</i> Setiap Pertanyaan	51
Gambar 4. 11 <i>Pie Chart</i> untuk Dimensi E-I.....	51
Gambar 4. 12 Jumlah Responden untuk Dimensi E-I.....	52
Gambar 4. 13 <i>Pie Chart</i> untuk Dimensi S-N	52
Gambar 4. 14 Jumlah Responden untuk Dimensi S-N.....	52
Gambar 4. 15 <i>Pie Chart</i> untuk Dimensi T-F.....	53
Gambar 4. 16 Jumlah Responden untuk Dimensi T-F	53
Gambar 4. 17 <i>Pie Chart</i> untuk Dimensi J-P.....	53
Gambar 4. 18 Jumlah Responden untuk Dimensi J-P.....	54
Gambar 4. 19 Rumus Pearson Correlation	54
Gambar 4. 20 Kode perintah untuk menghitung korelasi menggunakan Pearson.....	54
Gambar 4. 21 Visualisasi kepribadian E-I.....	55
Gambar 4. 22 Visualisasi Kepribadian S-N 1	2
Gambar 4. 23 Visualisasi Kepribadian T-F.....	58
Gambar 4. 24 Visualisasi Kepribadian J-P.....	60
Gambar 4. 25 Rumus <i>Custom Accuracy Score</i>	76
Gambar 4. 26 Hasil Testing untuk Semua Dimensi	76
Gambar 4. 27 Kode perintah menampilkan pertanyaan yang memiliki korelasi pada masing	

masing dimensi	77
Gambar 4. 28 Hasil Pertanyaan yang Memiliki Korelasi Tinggi Terhadap Setiap Dimensi	77
Gambar 4. 29 Hasil Testing untuk Pertanyaan yang Memiliki Korelasi Tinggi	77

BAB I

PENDAHULUAN

1.1 Latar Belakang

Di era digital yang terus berkembang, memahami kepribadian seseorang menjadi semakin penting, terutama dalam bidang pendidikan, pekerjaan, dan hubungan sosial. Kepribadian seseorang yang meliputi sikap, emosi, ekspresi, tempramen, dan perilakunya memegang peran penting dalam menentukan keberhasilan seseorang dalam berbagai aspek kehidupan.

Dalam laporannya WHO (2023) terjadi peningkatan sebesar 15% untuk kasus gangguan kesehatan mental yang terjadi sejak pandemi, yang mana kesehatan mental ini memiliki kaitan erat dengan masalah kepribadian seseorang. Tekanan sosial dan pengaruh media sosial juga termasuk alasan yang signifikan terhadap perkembangan kepribadian seseorang, terutama di kalangan generasi muda.

Pada lingkungan pekerjaan, Gallup (2024) juga melaporkan adanya 60% pekerja yang mengalami *burnout* di karenakan ketidaksesuaian antara kepribadian mereka dengan tuntutan pekerjaan. Situasi ini juga menjadi lebih sulit dengan meningkatnya sistem kerja *hybrid* dan *remote* dimana ini membutuhkan adaptasi kepribadian yang berbeda. Konflik interpersonal yang diakibatkan perbedaan kepribadian juga bisa menjadi tantangan tersendiri dalam kerjasama tim.

Kepribadian yang bersifat dinamis dapat disebut sebagai dinamika kepribadian. Karena sifatnya yang dinamis, setiap individu memiliki kepribadian yang dapat berubah dan berkembang sesuai dengan pengaruh lingkungan. Untuk memahami kepribadian seseorang, dapat dilakukan berbagai tes kepribadian, seperti *HTP (House Tree Person)* atau *DAP (Draw A Person)*. Namun, seiring dengan perkembangan, alat tes untuk mengukur kepribadian juga mengalami perubahan, salah satunya adalah *Myers-Briggs Type Indicator (MBTI)*.

Myers-Briggs Type Indicator (MBTI) merupakan salah satu alat ukur kepribadian yang sedang populer dan banyak digunakan di seluruh dunia. Menurut *The Myers Briggs Foundation* (2022), sudah ada lebih dari 2 juta orang yang menggunakan MBTI setiap

tahunnya untuk memahami kepribadian mereka, dan di tahun berikutnya meningkat sebesar 15%. MBTI mengklasifikasikan individu ke dalam 16 tipe kepribadian. MBTI membedakan karakter seseorang dari yang cenderung pendiam hingga yang sangat aktif dalam berkomunikasi dan bersosialisasi. Terdapat empat dimensi utama dalam MBTI, yaitu:

- a. *Ekstrovert-Introvert (E-I)*,
- b. *Sensing-iNntuition (S-N)*,
- c. *Thinking-Feeling (T-F)*,
- d. *Judging-Perceiving (J-P)*.

Pengenalan tipe kepribadian MBTI umumnya dilakukan melalui kuesioner yang mengukur kecenderungan individu pada setiap dimensi tersebut. Namun, proses ini dapat memakan waktu dan bersifat subjektif. Dengan kemajuan teknologi dan ilmu data, klasifikasi tipe kepribadian MBTI kini bisa dilakukan secara otomatis menggunakan algoritma *machine learning*. Salah satu *machine learning* yang dapat digunakan untuk klasifikasi MBTI adalah *Tree Based Model*.

Tree Based Model merupakan sekumpulan algoritma yang memanfaatkan struktur pohon keputusan untuk memprediksi atau mengklasifikasikan dengan membagi data ke dalam cabang- cabang keputusan berdasarkan pilihan tertentu menggunakan metrik seperti *Gini Index* atau *Entropy* yang bertujuan memisahkan data menjadi subset yang lebih homogen. Model ini juga memiliki kemampuan dalam penanganan data yang mampu menangani data numerik dan kategorikal, dan tidak memerlukan normalisasi data.

Tree Based Model telah banyak digunakan pada klasifikasi MBTI. Model ini cocok untuk kasus MBTI karena kemampuannya dalam menangani data *multi-class*, memiliki kemampuan interpretasi yang baik untuk memahami faktor-faktor yang mempengaruhi klasifikasi kepribadian, serta kemampuannya dalam menangani data teks yang telah melalui proses *vectorization*. Selain itu, *Tree Based Model* juga memiliki keunggulan yang signifikan dibandingkan algoritma lain untuk kasus MBTI. Dibandingkan dengan SVM, *Neural Networks*, *Naïve Bayes*, dan KNN, *Tree Based Model* menunjukkan keunggulan dalam menangani *multi-class classification*, efisiensi komputasi, interpretabilitas hasil, dan kemampuan menangani data kompleks tanpa *preprocessing*

yang ekstensif. Kombinasi antara performa yang baik, interpretabilitas yang tinggi, dan kemudahan implementasi menjadikan *Tree Based Model* sebagai pilihan yang optimal untuk klasifikasi tipe kepribadian MBTI.

Beberapa penelitian telah memanfaatkan berbagai algoritma *Tree-Based* untuk mengklasifikasikan tipe kepribadian MBTI. Sebagai contoh, penelitian yang dilakukan oleh Nisha et al. (2022) Dalam penelitiannya yang berjudul “*A Comparative Analysis of Machine learning Approaches in Personality Prediction Using MBTI.*” Sistem yang diusulkan mencapai hasil yang memuaskan dalam memprediksi kepribadian dengan menggunakan tweet. Dalam tiga pengklasifikasi, kinerja pengklasifikasi *Naive Bayes* lebih buruk sementara Linear SVM dan *XGBoost* memberikan hasil yang memuaskan. Mushtaq et al. (2020) Dalam penelitiannya yang berjudul “*Predicting MBTI Personality type with K-means Clustering and Gradient Boosting.*” Hasil penelitian menunjukkan bahwa kombinasi K-means clustering dan *Gradient Boosting* memberikan akurasi yang lebih tinggi dibandingkan dengan metode konvensional.

Di sisi lain, penelitian yang berjudul “*Social Media User Personality Classification Based on How User Live and Make Decisions*” oleh Faisal Amri et al., (2022) menggunakan algoritma *Random Forest* untuk melakukan klasifikasi kepribadian berdasarkan cara pengguna media sosial menjalani hidup dan mengambil keputusan. Hasil penelitian ini menunjukkan tingkat akurasi yang tinggi. Meskipun penelitian ini telah mencapai hasil akurasi yang baik dalam klasifikasi kepribadian, masih ada peluang untuk menerapkan algoritma lain dalam proses klasifikasi tersebut. Dalam penelitian ini, penulis akan mengeksplorasi beberapa algoritma turunan *Tree Based Model* dalam kasus tipe kepribadian MBTI, mengevaluasi kinerjanya pada kumpulan data yang berbeda, dan menganalisis faktor-faktor yang berkontribusi terhadap tingkat akurasi klasifikasi. Hasil penelitian ini diharapkan dapat memberikan wawasan baru mengenai penggunaan *Machine learning* untuk memahami kepribadian dan memberikan rekomendasi praktis untuk mengembangkan aplikasi pengenalan kepribadian yang efektif.

1.2 Rumusan Masalah dan Pertanyaan Penelitian

Berdasarkan latar belakang yang telah diuraikan maka dapat di rumuskan masalah

sebagai berikut:

- a. Bagaimana algoritma *Tree Based Model* dapat diterapkan untuk eksplorasi kasus tipe kepribadian MBTI?
- b. Apa saja algoritma *Tree Based Model* yang akurat untuk kasus tipe kepribadian MBTI?

1.3 Batasan Masalah dan Lingkup Penelitian

- a. Penelitian ini dilakukan menggunakan *machine learning* dengan jenis *supervised learning* yang dilakukan untuk mengeksplorasi masing-masing algoritma dalam kasus tipe kepribadian MBTI.
- b. Penelitian ini berfokus pada penerapan *Tree Based Model* untuk eksplorasi kasus tipe kepribadian MBTI.

1.4 Tujuan Penelitian dan Manfaat penelitian

Tujuan dari penelitian ini adalah sebagai berikut:

- a. Menganalisis serta mengimplementasikan algoritma *Tree Based* pada klasifikasi kepribadian MBTI
- b. Mengetahui model yang memberikan hasil prediksi yang paling akurat pada klasifikasi kepribadian MBTI

Manfaat dari penelitian ini adalah sebagai berikut:

- a. Dapat menerapkan ilmu tentang *Machine learning* terutama *Tree Based Model* yang di dapat selama masa perkuliahan.
- b. Dapat mengetahui metode yang lebih baik untuk digunakan pada klasifikasi tipe kepribadian MBTI.

1.5 Metodologi

Metode penelitian yang digunakan dalam pembuatan tes klasifikasi kepribadian dengan algoritma *Tree Based Model* berdasarkan tipe kepribadian MBTI terbagi menjadi beberapa tahapan antara lain:

- a. Analisis kebutuhan

Pada proses ini yang dilakukan adalah pengumpulan data dengan menjawab beberapa pertanyaan melalui *google form* dan website *16Personalities* secara bersamaan.
- b. Implementasi

Kemudian setelah melakukan analisis dan perancangan dilanjutkan dengan tahap implementasi. Pada tahap ini dilakukan olah data, mulai dari *pre-processing* data seperti *handling missing value* dan *split label*, korelasi dan analisis data seperti *statistic summary*, *response distribution of each answer*, *personality distribution*, *correlation values of each question responses to personalities*, dan *another correlation analysis*.

c. Pengujian

Dan tahapan yang terakhir adalah pengujian. Pada tahap ini dilakukan *modelling* dan *testing* dengan algoritma yang telah di tentukan.

1.6 Sistematika

Dalam laporan tugas akhir ini disusun secara sistematis dalam 5 bab diantaranya sebagai berikut:

BAB I PENDAHULUAN

Berisi latar belakang masalah yang mendasari dilakukannya penelitian ini, rumusan masalah, manfaat penelitian, tujuan penelitian, batasan masalah, metodologi penelitian, dan sistematika penulisan.

BAB II LANDASAN TEORI

Berisi penjelasan tentang masalah yang mendasari dilakukannya penelitian dan metode yang digunakan. Pada bab ini juga menjelaskan tentang penjelasan dan perbandingan dengan penelitian terdahulu.

BAB III METODOLOGI PENELITIAN

Berisi langkah-langkah yang digunakan dalam menyelesaikan penelitian, seperti pengumpulan data, *pre-processing data*, korelasi data, analisis data, *modelling*

BAB IV HASIL DAN PEMBAHASAN

Berisi hasil dari implementasi serta pembahasan yang telah dilakukan, mulai dari perancangan hingga pengujian model.

BAB V KESIMPULAN DAN SARAN

Berisi kesimpulan dari penelitian yang telah dilakukan serta berisi saran untuk penelitian selanjutnya.

BAB II

LANDASAN TEORI

2.1 Kepribadian

2.1.1 Definisi Kepribadian

Kepribadian adalah pola perilaku, pemikiran, dan emosi yang relatif stabil dan khas yang dimiliki oleh setiap individu. Kepribadian ini membedakan satu individu dengan lainnya dan mempengaruhi cara mereka beradaptasi dengan lingkungan sekitar. Beberapa ahli memberikan definisi mengenai kepribadian. Menurut Allport (1961), kepribadian adalah organisasi dinamis dalam sistem psikofisik individu yang menentukan penyesuaian uniknya terhadap lingkungan. Sementara itu, menurut Pervin & John (2001), kepribadian adalah karakteristik pola pikir, perasaan, dan perilaku yang membentuk kecenderungan individu untuk merespons secara konsisten terhadap berbagai situasi.

Woodworth juga menyampaikan pandangannya, yang menyatakan bahwa tindakan seseorang dipengaruhi oleh kepribadiannya. Menurutnya, kepribadian bukanlah substansi, melainkan gejala dan gaya hidup. Kepribadian tidak tercermin dalam jenis atau aktivitas tertentu, seperti berbicara, mengingat, atau berpikir, tetapi setiap individu dapat mengekspresikan kepribadiannya melalui aktivitas-aktivitas tersebut.

2.2 Faktor Kepribadian

Menurut Pervin & John (2001) kepribadian terbentuk dari interaksi antara faktor biologis (genetik, otak, dan sistem saraf), faktor lingkungan (keluarga, budaya, dan pengalaman hidup), serta faktor situasional. Di antaranya:

- a. Faktor genetik dan faktor biologis seperti tempramen bawaan dapat mempengaruhi kepribadian seseorang. Seperti kecenderungan untuk menjadi *Ekstrovert* atau *Introvert*. Hal ini juga dapat membatasi perkembangan kepribadian serta berkontribusi dalam keunikan individu dalam kepribadian mereka.
- b. Faktor lingkungan. Lingkungan dimana dia di besarkan, termasuk keluarga, budaya, dan pengalaman hidup akan sangat berpengaruh dalam pembentukan kepribadian.

- c. Faktor sosial. Bagaimana interaksi sosialnya, bagaimana hubungannya dengan orang lain, dan perannya dalam masyarakat dapat membentuk pola perilaku dan kepribadian seseorang.
- d. Faktor psikologis. Proses kognitif, emosi, motivasi, serta mekanisme pertahanan diri dapat mempengaruhi cara seseorang berfikir, merasakan, dan berperilaku.
- e. Faktor Situasional. Dalam situasi atau peristiwa tertentu yang dialami seseorang, seperti trauma, keberhasilan, atau bahkan kegagalan dapat membentuk atau mengubah kepribadian seseorang.

2.3 Teori Kepribadian

Sedangkan teori kepribadian merupakan sebuah pendekatan ilmiah yang bertujuan untuk menjelaskan bagaimana kepribadian seseorang terbentuk, terstruktur, dan berfungsi. Dimana teori kepribadian ini juga berupaya untuk memahami pola pikiran, perasaan, dan perilaku yang relatif stabil dan khas pada setiap individu. Berikut ini adalah beberapa penjelasan tentang teori kepribadian:

- a. Teori Psikodinamika (Sigmund Freud). Dalam teori ini menyatakan bahwa kepribadian dibentuk oleh interaksi antara dorongan naluriah, ego(prinsip realitas), dan superego (hati nurani). Konflik dari ketiga aspek ini yang dapat menyebabkan timbulnya mekanisme pertahanan diri yang menentukan pola perilaku seseorang.
- b. Teori Analitis (Carl Jung). Disini Jung memperkenalkan konsep kepribadian yang terdiri dari kesadaran dan ketidak sadaran kolektif. Menurutnya kepribadian manusia dapat dibagi menjadi dua kecenderungan memiliki kepribadian ekstrim melihat reaksi individu berdasarkan pengalamannya. Yang pertama ada kecenderungan manusia *Introvert*, yaitu manusia yang menarik diri dan larut dalam pengalaman batinnya sendiri. Dimana orang seperti itu biasanya tertutup, tidak memperhatikan orang sekitar, dan cenderung diam. Yang kedua itu ada *Ekstrovert*, dimana mereka termasuk pribadi yang suka membuka diri, berbicara dengan orang lain.
- c. Teori Behavioristik. Teori ini menekankan bahwa kepribadian dibentuk melalui proses belajar dari lingkungan. Kepribadian juga merupakan pola perilaku yang dipelajari melalui prinsip pengkondisian (klasik dan operan).
- d. Teori Humanistik (Abraham Maslow dan Carl Rogers). Pada teori ini menekankan pada

konsep aktualisasi, penerimaan, serta potensi positif pada diri manusia. Dimana kepribadian dibentuk oleh pengalaman yang subjektif serta persepsi individu tentang dirinya sendiri.

- e. Teori Trait (Raymond Cattell dan Hans Eysenck). Dimana teori ini mengklasifikasikan kepribadian ke dalam sifat-sifat atau trait yang relatif stabil. Misalnya, teori Lima Faktor Besar (Big Five) oleh Goldberg yang mencakup ekstroversion, agreeableness, conscientiousness, neuroticism, dan openness to experience.
- f. Teori Kognitif Sosial (Albert Bandura). Pada teori ini menekankan bahwa kepribadian dibentuk melalui interaksi timbal balik antara lingkungan, perilaku, dan proses kognitif seperti pemikiran, harapan, dan keyakinan individu.
- g. Teori Bioekologi (Bronfenbrenner). Dimana teori ini melihat bahwa kepribadian dibentuk oleh interaksi antara individu dengan lingkungan mikro (keluarga, sekolah) dan lingkungan makro (budaya, masyarakat) secara timbal balik.

2.4 Alat ukur Kepribadian

Para ahli memberikan pendapatnya tentang alat ukur kepribadian seseorang. Diantaranya:

2.4.1 Hippocrates

Sang “Bapak Ilmu Kedokteran” (460-377 SM) ini mengatakan bahwa manusia dapat dibagi kedalam empat golongan berdasarkan tempramennya.

- a. Sanguinis (*sanguine*) : Tipe ini dikaitkan dengan dominasi cairan darah (*sanguis*). Dimana orang dengan tipe ini merupakan orang yang ceria, mudah bergaul, penuh semangat, serta suka berbicara.
- b. Korelis (*choleric*) : Tipe ini dikaitkan dengan dominasi empedu kuning (*chole*). Dimana orang dengan tipe ini cenderung ambisius, tegas, mudah marah, dan sangat impulsif.
- c. Melankolis (*melancholic*) : Tipe ini dikaitkan dengan dominasi empedu hitam (*melaina chole*). Dimana orang dengan tipe melankolis cenderung pemurung, pesimis, perfeksionis, bahkan introspektif.
- d. Flegmatis (*phlegmatic*) : Tipe ini dikaitkan dengan dominasi lendir (*phlegma*). Dimana orang dengan tipe ini merupakan orang yang tenang, sabar, tidak mudah terganggu,

serta kurang responsif.

Kemudian tipologi ini disempurnakan oleh Galenus. Dimana Galenus mengatakan bahwa keempat cairan tersebut terdapat di dalam tubuh dengan proporsi yang berbeda beda setiap orang, dimana jika salah satu cairan yang lebih dominan dari cairan lainnya, maka cairan tersebut yang dapat membentuk kepribadian seseorang.

2.4.2 The Big Five

Tipe ini didefinisikan oleh Digman (1990), Goldberg (1992), dan John & Srivastava (1999) dimana tipe ini mengklasifikasikan kepribadian ke dalam lima faktor besar yang mana pembagiannya sebagai berikut:

- a. *Extraversion (Ekstrovert)*. Pada tipe ini mengukur tingkat kenyamanan seseorang dalam situasi sosial, keramahamahan, keaktifan, serta kebutuhan akan stimulasi. Orang yang memiliki skor tinggi cenderung ramah, asertif, dan mencari sensasi. Sedangkan yang memiliki skor rendah menunjukkan seseorang itu lebih pendiam, pemalu, dan penyendiri.
- b. *Agreeableness (keramahan/kebaikan hati)*. Pada tipe ini mengukur kualitas seperti kepercayaan, kepatuhan, rendah hati, serta tidak mudah menaruh curiga. Orang yang memiliki skor tinggi cenderung kooperatif, murah hati, dan dapat dipercaya. Skor yang rendah menunjukkan seseorang itu lebih mudah curiga, egois, dan antagonis.
- c. *Conscientiousness (Kesungguhan/Ketelitian)*. Pada tipe ini mengukur tingkat organisasi, persistensi, kontrol, serta motivasi dalam perilaku yang mengarah pada pencapaian tujuan. Skor *Conscientiousness* yang tinggi menunjukkan seseorang yang cenderung terorganisir, disiplin, dan dapat diandalkan. Sedangkan skor rendah menunjukkan seseorang yang lebih ceroboh dan tidak teratur.
- d. *Neuroticism (Neurotisme/Kecemasan)*. Pada tipe ini mengukur stabilitas emosi dan kontrol terhadap impuls. Seseorang dengan skor *Neuroticism* yang tinggi cenderung mudah cemas, temperamental, emosional, dan rentan terhadap stres. Skor rendah menunjukkan seseorang yang lebih tenang, stabil, dan terkendali.
- e. *Openness to Experience (Keterbukaan pada Pengalaman)*. Mengukur keingintahuan intelektual, kreativitas, dan preferensi terhadap hal-hal baru. Orang dengan skor tinggi cenderung imajinatif, penuh rasa ingin tahu, dan terbuka pada perubahan. Skor rendah menunjukkan seseorang yang lebih konvensional dan memilih rutinitas.

2.4.3 Tipe Enneagram

Merupakan sistem klasifikasi kepribadian yang membagi kepribadian manusia ke dalam sembilan tipe atau enneagram. Berikut penjelasan tentang sembilan tipe kepribadian dalam teori Enneagram menurut para ahli :

- a. Tipe Satu - *The Reformer* (Perfeksionis) Menurut Riso & Hudson (2000), tipe satu dicirikan dengan kecenderungan untuk selalu mencari kebenaran dan integritas. Individu menjadi cenderung perfeksionis, disiplin, dan berdedikasi pada sistem nilai tertentu.
- b. Tipe Dua - *The Helper* (Penolong) Naranjo (1994) menggambarkan tipe dua sebagai tipe yang memiliki rasa perhatian dan peduli pada kebutuhan orang lain. Mereka cenderung memiliki keinginan untuk dibutuhkan dan dihargai.
- c. Tipe Tiga - *The Achiever* (Pencari Prestasi) Menurut Riso & Hudson (2000), tipe tiga sangat memotivasi individu untuk memperoleh kesuksesan dan pengakuan dari lingkungan sekitarnya. Mereka cenderung ambisius dan berorientasi pada pencapaian.
- d. Tipe Empat - *The Individualist* (Romantis) Naranjo (1994) menggambarkan tipe empat sebagai tipe yang memprioritaskan autentisitas diri dan keunikan individu. Mereka cenderung kreatif, sensitif, dan introspektif.
- e. Tipe Lima - *The Investigator* (Pemikir) Riso & Hudson (2000) menjelaskan bahwa tipe lima sangat memiliki rasa ketertarikan pada pencarian pengetahuan dan pemahaman mendalam. Mereka cenderung analitis, cerdas, dan suka belajar.
- f. Tipe Enam - *The Loyalist* (Pengikut Setia) Menurut Naranjo (1994), tipe enam dicirikan dengan kecenderungan untuk mencari keamanan dan dukungan dari otoritas atau kelompok. Mereka cenderung waspada, loyal, dan bertanggung jawab.
- g. Tipe Tujuh - *The Enthusiast* (Optimis) Riso & Hudson (2000) menggambarkan tipe tujuh sebagai tipe yang selalu mencari kegembiraan dan pengalaman baru. Mereka cenderung optimis, spontan, dan bersemangat.
- h. Tipe Delapan - *The Challenger* (Pemimpin) Naranjo (1994) menjelaskan bahwa tipe delapan dicirikan dengan kekuatan, keberanian, dan keinginan untuk mengontrol lingkungan sekitarnya. Mereka cenderung tegas, asertif, dan langsung.
- i. Tipe Sembilan - *The Peacemaker* (Mediator) Menurut Riso & Hudson (2000), tipe sembilan sangat mementingkan keharmonisan dan menghindari konflik. Mereka

cenderung damai, mudah menyesuaikan diri, dan menerima.

2.4.4 The Myers-Briggs Type Indicator (MBTI)

Tipe kepribadian ini didasarkan kepada teori yang dikembangkan oleh Carl Jung. Dimana teori ini membagi kepribadian menjadi 16 tipe kepribadian yang berbeda yang didasarkan ke empat preferensi:

- a. *Extroversion* (E) vs. *Introversion* (I) Menurut Jung, *Ekstrovert* mendapatkan energi dari dunia luar, sedangkan *Introvert* dari dunia dalam dirinya sendiri.
- b. *Sensing* (S) vs. *Intuition* (N) *Sensing* mengacu pada cara memperoleh informasi melalui panca indera, sedangkan intuisi lebih memfokuskan pada pola, makna, dan kemungkinan.
- c. *Thinking* (T) vs. *Feeling* (F). *Thinking* mengacu pada pengambilan keputusan berdasarkan logika dan objektivitas, sedangkan *Feeling* lebih mengutamakan nilai-nilai subjektif dan pertimbangan emosional.
- d. *Judging* (J) vs. *Perceiving* (P). *Judging* cenderung hidup teratur dan terencana, sedangkan *Perceiving* lebih fleksibel dan spontan.

Dalam tipe kepribadian *The Myers-Briggs Type Indicator* (MBTI) memiliki 16 tipe kepribadian yaitu : ISTJ, ISFJ, INTJ, INFJ, INFP, ESTJ, ESFJ, ENTJ, ENFJ, ISTP, ISFP, INTP, ESFP, ESTP, ENTP, ENFP memiliki penjelasan sebagai berikut :

1. ISTJ (*Introvert, Sensing, Thinking, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat realistis, dapat dipercaya, rendah hati, dan sangat menjunjung tinggi tanggung jawab. Individu dengan tipe kepribadian ini cenderung cermat, pendiam atau lebih suka menyendiri, dan memiliki standar yang tinggi. Kepribadian ISTJ sangat dominan, sehingga orang lain seringkali merasa bahwa mereka dapat diandalkan, serta sangat berusaha untuk memenuhi janji yang telah dibuat.
2. ISFJ (*Introvert, Sensing, Feeling, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat pribadi yang hangat dan suka menolong. Individu dengan tipe kepribadian ini bekerja dengan sepenuh hati dan memiliki kemampuan dalam mengorganisasi berbagai hal dengan baik. Orang dengan kepribadian ISFJ merasa puas ketika mereka dapat berguna bagi orang lain dan memberikan perhatian serta

perawatan kepada orang di sekitarnya.

3. INTJ (*Introvert, INtuition, Thinking, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat pemikir yang mandiri. Individu dengan tipe kepribadian ini cenderung analitik, jenaka, serta memiliki rasa percaya diri yang tinggi dan tidak mudah terpengaruh oleh konflik atau kritik. Kepribadian ini sangat yakin dengan kemampuan yang dimilikinya.
4. INFJ (*Introvert, INtuition, Feeling, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan kemampuan berpikir mendalam dan perasaan yang kuat. Individu dengan tipe kepribadian ini biasanya memiliki sifat yang hangat, penuh perhatian, serta memiliki keinginan untuk memberikan motivasi dan dukungan kepada orang lain. Kepribadian INFJ juga ditandai dengan ketelitian dan komitmen yang kuat dalam setiap hal yang mereka lakukan.
5. ISTP (*Introvert, Sensing, Thinking, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat individualis. Individu dengan tipe kepribadian ini merasa percaya diri ketika sendirian dan memiliki rasa mandiri yang tinggi. Meskipun pendiam, mereka cenderung memiliki pemikiran yang realistis, rasional, dan tegas. Namun, mereka tidak terlalu suka dengan kewajiban atau rutinitas yang kaku. Kepribadian ini biasanya lebih suka mengamati situasi sebelum masalah muncul.
6. ISFP (*Introvert, Sensing, Feeling, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat peka terhadap perasaan orang lain, serta memiliki sifat lembut, rendah hati, dan pendiam. Individu dengan kepribadian ini cenderung optimis, tidak banyak bicara, namun menjadi pendengar yang baik dan sering dicari oleh orang lain untuk berbagi cerita. Kepribadian ISFP menciptakan kenyamanan di sekitarnya dan sering disukai oleh orang lain karena sifatnya yang perhatian dan menyenangkan.
7. INFP (*Introvert, INtuition, Feeling, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat idealis, terutama sebagai pemimpin yang berhati-hati. Meskipun tampak pemalu dan pendiam di mata orang lain, individu dengan tipe kepribadian ini sering kali suka membagikan kehidupan emosionalnya kepada orang-orang terdekat. Meskipun terkesan dingin dan pendiam, INFP memiliki sifat

suka menolong, setia, dan loyal, terutama dalam konteks organisasi.

8. INTP (*Introvert, INtuition, Thinking, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat pemikir dan analitis. Meskipun pendiam dan tidak banyak bicara, individu dengan tipe ini sangat fokus pada menyelesaikan pekerjaan dengan tuntas. Mereka cenderung menggali informasi secara mendalam untuk menyelesaikan masalah yang ada. Tipe ini memiliki rasa ingin tahu yang besar, fleksibel, mudah beradaptasi, namun sering kali merasa ragu. Selain itu, INTP juga memiliki sifat kritis dan analitis yang kuat.
9. ESTP (*Extrovert, Sensing, Thinking, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat periang dan spontan. Individu dengan tipe ini penuh pesona, enerjik, dan selalu terlibat dalam berbagai hal. Mereka memiliki pemikiran yang jernih, bahkan dalam situasi yang sulit. Kepribadian ESTP sering kali dianggap sebagai pemecah masalah yang handal, dengan pendekatan yang fleksibel dan toleran serta menggunakan pendekatan pragmatis dalam menghadapi tantangan.
10. ESFP (*Extrovert, Sensing, Feeling, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat santai, ramah, cerdas, dan pandai dalam berbicara serta mengungkapkan pendapat. Individu dengan tipe ini juga memiliki selera humor yang baik. Mereka senang menjadi pusat perhatian dan dikenal sebagai orang yang luwes, kreatif, serta cepat beradaptasi dengan berbagai situasi.
11. ENFP (*Extrovert, INtuition, Feeling, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat idealis, spontan, kreatif, berpikiran terbuka, dan penuh dengan humor. Individu dengan tipe ini memiliki kemampuan pengamatan yang tajam dan tidak ingin ketinggalan informasi atau kejadian yang terjadi di sekitar mereka.
12. ENTP (*Extrovert, INtuition, Thinking, Perceiving*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat antusias dan penuh energi. Individu dengan tipe ini suka berada dalam posisi yang menarik perhatian dan menikmati keragaman dalam kehidupan profesional maupun pribadi. Mereka memiliki ide-ide brilian, suka berbicara, dan memiliki semangat tinggi dalam menghadapi tantangan atau masalah yang ada di hadapan mereka.

13. ESTJ (*Extrovert, Sensing, Thinking, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat bertekad kuat, bertanggung jawab, dan antusias terhadap tantangan baru. Individu dengan tipe ini memiliki sifat stabil dan dapat diandalkan. Mereka suka berkomunikasi dengan orang lain, mudah berbaur, sangat aktif, serta memiliki pendekatan berpikir yang praktis dan realistis, mengandalkan fakta yang akurat. Tipe ini juga cepat dalam mengambil keputusan, peduli dengan rutinitas yang detail, dan memiliki standar logika yang jelas.
14. ESFJ (*Extrovert, Sensing, Feeling, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan kecenderungannya untuk senang berada dalam lingkungan sosial. Individu dengan tipe ini dapat diandalkan, terorganisir dengan baik, dan memiliki sifat pengertian serta peka terhadap kelemahan orang lain. Kepribadian ini mudah bergaul dengan berbagai tipe kepribadian lainnya. ESFJ sangat menghargai interaksi sosial dan cenderung terbuka serta jujur dalam menunjukkan perasaan mereka
15. ENTJ (*Extrovert, Intuition, Thinking, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan sifat pemikir yang dinamis, penuh percaya diri, dan mandiri. Individu dengan tipe ini mengerjakan segala sesuatu dengan semangat tinggi. Mereka menyukai tantangan baru, memiliki kemampuan kepemimpinan yang kuat, kompeten, penuh semangat, dan bertanggung jawab.
16. ENFJ (*Extrovert, Intuition, Feeling, Judging*) adalah tipe kepribadian dalam MBTI yang dikenal dengan energi tinggi dan kecenderungan untuk berinteraksi dengan orang lain serta dunia luar. Individu dengan tipe ini memiliki fokus yang tajam terhadap pola dan detail konkret, serta mampu membuat keputusan berdasarkan pertimbangan nilai, empati, dan harmoni. Mereka cenderung menjalani hidup dengan cara yang terorganisir, terencana, dan terstruktur.

Berdasarkan penjelasan tipe-tipe di atas, penulis memutuskan untuk menggunakan tipe kepribadian MBTI. Dengan alasan MBTI merupakan salah satu tipe kepribadian yang sedang populer untuk mengungkap kepribadian individu sehingga bisa digunakan untuk melihat kecocokan seseorang di suatu bidang tertentu. Tidak ada kepribadian yang lebih baik atau lebih buruk.

2.5 Aplikasi 16 Personalities

16Personalities adalah tes kepribadian yang menggunakan sistem tipe MBTI, yang

dikembangkan oleh *NERIS Type Explorer*. Tes ini menggunakan format akronim yang diperkenalkan oleh *Myers-Briggs* karena kesederhanaan dan kemudahannya. Tes kepribadian ini tersedia dalam 49 bahasa, termasuk bahasa Indonesia, dan dapat diakses secara gratis.

2.5.1 Koefisiensi Internal

Koefisiensi atau yang bisa disebut *alpha Cronbach* digunakan untuk mengukur apakah pertanyaan yang termasuk dalam skala yang sama akan menghasilkan skor yang sama.

Nilai alfa umumnya berada diantara 0,70 – 0,90. Dan nilai yang lebih rendah menunjukkan bahwa pertanyaan yang ditanyakan mungkin tidak mengukur konstruksi yang sama. Nilai yang lebih tinggi menandakan redudansi. Pada tabel 2.1 menunjukkan tabel nilai alfa untuk tes kepribadian *16Personalities*.

Tabel 2. 1 Nilai Cronbach Alpha

Scale	Alpha
Introverted vs. Extraverted	0.87
Observant vs. Intuitive	0.78
Thinking vs. Feeling	0.75
Judging vs. Prospecting	0.82
Assertive vs. Turbulent	0.86

Sample size: 10,000 respondents.

2.5.2 Keandalan Uji-Ulang

Keandalan uji ulang menunjukkan seberapa besar korelasi antara hasil awal dan pengujian ulang yang mana biasanya setelah jangka waktu yang lebih lama. Semakin tinggi koefisiensi keandalan, maka semakin sedikit variabilitas pada skala tertentu. Sama seperti Alpha Cronbach, semua koefisiensi diharapkan mencapai 0,70 atau lebih. Keandalan uji ulang dari alat ukur kepribadian ditampilkan dalam Tabel 2.2 untuk memastikan konsistensi hasil pengukuran.

Tabel 2. 2 Keandalan Uji Ulang

Scale	Coefficient
Introverted vs. Extraverted	0.83
Observant vs. Intuitive	0.74
Thinking vs. Feeling	0.80
Judging vs. Prospecting	0.79
Assertive vs. Turbulent	0.78

Sample size: 2,900 respondents, who took our assessment after a break of 5-7 months. $p < 0.001$.

2.5.3 Validitas Diskriminan

Pada langkah ini akan mengonfirmasi apakah skala yang seharusnya tidak terkait benar- benar tidak terkait. Nilai maksimum yang diterima untuk koefisiensi ini biasanya sekitar 0,70- 0,80, dan jika lebih dari nilai tersebut berarti ada cukup tumpang tindih antara kedua skala tersebut sehingga tidak valid. Tabel 2.3 menyajikan hasil analisis validitas diskriminan untuk menilai apakah skala yang digunakan dapat membedakan aspek yang diukur.

Tabel 2. 3 Validitas dan Diskriminan

	Introverted vs. Extraverted	Observant vs. Intuitive	Thinking vs. Feeling	Judging vs. Prospecting	Assertive vs. Turbulent
Introverted vs. Extraverted		-0.09	0.02	-0.01	-0.29
Observant vs. Intuitive	-0.09		0.09	0.37	0.22
Thinking vs. Feeling	0.02	0.09		0.08	0.25
Judging vs. Prospecting	-0.01	0.37	0.08		0.16
Assertive vs. Turbulent	-0.29	0.22	0.25	0.16	

Sample size: 10,000 respondents.

Berdasarkan penjelasan diatas pada penelitian ini penulis menggunakan website 16 Personalities sebagai tolak ukur tes kepribadian yang akan di klasifikasi pada algoritma *Tree Based Model*.

2.6 Kajian Pustaka

Kajian pustaka ini berisi definisi, kelebihan dan kekurangan dari algoritma yang akan digunakan serta penelitian-penelitian terdahulu sebagai bahan referensi bagi penulis selama penelitian.

2.6.1 Algoritma Tree Based Model

Dibawah ini merupakan definisi, kelebihan, dan kekurangan beberapa algoritma yang akan digunakan pada penelitian ini.

a. Decision Tree

Decision Tree merupakan algoritma yang menggunakan struktur pohon untuk membuat keputusan berdasarkan atribut-atribut input. Algoritma ini dikembangkan oleh Quinlan (1986). Quinlan mengembangkan algoritma ini sebagai bagian dari proyek penelitian yang bertujuan untuk mengembangkan sistem kecerdasan buatan yang dapat membuat keputusan berdasarkan data.

Kelebihan algoritma *Decision Tree* adalah dapat menangani data yang tidak terstruktur dan dapat digunakan untuk klasifikasi dan regresi. Namun, algoritma ini juga mempunyai kekurangan, yaitu dapat memiliki kesalahan yang signifikan jika data tidak terdistribusi secara acak dan dapat memiliki kompleksitas yang tinggi jika memiliki banyak atribut.

Berikut adalah contoh penelitian terdahulu yang meneliti klasifikasi tes kepribadian MBTI menggunakan *Decision Tree*. Feixiong Chen et al. (2023) dalam penelitiannya dengan judul “*Myers-Briggs Type Indicator analysis using decision-treerelated methods*”. Hasil yang di dapatkan adalah pada penelitian ini mencapai kinerja yang baik dalam hal akurasi klasifikasi dibandingkan dengan CNN (*Convolutional Neural Networks*) dan LSTM (*long short-term memory*) tradisional pada kumpulan data. Dalam penelitian ini, *Decision Tree* diterapkan pada klasifikasi empat huruf yang berbeda di MBTI. Hal ini menunjukkan bahwa model terkait *Decision Tree* efisien dalam beberapa kepribadian pengukuran, akan tetapi keakuratannya dapat bervariasi dalam dimensi yang berbeda.

b. Random Forest

Random Forest merupakan algoritma yang menggunakan kombinasi dari beberapa *Decision Tree* untuk membuat keputusan. Algoritma ini dikembangkan oleh Breiman (2001). Breiman mengembangkan algoritma ini sebagai bahan proyek penelitian yang bertujuan untuk mengembangkan sistem kecerdasan buatan untuk membuat keputusan berdasarkan data yang besar dan kompleks.

Kelebihan algoritma ini adalah dapat menangani data yang besar dan kompleks, serta dapat mengurasi kesalahan yang disebabkan oleh *Decision Tree* individu, dapat digunakan juga untuk klasifikasi dan regresi. Sedangkan kekurangannya, algoritma ini hanya dapat memiliki kompleksitas yang tinggi jika memiliki banyak atribut, dan memiliki waktu komputasi yang lama jika memiliki banyak *Decision Tree*.

Berikut adalah contoh penelitian terdahulu yang meneliti klasifikasi tes kepribadian MBTI menggunakan *Random Forest*. Faisal Amri et al. (2022) dalam penelitiannya yang berjudul “*Social Media User Personality Classification Based on How User Live and Make Decisions*”. Hasil penelitian yang dilakukan adalah percobaan dengan data oversample *Random Forest* memiliki hasil yang tinggi dibandingkan dengan percobaan lainnya. Kategori tipe kepribadian *Thinkers/Feelers* (TF) dan *Judgers/Perceivers* (JP) memberikan hasil yang paling baik dan serupa dengan *F1-score* dan akurasi sebesar 92%. Di sisi lain, untuk semua kelas, hasilnya cukup baik dengan *F1-score* dan akurasi sebesar 84%.

c. Extra Trees

Extra Trees (*Extremely Randomized Trees*) ditemukan oleh Geurts et al. (2006) merupakan variasi dari *Random Forest* yang menambahkan tingkat randomisasi dalam proses pembangunan tree. Berbeda dengan *Random Forest* yang mencari *split* terbaik untuk setiap fitur, *Extra Trees* memilih *split points* secara acak untuk setiap fitur dan memilih yang terbaik di antara *split points* tersebut. Pendekatan ini menghasilkan *training* yang lebih cepat dan *variance* yang lebih rendah dibanding *Random Forest*.

Extra Trees efektif dalam menangani *noise* dalam data dan memberikan *feature importance* yang stabil. Namun, algoritma ini mungkin memerlukan jumlah *trees* yang

lebih banyak untuk mencapai performa optimal dan bisa kurang akurat pada dataset yang memerlukan pemilihan *split point* yang presisi.

Berikut adalah contoh penelitian terdahulu yang meneliti klasifikasi tes kepribadian MBTI menggunakan *Extra Trees*. Garg et al. (2021) Dalam penelitiannya yang berjudul “*Comparison of Machine learning Algorithms for Content Based Personality Resolution of Tweets.*” Hasil penelitian ini adalah algoritma *Extra Trees* yang paling presisi untuk sumbu N-S, dan untuk sumbu yang lain algoritma ini juga mempunyai nilai akurasi yang cukup bagus walaupun bukan yang tertinggi.

d. AdaBoost

AdaBoost (Adaptive Boosting) dikembangkan oleh Freund & Schapire (1997) merupakan algoritma *boosting* yang fokus pada data yang sulit diklasifikasi dengan memberikan bobot lebih tinggi pada sampel yang salah klasifikasi. Algoritma ini membangun *weak learners* secara sekuensial dan menggabungkan prediksi mereka dengan pembobotan.

AdaBoost efektif dalam meningkatkan akurasi klasifikasi dan relatif sederhana dibanding algoritma *boosting* lainnya. Namun, algoritma ini sensitif terhadap *noise* dan *outlier* dalam data, serta bisa mengalami *overfitting* pada dataset yang kompleks.

Berikut adalah contoh penelitian terdahulu yang meneliti klasifikasi tes kepribadian MBTI menggunakan *AdaBoost*. Angsaweni & Maharani (2022) Dalam penelitiannya yang berjudul “*Identification of Big Five Personality on Twitter Users using the AdaBoost Method.*” Hasil penelitian menunjukkan bahwa Model terbaik yang dibangun dari percobaan adalah skenario pertama, yaitu model *AdaBoost* menggunakan seluruh fitur tanpa proses *stemming*, dengan akurasi sebesar 53,57%. Sementara itu, akurasi terbaik pada skenario kedua adalah kombinasi fitur emotion, PCA, dan sentiment tanpa *stemming* dengan akurasi sebesar 50%.

e. XGBoost

XGBoost (eXtreme Gradient Boosting) dikembangkan oleh T. Chen & Guestrin (2016) merupakan implementasi yang dioptimalkan dari *Gradient Boosting* yang menawarkan peningkatan performa dan efisiensi. Algoritma ini menambahkan regularisasi dan berbagai optimasi teknis untuk meningkatkan akurasi dan kecepatan *training*.

XGBoost unggul dalam penanganan *missing values*, paralelisasi komputasi, dan menghasilkan performa yang konsisten tinggi dalam berbagai kompetisi *machine learning*. Meskipun demikian, *XGBoost* memerlukan *tuning* parameter yang cermat dan pemahaman yang baik tentang *hyperparameter* untuk mencapai hasil optimal.

Berikut adalah contoh penelitian terdahulu yang meneliti klasifikasi tes kepribadian MBTI menggunakan *XGBoost*. Nisha et al. (2022) Dalam penelitiannya yang berjudul “*A Comparative Analysis of Machine learning Approaches in Personality Prediction Using MBTI.*” Sistem yang diusulkan mencapai hasil yang memuaskan dalam memprediksi kepribadian dengan menggunakan tweet. Dalam tiga pengklasifikasi, kinerja pengklasifikasi *Naive Bayes* lebih buruk sementara Linear SVM dan *XGBoost* memberikan hasil yang memuaskan. Hasil ini diperoleh dengan himpunan data yang tidak seimbang. Jadi kinerja pengklasifikasi ini dapat ditingkatkan dengan menyeimbangkan himpunan data atau membuat himpunan data keseimbangan baru.

f. Gradient Boosting Decision Trees (GBDT)

Gradient Boosting Decision Trees (GBDT) dikembangkan oleh Friedman (2001) merupakan algoritma yang membangun model secara sekuensial, dimana setiap *tree* baru berusaha memperbaiki kesalahan prediksi dari *tree* sebelumnya menggunakan *gradient descent*.

Algoritma ini memiliki kemampuan untuk mengoptimalkan berbagai fungsi loss dan memberikan performa yang sangat baik untuk berbagai tipe data. GBDT unggul dalam akurasi prediksi dan fleksibilitas dalam penanganan berbagai masalah *machine learning*. Kelemahannya termasuk sensitifitas terhadap *noise* dan *outlier*, kecenderungan *overfitting* jika tidak di-*tune* dengan tepat, serta waktu *training* yang relatif lama.

Berikut adalah contoh penelitian terdahulu yang meneliti klasifikasi tes kepribadian MBTI menggunakan *Gradient Boosting Decision Trees* (GBDT). Mushtaq et al. (2020) Dalam penelitiannya yang berjudul “*Predicting MBTI Personality type with K-means Clustering and Gradient Boosting.*” Hasil penelitian menunjukkan bahwa kombinasi *K-means clustering* dan *Gradient Boosting* memberikan akurasi yang lebih tinggi dibandingkan dengan metode konvensional.

Penulis mencatat bahwa penggunaan *K-means* sebagai langkah awal dalam pengelompokan data membantu dalam mengurangi kompleksitas dan meningkatkan kinerja model prediksi.

2.6.2 Penelitian Terdahulu

Ringkasan penelitian terdahulu yang digunakan sebagai referensi dalam penelitian ini disajikan dalam Tabel 2.4.

Tabel 2. 4 Penelitian Terdahulu

Judul	Publikasi	Algoritma	Hasil
“Penggunaan Indikator <i>Myers-Briggs Type Indicator</i> (MBTI) untuk Penempatan Karyawan Memanfaatkan Sistem Pakar Metode <i>Forward Chaining</i>	Ikhran & Ramsari (2023)	<i>Forward Chaining</i>	metode <i>forward chaining</i> adalah metode yang tepat jika digunakan untuk tes kepribadian karena karakteristik dari metode <i>forward chaining</i> sesuai dengan kemampuan dari aplikasi sistem pakar kepribadian itu sendiri, yakni pengguna akan memilih fakta fakta yang sesuai dengan sifat yang dimilikinya kemudian <i>Inference Engine</i> akan mengambil kesimpulan berdasarkan aturan aturan yang sebelumnya sudah dibuat.
<i>MBTI Personality Prediction Using Machine learning and SMOTE for Balancing Data Based on Statement Sentences</i>	Ryan et al., (2023)	<i>Logistic Regression</i> (LR) SMOTE	metode <i>Machine learning</i> terbaik untuk memprediksi tipe MBTI adalah <i>Logistic Regression</i> (LR), dengan skor F1 rata-rata sebesar 0,8282. Teknik SMOTE yang diterapkan juga memberikan hasil yang lebih baik, dengan skor F1 meningkat menjadi 0,8337, dan dimensi 3 (N/S) mencatatkan skor tertinggi sebesar 0,8821. Ambang batas skor F1 yang diterima bervariasi tergantung pada aplikasinya, tetapi skor F1 yang mendekati 1 umumnya dianggap sangat baik untuk klasifikasi data. Oleh karena itu, hasil ini lebih unggul dibandingkan model lain yang diuji, yang menunjukkan bahwa pendekatan yang diusulkan dapat meningkatkan pemahaman tentang MBTI dan berpotensi digunakan dalam berbagai aplikasi yang membutuhkan klasifikasi kepribadian.

<i>MBTI Personality Predictio Based on BERT Classification</i>	Zhang (2023)	BERT	Hasil penelitian menunjukkan penggunaan metode BERT Classification dan Vektorizer TF-IDF untuk memprediksi kepribadian MBTI. Namun, tantangan utama dalam penelitian ini adalah ketidakseimbangan Sampel yang dikumpulkan, yang menyebabkan penurunan performa model, sehingga metode tersebut menjadi kurang efektif meskipun sampel yang digunakan dalam penelitian ini sepuluh kali lebih banyak dibandingkan dengan penelitian sebelumnya.
<i>Myers-Briggs TypeIndicator analysis using decision-tree- related methods</i>	F. Chen et al. (2023)	<i>Decision Tree</i>	Hasil penelitian menunjukkan bahwa model ini mencapai kinerja yang baik dalam hal akurasi klasifikasi, lebih unggul dibandingkan dengan CNN (<i>Convolutional Neural Networks</i>) dan LSTM (<i>Long Short-Term Memory</i>) tradisional pada kumpulan data yang digunakan. Dalam penelitian ini, <i>Decision Tree</i> diterapkan untuk mengklasifikasikan empat huruf yang berbeda dalam MBTI. Temuan ini menunjukkan bahwa model berbasis <i>Decision Tree</i> efisien untuk beberapa pengukuran kepribadian, meskipun tingkat akurasinya dapat bervariasi pada dimensi yang berbeda.
<i>Social Media User Personality Classification Based on How User Live and Make Decisions"</i>	Faisal Amri et al. (2022)	<i>Random Forest</i>	Hasil penelitian yang dilakukan adalah percobaan dengan data <i>oversample</i> memiliki hasil yang tinggi dibandingkan dengan percobaan lainnya. Kategori tipe kepribadian <i>Thinkers/Feelers</i> (TF) dan <i>Judgers/Perceivers</i> (JP) memberikan hasil yang paling baik dan serupa dengan <i>F1-score</i> dan akurasi sebesar 92%. Di sisi lain, untuk semua kelas, hasilnya cukup baik dengan <i>F1-score</i> dan akurasi sebesar 84%.
<i>A Comparative Analysis Of Machine Learning Approaches in Personality Prediction Using MBTI</i>	Nisha et al (2022)	SVM, <i>Naïve Bayes</i> , <i>XGBoost</i>	Sistem yang diusulkan mencapai hasil yang memuaskan dalam memprediksi kepribadian dengan menggunakan tweet. Dalam tiga pengklasifikasi, kinerja pengklasifikasi Naive Bayes lebih buruk sementara Linear SVM dan XGBoost memberikan hasil yang memuaskan. Hasil ini diperoleh dengan himpunan data yang tidak seimbang. Jadi kinerja pengklasifikasi ini dapat ditingkatkan dengan menyeimbangkan himpunan data atau membuat himpunan data keseimbangan baru.

<i>Predicting MBTI Personality type with K-means Clustering and Gradient Boosting</i>	Mushtaqet al. (2020)	<i>K-means Clustering, Gradient Boosting</i>	Hasil penelitian menunjukkan bahwa Kombinasi <i>K-means clustering</i> dan <i>Gradient Boosting</i> memberikan akurasi yang lebih tinggi dibandingkan dengan metode konvensional. Penulis mencatat bahwa penggunaan <i>K-means</i> sebagai langkah awal dalam pengelompokan data membantu dalam mengurangi kompleksitas dan meningkatkan kinerja model prediksi.
<i>Comparison of Machine learning Algorithms for Content Based Personality Resolution of Tweets</i>	Garg et al. (2021)	<i>Logistic Regression (LR), KNN, Naive Bayes (NB), SVM, Multilayer Perceptron, Extra Tree, Bagging</i>	Hasil Penelitian menunjukkan bahwa <i>Multilayer Perceptron</i> bekerja lebih efisien daripada algoritma pembelajaran mesin lainnya yang diuji. Dan untuk masing-masing sumbu, <i>Multilayer Perceptron</i> adalah yang paling presisi untuk sumbu I-E, sedangkan <i>Extra Tree</i> adalah yang paling presisi pada sumbu N-S, <i>Random Forest</i> ditemukan sebagai yang paling presisi pada sumbu T-F. Terakhir, GNB ditemukan sebagai yang paling presisi pada sumbu J-P.
<i>Identification of Big Five Personality on Twitter Users using the AdaBoost Method</i>	Angsaweni & Maharani (2022)	<i>AdaBoost</i>	Hasil penelitian menunjukkan bahwa Model terbaik yang dibangun dari percobaan adalah skenario pertama, yaitu model <i>AdaBoost</i> menggunakan seluruh fitur tanpa proses <i>stemming</i> , dengan akurasi sebesar 53,57%. Sementara itu, akurasi terbaik pada skenario kedua adalah kombinasi fitur emotion, PCA, dan sentiment tanpa <i>stemming</i> dengan akurasi sebesar 50%.

Berdasarkan penelitian-penelitian terdahulu, penggunaan *Tree Based Model* untuk klasifikasi MBTI menunjukkan beberapa keunggulan yang signifikan. Dari sisi performa dan akurasi, penelitian F. Chen et al. (2023) membuktikan bahwa *Decision Tree* memberikan hasil yang lebih baik dibandingkan CNN dan LSTM dalam klasifikasi MBTI. Hal ini diperkuat dengan penelitian Faisal Amri et al. (2022) yang berhasil mencapai akurasi hingga 92% menggunakan *Random Forest* untuk dimensi TF dan JP, menunjukkan konsistensi *Tree Based Model* dalam memberikan akurasi yang tinggi.

Tree Based Model juga menunjukkan kemampuan superior dalam menangani kompleksitas data MBTI yang memiliki 16 tipe kepribadian. Algoritma ini mampu mengolah data kategorikal dan numerik secara efektif, serta menangani *missing values* dan *outliers* dengan baik. Keunggulan lain yang menonjol adalah aspek interpretabilitas, dimana hasil penelitian menunjukkan bahwa *Tree Based Model* memberikan interpretasi yang lebih mudah dipahami, memungkinkan analisis *feature importance* untuk

memahami faktor-faktor yang mempengaruhi klasifikasi, dan menawarkan transparansi dalam proses pengambilan keputusan. Fleksibilitas *Tree Based Model* terlihat dari tersedianya berbagai varian algoritma seperti *Decision Tree*, *Random Forest*, dan *Extra Trees* yang memungkinkan optimasi sesuai kebutuhan spesifik. Algoritma-algoritma ini dapat dikombinasikan dengan teknik lain seperti boosting dan bagging, serta bersifat adaptif terhadap berbagai jenis dataset. Dari segi efisiensi komputasi, *Tree Based Model* tidak memerlukan *preprocessing* data yang ekstensif, memiliki *training time* yang relatif cepat dibandingkan *deep learning*, dan menggunakan memori secara efisien. Berdasarkan hasil-hasil penelitian terdahulu tersebut, *Tree Based Model* terbukti menjadi pilihan yang tepat karena menawarkan kombinasi optimal antara performa tinggi, interpretabilitas yang baik, kemampuan menangani data kompleks, dan efisiensi dalam implementasi.

BAB III METODE PENELITIAN

3.1 Desain Penelitian

Pada penelitian ini menggunakan pendekatan kuantitatif dengan metode eksperimen untuk mengeksplorasi dan membandingkan performa berbagai algoritma turunan dari *Tree Based Model* dalam kasus klasifikasi tipe kepribadian MBTI.

Fraenkel dan Wallen (2009) menyatakan eksperimen itu mencoba, mencari, serta mengkonfirmasi. Gordon L Patzer (1996) mengatakan bahwasanya hubungan kausal atau sebab akibat adalah inti dari penelitian eksperimen. Dari penjelasan diatas dapat di simpulkan bahwa metode penelitian eksperimen merupakan metode penelitian kuantitatif yang digunakan untuk mengetahui pengaruh dari variabel *independen* terhadap variabel *dependen* dalam kondisi yang terkendalikan.

Gambar 3.1 menunjukkan diagram alur penelitian dimana menggambarkan semua tahapan-tahapan yang dilakukan pada penelitian ini, mulai dari pengumpulan data, *pre-processing data*, analysis, hingga *modelling* dan *testing*.

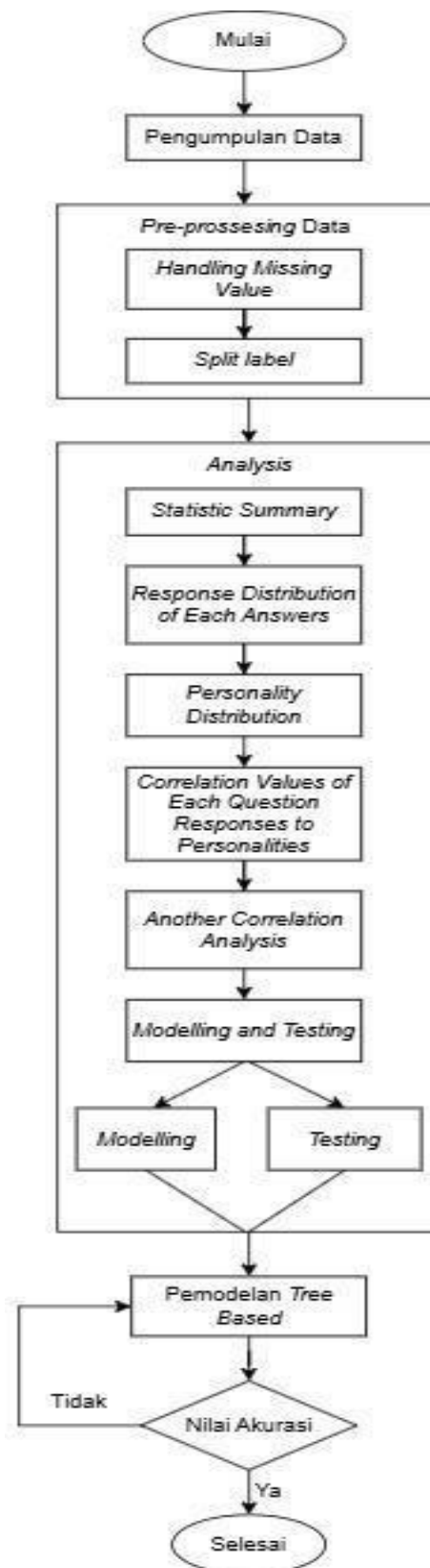
3.2 Variabel Penelitian

Variabel merupakan suatu atribut atau sifat dari seseorang, objek atau kegiatan yang memiliki variasi tertentu yang ditetapkan oleh peneliti untuk di pelajari serta disimpulkan (Sugiono, 2015: 38). Dalam penelitian ini terdapat dua variabel, antara lain:

a. Variabel Bebas (*Independent Variable*)

Variabel bebas merupakan variabel yang mempengaruhi atau menjadi penyebab timbulnya variabel terikat (*dependent variable*) (Sugiono, 2015: 39). Pada penelitian ini yang menjadi variabel bebas adalah tipe kepribadian MBTI yang terdiri dari 4 dimensi, yaitu:

1. *Extroversion* (E) - *Introversion* (I)
2. *Sensing* (S) - *iNtuition* (N)
3. *Thinking* (T) - *Feeling* (F)
4. *Judging* (J) - *Perceiving* (P)



Gambar 3. 1 Diagram Penelitian

b. Variabel Terikat (*Dependent Variable*)

Variabel terikat merupakan variabel yang dipengaruhi atau yang menjadi akibat, karena adanya variabel bebas (*independent variable*) (Sugiono,2015: 39). Pada penelitian ini yang menjadi variabel terikat adalah klasifikasi tipe kepribadian, yang merupakan hasil dari prediksi model berdasarkan tipe kepribadian MBTI.

3.3 Subjek Penelitian

Dalam penelitian ini, subjek penelitian terdiri dari individu yang berpartisipasi dalam pengisian kuesioner MBTI yang dipilih secara random.

3.4 Pengumpulan Data

3.4.1 Pengisian Kuesioner

Pada penelitian ini data yang digunakan didapatkan dari pengisian kuesioner, dimana pertanyaan yang di tanyakan mengacu pada salah satu website tes kepribadian MBTI *16Personalities*. Sebelum mengisi kuesioner yang ada peserta diharuskan untuk menjawab tes terlebih dahulu pada *website* tersebut kemudian mengisi hal yang serupa pada kuesioner yang ada dan di akhiri dengan tangkapan layar hasil dari tes kepribadian pada *website 16Personalities*.

3.5 Pre-Processing Data

Pre-Processing Data pada penelitian ini dilakukan untuk menyiapkan data agar siap untuk digunakan dalam melakukan pemodelan data. Pada tahap *Pre-Processing Data* ini terdapat beberapa langkah, yaitu:

3.5.1 Penanganan Missing Value

Penanganan *Missing Value* dalam penelitian dilakukan untuk menggantikan nilai yang hilang. *Missing Value* merujuk pada data yang tidak tersedia untuk suatu objek. Hal ini bisa terjadi karena informasi terkait objek tersebut tidak disediakan, sulit diperoleh, atau memang tidak ada sama sekali.

3.5.2 Split Label

Penilaian MBTI mengidentifikasi orang sebagai salah satu dari 16 tipe kepribadian MBTI. Setiap tipe mencerminkan bagaimana seseorang secara alami lebih suka mengarahkan dan menerima energi, menyerap informasi, membuat keputusan, dan mendekati dunia luar. Mengetahui hal ini memberikan kerangka kerja yang kuat untuk memahami dan berhubungan dengan orang lain. Penilaian ini menghasilkan empat dimensi:

- a. *Ekstrovert* [E] - *Introvert* [I]
- b. *Sensing* [S] - *iNtuition* [N]
- c. *Thinking* [T] - *Feeling* [F]
- d. *Judging* [J] - *Perceiving* [P]

Pada penelitian ini hanya akan menggunakan 4 dimensi inti MBTI, dan tidak akan menyertakan komponen terakhir (-A/-T). Tabel 3.1 menunjukkan hasil proses pemisahan label dalam dataset untuk setiap dimensi.

Tabel 3. 1 Split Label

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q55	q56	q57	q58	q59	q60	E-I	S-N	T-F	J-P
0	6	5	3	3	3	4	6	6	6	3	...	7.0	7.0	3	3.0	1	6.0	E	N	T	J
1	7	7	6	7	7	6	6	5	6	6	...	7.0	7.0	7	1.0	1	1.0	I	N	F	J
2	1	1	6	5	5	4	2	2	1	2	...	1.0	1.0	1	1.0	4	4.0	E	S	T	P
3	5	5	5	4	4	3	2	2	3	7	...	4.0	4.0	3	3.0	3	1.0	E	S	T	P
4	5	5	5	4	4	4	6	6	5	5	...	7.0	7.0	7	7.0	4	4.0	I	N	F	J

5 rows × 64 columns

3.5.3 Analisis Statistic Summary

Pada tahap ini hal pertama yang di lakukan adalah menampilkan *Statistic Summary*. *Statistic Summary* dari dataset yang digunakan dalam penelitian dapat dilihat pada tabel 3.2.

Tabel 3. 2 Statistic Summary

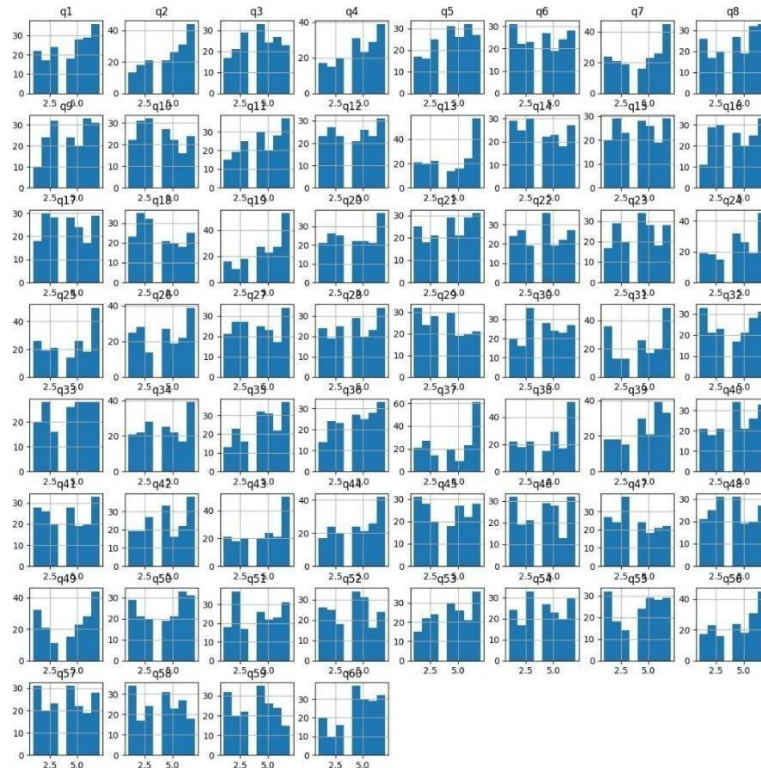
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q51	
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	...	174.000000	174.000000
mean	4.402299	4.712644	4.143678	4.557471	4.362069	3.948276	4.442529	4.287356	4.396552	3.804598	...	4.091954	3.912644
std	2.053953	1.964471	1.864496	1.966372	1.887270	2.091179	2.173042	2.084393	1.895596	1.937396	...	2.037976	1.912644
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000	1.000000
25%	3.000000	3.000000	3.000000	3.000000	3.000000	2.000000	2.000000	3.000000	3.000000	2.000000	...	2.000000	2.000000
50%	5.000000	5.000000	4.000000	5.000000	4.000000	4.000000	5.000000	4.000000	4.000000	4.000000	...	4.000000	4.000000
75%	6.000000	6.750000	6.000000	6.000000	6.000000	6.000000	7.000000	6.000000	6.000000	5.000000	...	6.000000	5.000000
max	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	...	7.000000	7.000000

8 rows × 60 columns

3.5.4 Response Distribution of Each Answers

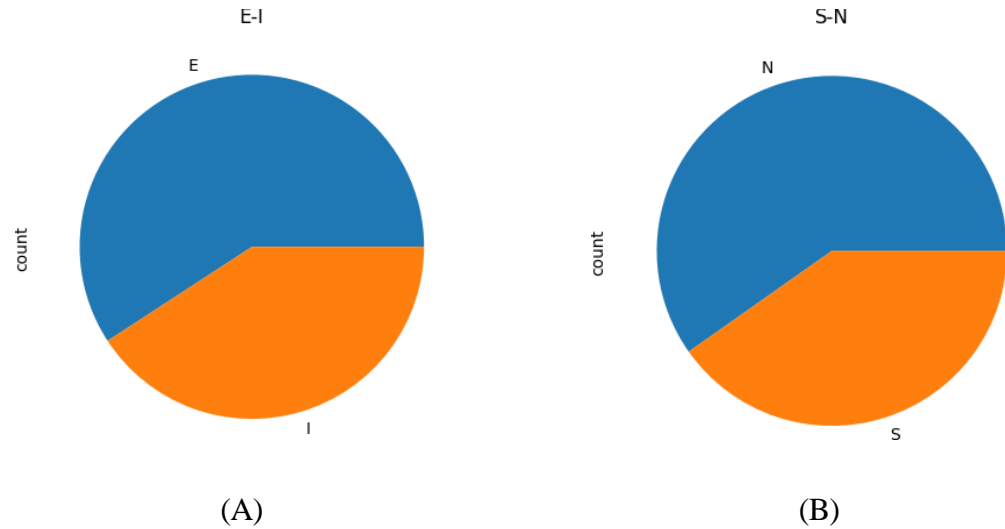
Setelah menampilkan *Statistic Summary* kemudian menampilkan respons peserta pada setiap pertanyaan yang ada. Distribusi respons peserta terhadap setiap pertanyaan kuesioner disajikan pada tabel 3.3.

Tabel 3. 3 Response Distribution of Each Answer

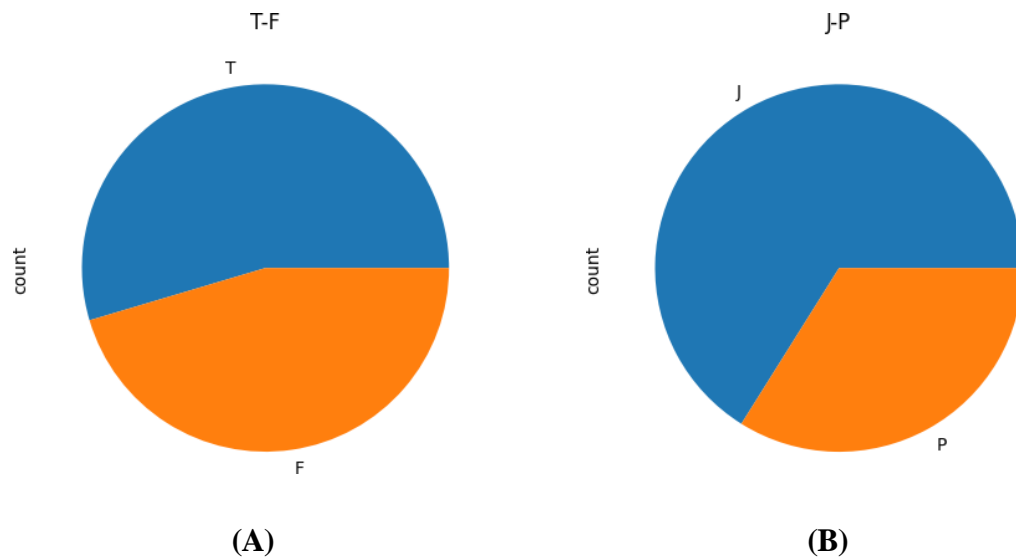


3.5.5 Personality Distribution

Kemudian menampilkan *Personality distribution* untuk setiap dimensi yang ada. Distribusi kepribadian berdasarkan dimensi MBTI ditampilkan dalam Gambar 3.2 dan Gambar 3.3. Gambar-gambar ini menunjukkan bagaimana data responden terdistribusi dalam empat dimensi utama MBTI, yaitu E-I, S-N, T-F, dan J-P, yang menjadi dasar dalam analisis klasifikasi kepribadian."



Gambar 3. 2 (A) E-I (B) S-N



Gambar 3. 3 (A) T-F (B) J-P

- Ekstrovert [E] - Introvert [I]
- Sensing [S] - iNtuition [N]
- Thinking [T] - Feeling [F]
- Judging [J] - Perceiving [P]

3.5.6 Correlation values Of Each Question Responses to Personalities

Setelah menampilkan *personality distribution*, langkah selanjutnya adalah membuat *Correlation values* dari setiap jawaban pertanyaan kepada kepribadian yang sudah ditentukan. Sebelumnya untuk membuat matriks korelasi, perlu dilakukan perubahan kepribadian yang ada kedalam bentuk angka. Dan dalam kasus ini akan diubah menjadi

boolean (0/1). Proses konversi data kepribadian menjadi format boolean ditampilkan dalam gambar 3.4.

```
C:\Users\USER\AppData\Local\Temp\ipykernel_16480\4032763019.py:8:
df_binary['E-I'] = df_binary['E-I'].replace({'E':1, 'I':0})
C:\Users\USER\AppData\Local\Temp\ipykernel_16480\4032763019.py:9:
df_binary['S-N'] = df_binary['S-N'].replace({'S':1, 'N':0})
C:\Users\USER\AppData\Local\Temp\ipykernel_16480\4032763019.py:10:
df_binary['T-F'] = df_binary['T-F'].replace({'T':1, 'F':0})
C:\Users\USER\AppData\Local\Temp\ipykernel_16480\4032763019.py:11:
df_binary['J-P'] = df_binary['J-P'].replace({'J':1, 'P':0})
```

Gambar 3. 4 Perubahan kepribadian menjadi boolean

Contoh tampilan dataset setelah dilakukan preprocessing dapat dilihat dalam Tabel 3.4.

Tabel 3. 4 Tampilan tabel dataset teratas

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q55	q56	q57	q58	q59	q60	E-I	S-N	T-F	J-P
0	6	5	3	3	3	4	6	6	6	3	...	7.0	7.0	3	3.0	1	6.0	1	0	1	1
1	7	7	6	7	7	6	6	5	6	6	...	7.0	7.0	7	1.0	1	1.0	0	0	0	1
2	1	1	6	5	5	4	2	2	1	2	...	1.0	1.0	1	1.0	4	4.0	1	1	1	0
3	5	5	5	4	4	3	2	2	3	7	...	4.0	4.0	3	3.0	3	1.0	1	1	1	0
4	5	5	5	4	4	4	6	6	5	5	...	7.0	7.0	7	7.0	4	4.0	0	0	0	1

5 rows × 64 columns

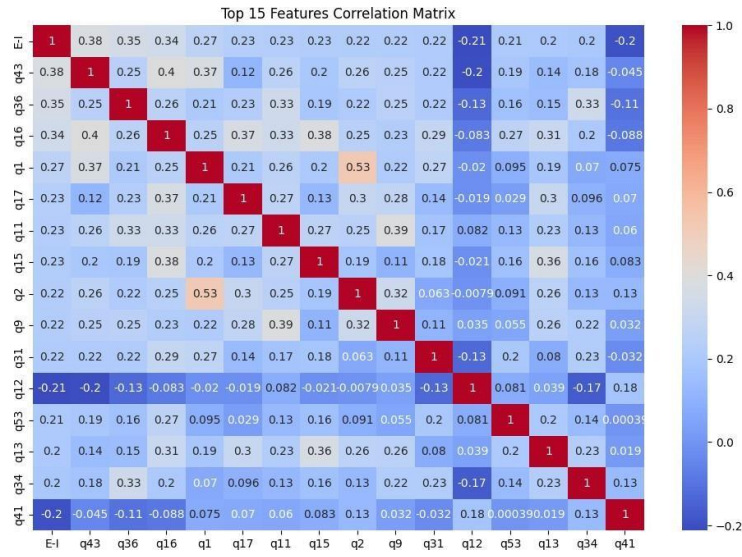
Setelah berubah menjadi *boolean* maka akan dilakukan korelasi matriks seperti yang telah disebutkan di atas. Pada tabel 3.5 menampilkan hasil perhitungan korelasi antara respons peserta dengan tipe kepribadian MBTI.

Tabel 3. 5 Hasil Correlation values

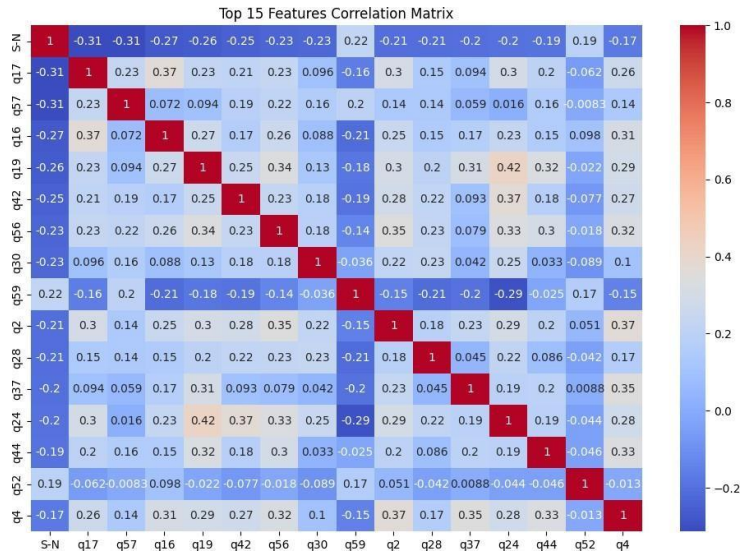
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q55	q56	q57
q1	1.000000	0.525920	0.203681	0.387820	0.297721	0.088311	0.104931	0.110558	0.215629	0.189823	...	0.011377	0.196173	0.049045
q2	0.525920	1.000000	0.304872	0.366424	0.383700	0.194758	0.288588	0.229207	0.321049	0.096031	...	-0.006308	0.345671	0.135329
q3	0.203681	0.304872	1.000000	0.241322	0.177327	0.305834	0.071243	0.188620	0.126073	0.097428	...	0.051750	0.095922	0.219754
q4	0.387820	0.366424	0.241322	1.000000	0.485781	0.070310	0.365346	0.155310	0.324936	0.131935	...	0.031316	0.318236	0.138255
q5	0.297721	0.383700	0.177327	0.485781	1.000000	0.231791	0.269376	0.208503	0.147061	0.160160	...	0.125531	0.238689	0.159924
...
q60	0.079327	0.220184	0.131016	0.319926	0.150023	-0.099503	0.307846	0.184802	0.327318	-0.059319	...	-0.115331	0.427440	0.106568
E-I	0.265862	0.224450	0.133353	0.045210	0.091388	-0.115932	0.077819	0.047274	0.223681	-0.041607	...	-0.163154	0.172676	-0.005114
S-N	-0.115373	-0.208728	-0.063403	-0.173485	-0.139165	0.110283	-0.167555	-0.130346	-0.134919	-0.008019	...	-0.030377	-0.231925	-0.309726
T-F	0.026950	-0.010025	-0.134419	0.012011	0.052773	-0.033692	-0.096106	-0.246031	-0.028535	-0.092239	...	-0.094226	0.007294	-0.064466
J-P	0.116987	0.161425	-0.003415	0.327485	0.176519	-0.046879	0.297563	0.151601	0.355806	0.134932	...	0.137255	0.348847	0.058814

64 rows × 64 columns

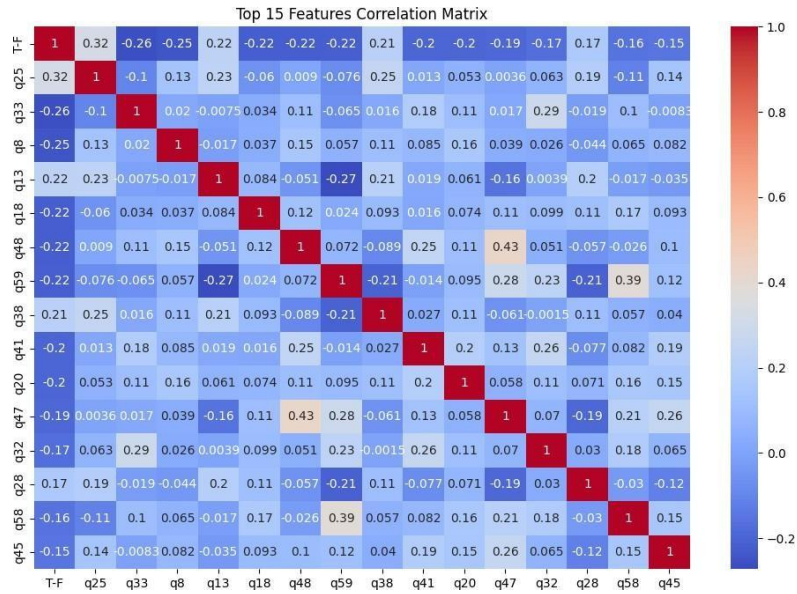
Gambar 3.5 hingga Gambar 3.8 menampilkan visualisasi distribusi kepribadian berdasarkan masing-masing dimensi MBTI, yaitu E-I, S-N, T-F, dan J-P. Visualisasi ini membantu dalam memahami pola distribusi tipe kepribadian responden serta memberikan gambaran yang lebih jelas mengenai kecenderungan karakteristik individu dalam setiap dimensi MBTI.



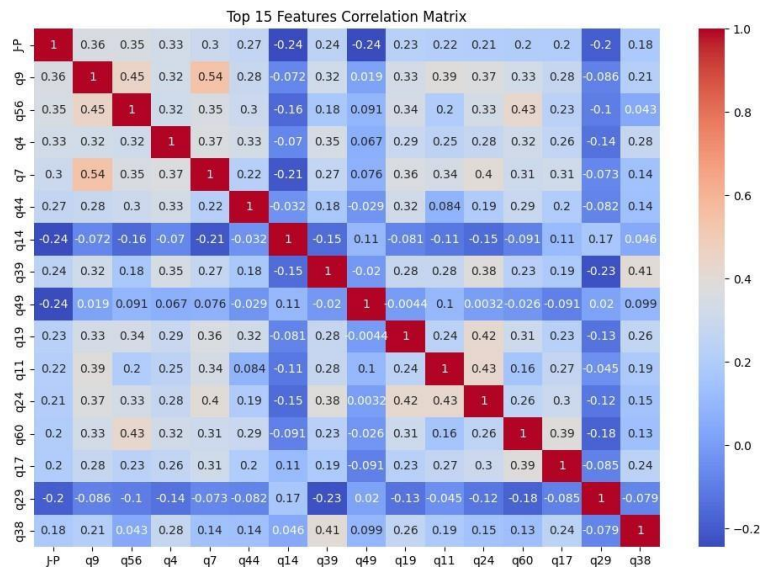
Gambar 3. 5 Visualisasi kepribadian E-I



Gambar 3. 6 Visualisasi kepribadian S-N



Gambar 3. 7 Visualisasi kepribadian T-F



Gambar 3. 8 Visualisasi kepribadian J-P

Dan pada tabel 3.6 berisi blueprint MBTI yang digunakan sebagai acuan dalam proses klasifikasi.

Tabel 3. 6 Blue Print MBTI

No	Dimensi	Indikator	Butir Item		N
			F	UF	
1	Energi (<i>Energy</i>), dimensi ini menunjukkan bagaimana seseorang berinteraksi dengan lingkungan sekitarnya.	Diukur dengan bagaimana respons mereka yang lebih luas terhadap lingkungan mereka, yang juga melibatkan orang lain bukan hanya tentang mereka. <i>Introvert</i> cenderung lebih suka melakukan aktivitas sendiri dan mudah lelah dengan interaksi sosial.	1 2 9 11 12 13 15 16 17 31 34 36 41 43 53		15
2	Pikiran (<i>Mind</i>), dimensi ini menentukan bagaimana seseorang melihat dunia dan memproses sebuah informasi.	Seseorang dengan kepribadian <i>observant</i> cenderung memiliki kebiasaan yang kuat dan fokus pada apa yang sedang atau telah terjadi. Sedangkan untuk seseorang dengan kepribadian intuitif lebih menyukai hal-hal yang baru daripada stabilitas dan cenderung fokus pada makna tersembunyi serta kemungkinan masa depan.	2 4 16 17 19 24 28 30 37 42 44 52 56 57 59		15

3	<p>Alam (<i>Nature</i>), dimensi ini menentukan bagaimana seseorang membuat keputusan dan mengatasi emosi.</p>	<p>Seseorang dengan kepribadian <i>Thinking</i> lebih mengutamakan logika daripada emosi, mereka cenderung menyembunyikan perasaan mereka dan menganggap efisiensi lebih penting daripada kerja sama. Sedangkan untuk seseorang dengan kepribadian <i>Feeling</i> lebih sensitif dan ekspresif secara emosional, mereka akan berfokus pada harmoni dan kerja sama sosial.</p>	<p>8 13 18 20 25 28 32 33 38 41 45 47 48 58 59</p>		15
4	<p>Taktik (<i>Tactics</i>), dimensi ini mencerminkan pendekatan seseorang terhadap pekerjaan, perencanaan, dan pengambilan</p>	<p>Seseorang dengan kepribadian <i>Judging</i> merupakan orang yang tegas, teliti, dan sangat terorganisir. Mereka lebih menyukai struktur dan perencanaan daripada spontanitas. Sedangkan seseorang dengan kepribadian <i>Prospecting</i> pandai berimprovokasi dan menemukan peluang. Mereka cenderung fleksibel, santai, tidak suka mengikuti arus, dan lebih suka membiarkan pilihan mereka terbuka.</p>	<p>4 7 9 11 14 17 19 24 29 38 39 44 49 56 60</p>		15

Keterangan:

F : Favorable

UF : Unfavorable

N : Total keputusan.

3.5.7 Another Correlation Analysis

Setelah itu dapat di analisis lebih jauh, jika *big correlation* memiliki lebih dari 0,1 *corr value*, maka dapat dibuatkan beberapa list pertanyaan yang memiliki *big correlation* kepada setiap dimensi. Dan didapat pertanyaan yang mempunyai *big correlation* kepada setiap dimensi adalah pertanyaan 32 dan 59. Dengan pertanyaan sebagai berikut:

32. Anda merasa bosan atau kehilangan minat saat diskusi menjadi teori
 59. Saat membuat pilihan, anda lebih cenderung mengandalkan intuisi emosional daripada penalaran logis.

Dan pada tabel 3.7 merangkum pertanyaan yang memiliki korelasi tinggi terhadap setiap dimensi.

Tabel 3. 7 Jumlah korelasi Pertanyaan untuk setiap dimensi

	q32	q59
E-I	-0.105590	-0.133797
S-N	0.143905	0.224520
T-F	-0.172063	-0.215205
J-P	-0.106335	-0.115518

Dan sebaliknya pertanyaan yang tidak mempunyai *big correlation* untuk setiap dimensi adalah pertanyaan nomor 26, 40, 50, dan 54. Dengan pertanyaan sebagai berikut.

26. Saat ada acara sosial, anda biasanya menunggu orang lain untuk memperkenalkan diri terlebih dahulu
 40. Anda jarang merasa cemas
 50. Ketika seseorang mengagumi anda, anda penasaran kapan mereka akhirnya merasa kecewa pada anda
 54. Jika anda merasa itulah keputusan yang terbaik, anda sering bertindak berdasarkan pertimbangan tersebut

Dan pada tabel 3.8 menunjukkan pertanyaan yang memiliki korelasi rendah untuk setiap dimensi kepribadian.

Tabel 3. 8 Jumlah korelasi pertanyaan untuk setiap dimensi

	q26	q40	q50	q54
E-I	-0.052935	-0.063383	0.063872	0.068835
S-N	0.081449	-0.046663	-0.090514	0.031618
T-F	0.004790	-0.002268	-0.021245	-0.090755
J-P	0.027527	0.083200	-0.065395	0.065573

3.5.8 Modelling and Testing Modelling

Dalam tahap ini pertama adalah import ensemble classifiers. Dan beberapa algoritma *Tree Based* yang digunakan adalah:

- Decision Tree Classifier*
- Random Forest Classifier*
- Extra Trees Classifier*
- Gradient Boosting Classifier*
- XGBoost Classifier*
- AdaBoost Classifier*

Dataset yang ada akan dibagi menjadi *train sets* dan *test sets*, dan data yang digunakan untuk *test sets* adalah 20% dari keseluruhan dataset yang ada. Disini dataset yang digunakan untuk *train sets* berjumlah 139 sampel dan untuk *test sets* berjumlah 35 sampel. Hasil analisis model untuk masing-masing dimensi MBTI, yaitu E-I, S-N, T-F, dan J-P, disajikan dalam Tabel 3.9 hingga Tabel 3.12. Tabel-tabel ini menampilkan performa model dalam mengklasifikasikan tipe kepribadian berdasarkan setiap dimensi yang dianalisis, menunjukkan bagaimana algoritma yang digunakan membedakan karakteristik kepribadian pada masing-masing aspek.

Tabel 3. 9 Dimensi E-1

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.614286	0.774194	0.545455	0.64
RandomForestClassifier	0.757143	0.829268	0.772727	0.8
ExtraTreesClassifier	0.771429	0.833333	0.795455	0.813953
GradientBoostingClassifier	0.771429	0.85	0.772727	0.809524
AdaBoostClassifier	0.757143	0.813953	0.795455	0.804598
xgb_classifier	0.742857	0.795455	0.795455	0.795455

Tabel 3. 10 Dimensi S-N

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.714286	0.711238	0.714286	0.710348
RandomForestClassifier	0.614286	0.601819	0.614286	0.584962
ExtraTreesClassifier	0.714286	0.718709	0.714286	0.700052
GradientBoostingClassifier	0.714286	0.714286	0.714286	0.704127
AdaBoostClassifier	0.685714	0.681651	0.685714	0.681383
xgb_classifier	0.7	0.696853	0.7	0.697112

Tabel 3. 11 Dimensi T-F

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.528571	0.516812	0.528571	0.509453
RandomForestClassifier	0.6	0.607407	0.6	0.568116
ExtraTreesClassifier	0.614286	0.613215	0.614286	0.603147
GradientBoostingClassifier	0.614286	0.615327	0.614286	0.614681
AdaBoostClassifier	0.642857	0.646531	0.642857	0.643514
xgb_classifier	0.628571	0.627143	0.628571	0.627343

Tabel 3. 12 Dimensi J-P

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.7	0.69	0.7	0.691228
RandomForestClassifier	0.657143	0.632157	0.657143	0.623176
ExtraTreesClassifier	0.7	0.692529	0.7	0.666229
GradientBoostingClassifier	0.685714	0.67013	0.685714	0.662143
AdaBoostClassifier	0.657143	0.646259	0.657143	0.649531
xgb_classifier	0.685714	0.670908	0.685714	0.654578

3.5.9 Testing

Setelah proses pemodelan untuk semua dimensi selesai, langkah berikutnya adalah pengujian. Pengujian ini menggunakan rumus *Custom Accuracy Score Formula*, dan rumus yang digunakan untuk perhitungan *Custom Accuracy Score* ditunjukkan dalam gambar 3.9.

$$\text{Custom Accuracy Score} = \frac{1}{N} \sum_{i=1}^N \left(\frac{1}{M} \sum_{j=1}^M 1(p_{ij} = t_{ij}) \right)$$

Where:

- N is the number of predictions,
- M is the number of binary labels in each prediction or true label pair,
- p_{ij} and t_{ij} are the j -th binary labels in the i -th prediction and true label, respectively,
- 1 is the indicator function, which returns 1 if $p_{ij} = t_{ij}$ (prediction matches true label) and 0 otherwise.

Gambar 3. 9 Rumus Custom Accuracy Score

Hasil akhir pengujian model untuk semua dimensi MBTI ditampilkan dalam gambar 3.10.

```
{'DecisionTreeClassifier': 0.6392857142857142,
 'RandomForestClassifier': 0.6571428571428571,
 'ExtraTreesClassifier': 0.7,
 'GradientBoostingClassifier': 0.6964285714285714,
 'AdaBoostClassifier': 0.6857142857142857,
 'xgb_classifier': 0.6892857142857143}
```

Gambar 3. 10 Hasil Testing untuk semua Dimensi

3.5.10 Explore More

Setelah memperoleh hasil dari pemodelan dan pengujian yang telah dilakukan, pada tahap ini penulis hanya akan menggunakan pertanyaan-pertanyaan yang memiliki korelasi tinggi untuk setiap dimensi. Pertanyaan yang memiliki korelasi tinggi dengan masing-masing dimensi ditampilkan dalam gambar 3.11.

```
Number of High Corr Question of E-I: 29
Number of High Corr Question of S-N: 33
Number of High Corr Question of T-F: 25
Number of High Corr Question of J-P: 31
```

Gambar 3. 11 Pertanyaan yang memiliki korelasi tinggi untuk setiap dimensi

Setelah mendapatkan pertanyaan-pertanyaan tersebut maka dapat dilakukan pemodelan kembali. Hasil pengujian model klasifikasi untuk masing-masing dimensi MBTI, yaitu E-I, S-N, T-F, dan J-P, disajikan dalam Tabel 3.13 hingga Tabel 3.16. Tabel-tabel ini menampilkan performa model berdasarkan masing-masing dimensi kepribadian, menunjukkan efektivitas pendekatan yang digunakan dalam membedakan karakteristik individu dalam setiap kategori MBTI.

Tabel 3. 13 Dimensi E-I

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.6	0.735294	0.568182	0.641026
RandomForestClassifier	0.8	0.840909	0.840909	0.840909
ExtraTreesClassifier	0.814286	0.897436	0.795455	0.843373
GradientBoostingClassifier	0.757143	0.829268	0.772727	0.8
AdaBoostClassifier	0.757143	0.846154	0.75	0.795181
xgb_classifier	0.8	0.875	0.795455	0.833333

Tabel 3. 14 Dimensi S-N

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.642857	0.583333	0.482759	0.528302
RandomForestClassifier	0.671429	0.714286	0.344828	0.465116
ExtraTreesClassifier	0.7	0.681818	0.517241	0.588235
GradientBoostingClassifier	0.7	0.666667	0.551724	0.603774
AdaBoostClassifier	0.657143	0.6	0.517241	0.555556
xgb_classifier	0.7	0.7	0.482759	0.571429

Tabel 3. 15 Dimensi T-F

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.542857	0.575	0.605263	0.589744
RandomForestClassifier	0.671429	0.674419	0.763158	0.716049
ExtraTreesClassifier	0.685714	0.7	0.736842	0.717949
GradientBoostingClassifier	0.628571	0.676471	0.605263	0.638889
AdaBoostClassifier	0.685714	0.75	0.631579	0.685714
xgb_classifier	0.614286	0.627907	0.710526	0.666667

Tabel 3. 16 Dimensi J-P

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.628571	0.686275	0.777778	0.729167
RandomForestClassifier	0.728571	0.740741	0.888889	0.808081
ExtraTreesClassifier	0.714286	0.727273	0.888889	0.8
GradientBoostingClassifier	0.714286	0.745098	0.844444	0.791667
AdaBoostClassifier	0.714286	0.755102	0.822222	0.787234
xgb_classifier	0.742857	0.764706	0.866667	0.8125

Setelah dilakukan modeling maka kembali dilakukan *Testing* untuk keseluruhan dimensi menggunakan pertanyaan-pertanyaan yang memiliki korelasi paling tinggi. Dan hasil pengujian model menggunakan pertanyaan yang memiliki korelasi tinggi dapat dilihat dalam gambar 3.12.

```
{'DecisionTreeClassifier': 0.6035714285714285,
'RandomForestClassifier': 0.7178571428571429,
'ExtraTreesClassifier': 0.7285714285714285,
'GradientBoostingClassifier': 0.7,
'AdaBoostClassifier': 0.7035714285714286,
'xgb_classifier': 0.7142857142857143}
```

Gambar 3. 12 Hasil dari Testing

BAB IV

HASIL DAN PEMBAHASAN

4.1 Analisis Pengumpulan Data

Pengumpulan data dilakukan dengan menggunakan dua tahapan, yang pertama responden harus menjawab tes kepribadian melalui website *16Personalities* yang dapat di akses melalui link berikut <https://www.16Personalities.com/free-personality-test>.

Gambar 4.1 memperlihatkan tampilan pertanyaan yang ada dalam tes 16Personalities.

Gambar 4. 1 Tampilan pertanyaan pada Website 16Personalities

Setelah menjawab pada website tersebut responden diharapkan untuk menjawab dengan jawaban yang sama pada *google form* yang telah disediakan. *Google form* dapat di akses melalui link <https://forms.gle/SGiMR2kwTjNU8Y3j7>.

Tampilan pertanyaan dalam google form yang digunakan dalam pengumpulan data dapat dilihat dalam gambar 4.2.

Gambar 4. 2 Tampilan pertanyaan teratas pada google form

Pengumpulan data ini terdiri dari individu yang berpartisipasi dalam pengisian kuesioner MBTI yang dipilih secara random. Di bawah ini merupakan pertanyaan-pertanyaan yang ditanyakan kepada responden:

1. Anda rutin mendapatkan teman baru.
2. Ide-ide kompleks dan baru lebih membuat Anda bersemangat daripada yang sederhana dan terang-terangan.
3. Anda biasanya merasa lebih terpengaruh oleh apa yang beresonansi secara emosional dengan Anda daripada oleh argumen faktual.
4. Ruang hidup dan kerja Anda bersih dan terorganisir.
5. Anda biasanya tetap tenang sekalipun sedang dalam banyak tekanan.
6. Anda merasa konsep berjejaring atau memperkenalkan diri Anda kepada orang asing sangat menakutkan.
7. Anda memprioritaskan dan merencanakan tugas dengan efektif, seringkali menyelesaikannya jauh sebelum tenggat waktu.
8. Kisah dan penghayatan emosi seseorang mempunyai dampak lebih signifikan bagi Anda daripada angka atau data.
9. Anda suka menggunakan alat pengorganisasian seperti jadwal dan daftar tugas.
10. Kesalahan kecil pun dapat menyebabkan Anda ragu dengan kemampuan dan pengetahuan Anda.
11. Anda merasa nyaman untuk mendekati seseorang yang menurut Anda menarik dan memulai percakapan.
12. Anda tidak terlalu tertarik dalam diskusi tentang berbagai interpretasi karya kreatif.
13. Anda mengutamakan fakta daripada perasaan orang ketika menentukan langkah tindakan.
14. Anda sering membiarkan hari berlalu tanpa jadwal sama sekali.
15. Anda jarang mengkhawatirkan kesan orang terhadap diri Anda.
16. Anda menikmati berpartisipasi dalam kegiatan berbasis tim.
17. Anda menikmati bereksperimen dengan pendekatan baru dan belum teruji.
18. Anda lebih mengutamakan untuk bersikap peka daripada sepenuhnya jujur.
19. Anda secara aktif mencari pengalaman dan area pengetahuan baru untuk dijelajahi.
20. Anda mudah merasa khawatir bahwa segala hal akan berubah menjadi lebih buruk.

21. Anda lebih menikmati hobi atau aktivitas individual daripada yang perlu dilakukan berkelompok.
22. Anda tidak dapat membayangkan diri Anda menulis cerita fiksi sebagai mata pencaharian.
23. Anda lebih memilih efisiensi dalam pengambilan keputusan, meskipun itu berarti mengabaikan beberapa aspek emosional.
24. Anda lebih suka menyelesaikan tugas-tugas terlebih dahulu sebelum bersantai.
25. Saat terjadi perselisihan, Anda lebih mengutamakan membuktikan pendapat Anda daripada menjaga perasaan orang lain.
26. Saat ada acara sosial, Anda biasanya menunggu orang lain untuk memperkenalkan diri terlebih dahulu.
27. Suasana hati Anda bisa berubah dengan sangat cepat.
28. Anda tidak mudah terpengaruh oleh argumen yang berlandaskan emosi.
29. Anda sering menunda pekerjaan hingga detik-detik terakhir.
30. Anda senang berdebat tentang dilema etika.
31. Anda biasanya lebih suka berada di sekitar orang lain daripada sendirian.
32. Anda merasa bosan atau kehilangan minat saat diskusi menjadi terlalu teoretis.
33. Saat ada pertentangan antara kenyataan dan perasaan, Anda biasanya lebih memilih untuk mengikuti kata hati.
34. Bagi Anda, menjaga konsistensi jadwal kerja atau belajar adalah sebuah tantangan.
35. Anda jarang mempertanyakan keputusan yang telah Anda buat.
36. Teman-teman Anda menggambarkan diri Anda sebagai orang yang energik dan ramah.
37. Anda tertarik pada berbagai bentuk ekspresi kreatif, seperti menulis.
38. Anda biasanya membuat pilihan berdasarkan fakta objektif daripada kesan emosional.
39. Anda suka menyusun daftar tugas sehari-hari.
40. Anda jarang merasa cemas.
41. Anda menghindari melakukan panggilan telepon.
42. Anda senang menggali ide dan sudut pandang yang kurang dikenal.
43. Anda mudah berinteraksi dengan orang yang baru dikenal.
44. Jika rencana Anda terganggu, prioritas utama Anda adalah kembali ke rencana tersebut secepat mungkin.
45. Anda masih merasa terganggu dengan kesalahan yang Anda lakukan di masa lalu.

46. Anda tidak terlalu tertarik untuk membahas teori tentang masa depan dunia.
47. Emosi lebih dominan mengendalikan diri Anda.
48. Saat mengambil keputusan, Anda lebih mempertimbangkan perasaan orang yang terkena dampak keputusan itu daripada pertimbangan yang paling logis atau efisien.
49. Gaya kerja pribadi Anda lebih mendekati letupan energi spontan daripada upaya yang teratur dan konsisten.
50. Ketika seseorang mengagumi Anda, Anda penasaran kapan mereka akhirnya merasa kecewa pada Anda.
51. Anda akan menyukai pekerjaan yang mengharuskan untuk lebih banyak bekerja sendiri.
52. Anda percaya bahwa merenungkan pertanyaan filosofis abstrak hanya membuang-buang waktu.
53. Anda lebih tertarik pada suasana ramai dan sibuk daripada tempat tenang dan akrab.
54. Jika Anda merasa itulah keputusan yang terbaik, Anda sering bertindak berdasarkan pertimbangan tersebut
55. Anda sering merasa kewalahan.
56. Anda menyelesaikan berbagai hal menurut metode tanpa melewatkan langkah apa pun.
57. Anda lebih menyukai tugas yang memerlukan solusi kreatif daripada mengikuti prosedur yang sudah ditentukan.
58. Saat membuat pilihan, Anda lebih cenderung mengandalkan intuisi emosional daripada penalaran logis.
59. Anda kesulitan memenuhi tenggat.
60. Anda yakin bahwa segalanya akan berjalan dengan baik.

Disediakan tujuh pilihan jawaban untuk setiap pertanyaan serta skor untuk masing-masing pilihan jawaban. Pada tabel 4.1 menampilkan pilihan jawaban yang disediakan untuk setiap pertanyaan.

Tabel 4. 1 Jawaban pertanyaan

Pilihan Jawaban	Keterangan	Skor	
		<i>Favorable</i>	<i>Unfavorable</i>
1	Sangat Tidak Setuju	1	7
2	Tidak Setuju	2	6
3	Cukup Tidak Setuju	3	5
4	Netral	4	4
5	Cukup Setuju	5	3
6	Setuju	6	2
7	Sangat Setuju	7	1

4.1.1 Analisis Informasi Data

Untuk penelusuran informasi data dapat menggunakan perintah `df_init.info()` yang akan menampilkan ringkasan yang ada pada dataset yang sudah dikumpulkan. Informasi data yang dikumpulkan dalam penelitian ini disajikan dalam tabel 4.2.

Tabel 4. 2 Tabel Hasil Informasi Data

Jumlah kolom	Jumlah baris	Nomor baris awal	Nomor baris akhir	Jenis data	Jumlah baris float64	Jumlah baris int64	Jumlah baris object	Jumlah memori (KB)
64	174	0	173	float 64 int6 4	8	53	3	87.1+
				obje ct				

Tipe kepribadian MBTI terbagi menjadi 16 tipe. Dengan menggunakan perintah `df_splitting['MBTI'].value_counts()` akan menampilkan jumlah responden untuk tiap tipe kepribadian. Distribusi jumlah responden yang berpartisipasi dalam penelitian ditampilkan dalam gambar 4.3.

```

Jumlah tiap kepribadian MBTI:
MBTI
ENTJ    28
ENFJ    24
INFJ    14
ESTP    13
ENTP    13
INTJ    12
ISFJ    12
ISTJ    11
ESFJ    10
ISTP    9
ESFP    6
ENFP    5
INTP    5
ISFP    5
ESTJ    4
INFP    3
Name: count, dtype: int64

```

Gambar 4. 3 Jumlah Responden

4.2 Analisis Pre-Processing Data

Pada tahap ini analisis dilakukan untuk membahas mengenai hasil dari pengumpulan dan pengolahan data yang telah dilakukan. *Pre-Processing Data* yang dilakukan bertujuan untuk menyiapkan data agar siap digunakan untuk pemodelan. Berikut ini merupakan langkah-langkah *Pre-Processing Data* yang telah dilakukan:

4.2.1 Analisis Hasil Penanganan *Missing Value*

Pada tahap ini analisis pertama adalah memeriksa ada atau tidak angka null pada dataset yang digunakan. Menggunakan perintah

`df_rename.isnul().sum().sum()` dapat menampilkan jumlah data yang memiliki angka null. Jumlah data yang memiliki nilai null sebelum dilakukan pre-processing dapat dilihat dalam gambar 4.4.

```
np.int64(9)
```

Gambar 4. 4 Jumlah Angka Null

Setelah mengetahui jumlah tersebut langkah selanjutnya adalah mengisi *Missing Value* dengan menggunakan nilai Modus pada setiap kolom. Seperti terlihat pada gambar 4.5.

```
df_miss_handled = df_rename.fillna(df_rename.mode().iloc[0])
```

Gambar 4. 5 Kode untuk mengisi nilai null dengan nilai modus

`df_rename.mode()` merupakan perintah untuk mengembalikan nilai modus untuk setiap kolom yang memiliki angka null. `.iloc[0]` merupakan perintah untuk mengambil modus pertama pada setiap kolom. `df_rename.fillna()` merupakan perintah untuk mengganti angka null dengan nilai yang sudah di tentukan sebelumnya.

Kemudian menampilkan kembali jumlah data yang memiliki angka null dengan menggunakan perintah

`df_miss_handled.isnull().sum().sum()`. Jumlah dari angka null setelah dilakukan penanganan dapat dilihat dalam gambar 4.6.

```
np.int64(0)
```

Gambar 4. 6 Jumlah Angka Null

4.3 Analisis Hasil *Split Label*

Pada website *16Personalities* tes kepribadian yang di lakukan akan menghasilkan komponen tambahan (-A/-T), sebagai contoh ENFP-A, INTJ-T. dan pada penelitian ini komponen tersebut tidak akan digunakan. Langkah pertama yang dilakukan adalah memisah antara komponen pertama (ENFP) dengan komponen kedua (-A). kode perintah yang digunakan untuk memisahkan komponen pertama dan kedua dari data kepribadian ditampilkan dalam gambar 4.7.

```
dfSplitted['MBTI'] =
dfSplitted['MBTI'].apply(lambda x: x.split('-')[0])
```

Gambar 4. 7 Kode perintah untuk memisah komponen pertama dan kedua

`dfSplitted['MBTI'].apply` digunakan untuk menerapkan fungsi lambda untuk setiap nilai dalam kolom MBTI. `lambda x:` merupakan fungsi lambda yang didefinisikan untuk memproses setiap elemen pada MBTI, `x.split('-')[0]` merupakan perintah untuk membagi string 'x' menjadi list berdasarkan pemisah '-'. Dan angka 0 adalah perintah untuk mengambil elemen pertama dari daftar list yang dihasilkan.

Kemudian menampilkan hasil dari *Split Label* yang telah dilakukan dengan menggunakan perintah `dfSplitted.MBTI.tolist()`. Hasil split label dalam dataset ditunjukkan dalam gambar 4.8.

```

['ENTJ',
 'INFJ',
 'ESTP',
 'ESTP',
 'INFJ',
 'INTJ',
 'ENFJ',
 'ISFJ',
 'INTJ',
 'ENFP',
 'ISFJ',
 'INTP',
 'ISFJ',
 'INFJ',
 'INTP',
 'ENTJ',
 'ENFP',
 'ISTJ',
 'ESTP',
 'INTJ',
 'ENFJ',
 'ISFJ',
 'INTJ',
 'INFP',
 'INFJ',
 ...
 'INFJ',
 'ENFJ',

```

Gambar 4. 8 Hasil Split Label

Setelah pemisahan diatas, label MBTI akan di pisah lagi menjadi 4 komponen. Kode yang digunakan untuk memisahkan label ke dalam empat bagian ditampilkan dalam gambar 4.9.

```

def split_mbti(label):
    return pd.Series([label[0], label[1], label[2], label[3]])

```

Gambar 4. 9 Kode perintah untuk memisahkan label menjadi 4 bagian

`def split_mbti(label)` merupakan perintah untuk menerima argumen dan pada kode tersebut hanya menerima argumen label. `return pd.Series` merupakan fungsi untuk mengekstrak setiap karakter dari label MBTI dan dikembalikan dalam bentuk `pandas.series`. sedangkan label 0-3 menunjukkan list dari karakter-karakter dari MBTI. Dan contoh lima data teratas dari dataset yang digunakan dapat dilihat dalam tabel 4.3.

Tabel 4. 3 Data Lima Teratas dari Dataset

	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q55	q56	q57	q58	q59	q60	E-I	S-N	T-F	J-P
0	6	5	3	3	3	4	6	6	6	3	...	7.0	7.0	3	3.0	1	6.0	E	N	T	J
1	7	7	6	7	7	6	6	5	6	6	...	7.0	7.0	7	1.0	1	1.0	I	N	F	J
2	1	1	6	5	5	4	2	2	1	2	...	1.0	1.0	1	1.0	4	4.0	E	S	T	P
3	5	5	5	4	4	3	2	2	3	7	...	4.0	4.0	3	3.0	3	1.0	E	S	T	P
4	5	5	5	4	4	4	6	6	5	5	...	7.0	7.0	7	7.0	4	4.0	I	N	F	J

5 rows × 64 columns

4.4 Analisis Hasil Statistic Summary

Statistic Summary dilakukan untuk memahami karakteristik dari dataset yang digunakan.

Pad tabel 4.4 berisi ringkasan statistik sari dataset yang telas di analisis.

Tabel 4. 4 Hasil *Statistic Summary*

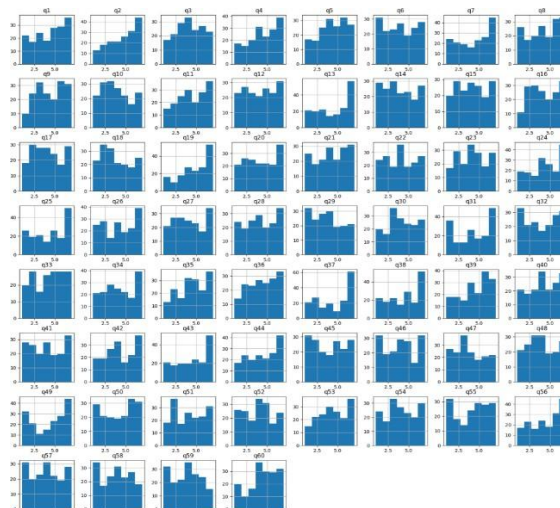
	q1	q2	q3	q4	q5	q6	q7	q8	q9	q10	...	q51
count	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	174.000000	...	174.000000
mean	4.402299	4.712644	4.143678	4.557471	4.362069	3.948276	4.442529	4.287356	4.396552	3.804598	...	4.091954
std	2.053953	1.964471	1.864496	1.966372	1.887270	2.091179	2.173042	2.084393	1.895596	1.937396	...	2.037976
min	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	1.000000	...	1.000000
25%	3.000000	3.000000	3.000000	3.000000	3.000000	2.000000	2.000000	3.000000	3.000000	2.000000	...	2.000000
50%	5.000000	5.000000	4.000000	5.000000	4.000000	4.000000	5.000000	4.000000	4.000000	4.000000	...	4.000000
75%	6.000000	6.750000	6.000000	6.000000	6.000000	6.000000	7.000000	6.000000	6.000000	5.000000	...	6.000000
max	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	7.000000	...	7.000000

8 rows × 60 columns

Berdasarkan tabel 4.4 di atas dapat dilihat *Statistic Summary* untuk setiap pertanyaan. Pada q1 memiliki jumlah data 174 yang didalamnya terdapat nilai minimal yaitu 1 dan nilai maksimal yaitu 7. Variabel ini juga memiliki nilai mean atau nilai rata-rata yaitu 4,402299 dan memiliki nilai standar deviasi yaitu 2,053953 yang mana nilai ini masih berada di bawah dari nilai mean, sehingga menyebabkan kurangnya persebaran dan keragaman pada variabel ini. Variabel ini memiliki nilai kuartil yang terdiri dari kuartil 1 (25%) memiliki nilai 3 yang artinya sebanyak 25% dari nilai yang ada pada kolom q1 adalah 3 atau bisa lebih rendah. Kuartil 2 (50%) memiliki nilai 5 yang artinya setengah dari nilai pada kolom q1 adalah 5 atau bisa lebih rendah. Yang terakhir kuartil 3 (75%) memiliki nilai 6 yang artinya 75% nilai pada kolom q1 adalah 6 atau bisa lebih rendah. Begitupun seterusnya untuk variabel pertanyaan yang lain.

4.5 Analisis hasil Response Distribution

Pada tahap ini yang dilakukan adalah menampilkan *Histogram* untuk setiap jawaban responden di setiap pertanyaan. *Histogram* distribusi jawaban untuk setiap pertanyaan kuesioner dapat dilihat dalam gambar 4.10.

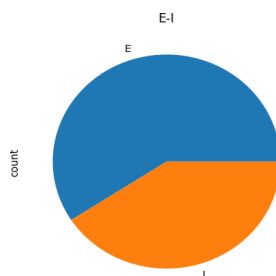


Gambar 4. 10 *Histogram* Setiap Pertanyaan

Dari histogram di atas dapat dilihat jumlah jawaban responden untuk setiap pertanyaan. *Histogram* menampilkan jawaban dari 1-7 sesuai jawaban yang di sediakan untuk setiap pertanyaan.

4.6 Analisis Hasil Personality Distribution

Pada tahap ini dilakukan perbandingan jumlah responden untuk setiap dimensi. Hasilnya akan ditampilkan dalam bentuk *Pie Chart*, yang akan memudahkan melihat perbandingan untuk setiap dimensi kepribadian. Gambar 4.11 menunjukkan pie chart untuk dimensi E-I dan jumlah responden dimensi E-I dapat dilihat dalam gambar 4.12.



Gambar 4. 11 *Pie Chart* untuk Dimensi E-I

```

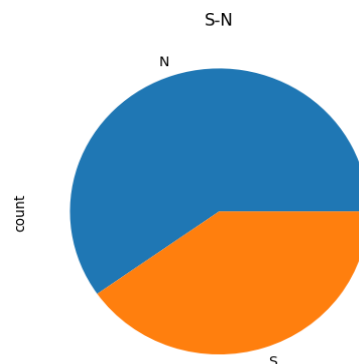
Jumlah untuk dimensi E-I:
E-I
E    103
I     71
Name: count, dtype: int64

```

Gambar 4. 12 Jumlah Responden untuk Dimensi E-I

Berdasarkan gambar diatas jumlah responden yang memilki kepribadian *Ekstrovert* (E) sebanyak 103 orang, sedangkan jumlah responden yang memilki kepribadian *Introvert* (I) sebanyak 71 orang. Dari jumlah ini dapat diketahui bahwa responden yang memilki kepribadian *Ekstrovert* lebih banyak 32 responden daripada responden yang memilki kepribadian *Introvert*.

Gambar 4.13 menunjukkan pie chart untuk dimensi S-N dan jumlah responden dimensi S-N dapat dilihat dalam gambar 4.14.



Gambar 4. 13 Pie Chart untuk Dimensi S-N

```

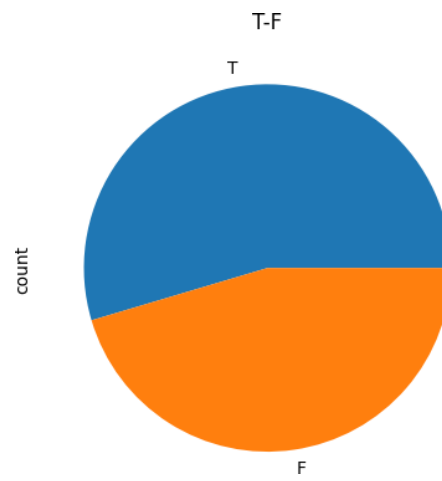
Jumlah untuk dimensi S-N:
S-N
N    104
S     70
Name: count, dtype: int64

```

Gambar 4. 14 Jumlah Responden untuk Dimensi S-N

Berdasarkan gambar diatas jumlah responden yang memilki kepribadian *iNtuition* (N) sebanyak 104 orang, sedangkan jumlah responden yang memilki kepribadian *Sensing* (S) sebanyak 70 orang. Dari jumlah ini dapat diketahui bahwa responden yang memilki kepribadian *iNtuition* lebih banyak 33 responden daripada responden yang memilki kepribadian *iNtuition*.

Gambar 4.15 menunjukkan pie chart untuk dimensi T-F dan jumlah responden dimensi T-F dapat dilihat dalam gambar 4.16.

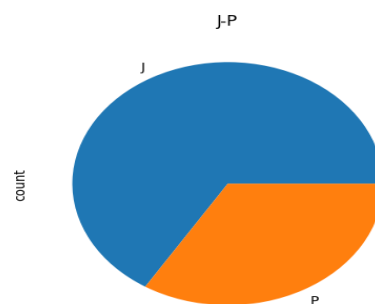


Gambar 4. 15 *Pie Chart* untuk Dimensi T-F

```
Jumlah untuk dimensi T-F:
T-F
T    95
F    79
Name: count, dtype: int64
```

Gambar 4. 16 Jumlah Responden untuk Dimensi T-F

Berdasarkan gambar diatas jumlah responden yang memiliki kepribadian *Thinking* (T) sebanyak 95 orang, sedangkan jumlah responden yang memiliki kepribadian *Feeling* (F) sebanyak 79 orang. Dari jumlah ini dapat diketahui bahwa responden yang memiliki kepribadian *Thinking* lebih banyak 16 responden daripada responden yang memiliki kepribadian *Feeling* dengan perbandingan tidak terlalu besar seperti dimensi sebelumnya.



Gambar 4. 17 *Pie Chart* untuk Dimensi J-P

```

Jumlah untuk dimensi J-P:
J-P
J      115
P       59
Name: count, dtype: int64

```

Gambar 4. 18 Jumlah Responden untuk Dimensi J-P

Gambar 4.17 menunjukkan pie chart untuk dimensi J-P dan jumlah responden dimensi J-P dapat dilihat dalam gambar 4.18.

Berdasarkan gambar 4.17 dan gambar 4.18 jumlah responden yang memiliki kepribadian *Judging* (J) sebanyak 115 orang, sedangkan jumlah responden yang memiliki kepribadian *Perceiving* (P) sebanyak 59 orang. Dari jumlah ini dapat diketahui bahwa responden yang memiliki kepribadian *Judging* lebih banyak 56 responden daripada responden yang memiliki kepribadian *Perceiving* dan merupakan perbandingan dengan jumlah paling banyak diantara dimensi lainnya.

4.7 Analisis hasil Correlation Value

Setelah mengubah variabel *personalities* menjadi *boolean*, maka akan dilakukan *correlation matrix*. Korelasi yang digunakan yaitu *Pearson Correlation*, rumus *Pearson Correlation* yang digunakan dalam penelitian ditampilkan dalam gambar 4.19. dan kode untuk menghitung korelasi menggunakan *Pearson* ditampilkan dalam gambar 4.20.

$$r = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sqrt{\sum(X_i - \bar{X})^2 \sum(Y_i - \bar{Y})^2}}$$

Gambar 4. 19 Rumus Pearson Correlation

```

corr_matrix = df_binary.corr(method='pearson')
corr_matrix

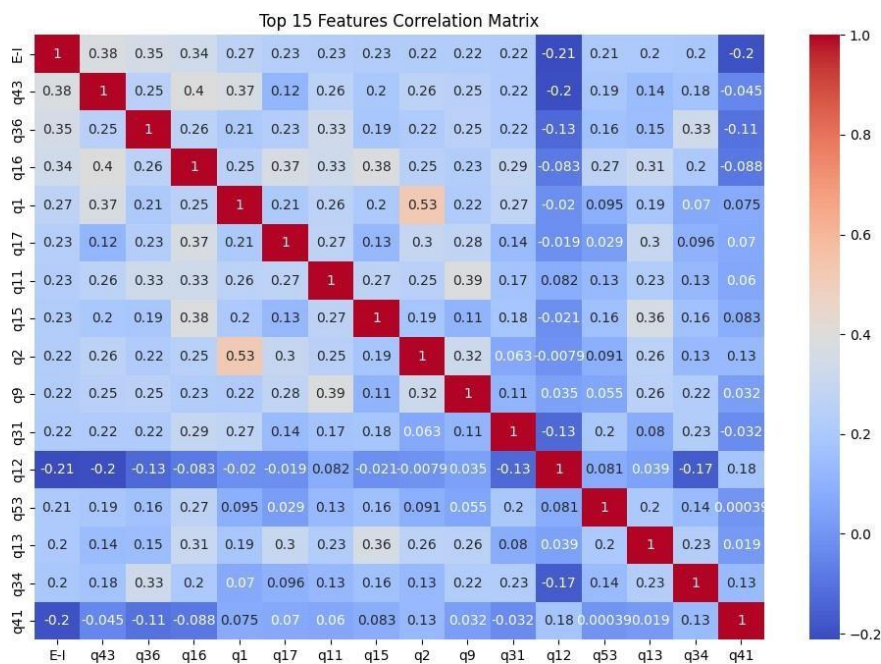
```

Gambar 4. 20 Kode perintah untuk menghitung korelasi menggunakan Pearson

`df_binary.corr()` merupakan perintah untuk menghitung *correlation value*, sedangkan `method='pearson'` merupakan perintah spesifik menggunakan *Pearson Correlation*. Dan di bawah ini merupakan visualisasi dan penjelasan tentang *correlation matrix* untuk setiap dimensi.

4.7.1 Correlation Value dimensi E-I

Visualisasi kepribadian berdasarkan dimensi E-I dapat dilihat dalam gambar 4.21.



Gambar 4. 21 Visualisasi kepribadian E-I

Pada dimensi E-I pertanyaan yang ditanyakan mewakili bagaimana seseorang berinteraksi dengan lingkungan sekitar. *Ekstrovert* merupakan kepribadian yang bisa mendapatkan energi dari dunia luar yang bisa berupa orang atau pengalaman, mereka juga memfokuskan energi ke luar dalam tindakan. Sedangkan *Introvert* merupakan kepribadian yang mendapatkan energi dari dunia batin yang berupa refleksi dan juga pikiran. Mereka lebih memfokuskan pada dunia mereka sendiri dan kurang memperhatikan sekitar.

Korelasi yang memiliki nilai positif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah *Ekstrovert*. Sebaliknya korelasi yang memiliki nilai negatif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah *Introvert*. Dan di bawah ini merupakan 15 pertanyaan ter atas yang memiliki korelasi paling besar untuk dimensi E-I:

q43: Anda mudah berinteraksi dengan orang yang baru dikenal.

q36: Teman-teman Anda menggambarkan diri Anda sebagai orang yang energik dan ramah.

q16: Anda menikmati berpartisipasi dalam kegiatan berbasis tim.

q1: Anda rutin mendapatkan teman baru.

q17: Anda menikmati bereksperimen dengan pendekatan baru dan belum teruji.

q11: Anda merasa nyaman untuk mendekati seseorang yang menurut Anda menarik dan memulai percakapan.

q15: Anda jarang mengkhawatirkan kesan orang terhadap diri Anda.

q2: Ide-ide kompleks dan baru lebih membuat Anda bersemangat daripada yang sederhana dan terang-terangan.

q9: Anda suka menggunakan alat pengorganisasian seperti jadwal dan daftar tugas.

q31: Anda biasanya lebih suka berada di sekitar orang lain daripada sendirian.

q12: Anda tidak terlalu tertarik dalam diskusi tentang berbagai interpretasi karya kreatif.

q53: Anda lebih tertarik pada suasana ramai dan sibuk daripada tempat tenang dan akrab.

q13: Anda mengutamakan fakta daripada perasaan orang ketika menentukan langkah tindakan.

q34: Bagi Anda, menjaga konsistensi jadwal kerja atau belajar adalah sebuah tantangan.

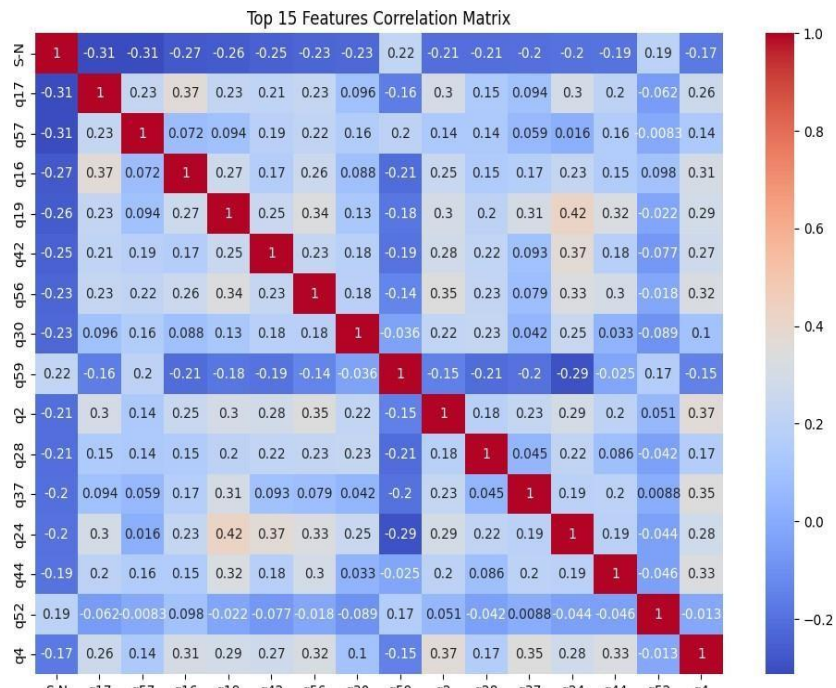
q41: Anda menghindari melakukan panggilan telepon.

4.7.2 Correlation Value dimensi S-N

Pada dimensi S-N pertanyaan yang ditanyakan mewakili bagaimana seseorang melihat dunia dan memproses sebuah informasi. Sensing merupakan kepribadian yang lebih menyukai informasi nyata yang berasal dari kelima indra, mereka berfokus pada apa yang nyata. Sedangkan iNtuition lebih menyukai informasi yang berasal dari asosiasi, mereka lebih berfokus pada kemungkinan dan apa yang mungkin terjadi.

Korelasi yang memiliki nilai positif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah Sensing. Sebaliknya korelasi yang memiliki nilai negatif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah iNtuition.

Distribusi kepribadian berdasarkan dimensi T-F ditampilkan dalam gambar 4.22.



Gambar 4. 22 Visualisasi Kepribadian S-N 1

Dan di bawah ini merupakan 15 pertanyaan ter atas yang memiliki korelasi paling besar untuk dimensi S-N:

q17: Anda menikmati bereksperimen dengan pendekatan baru dan belum teruji.

q57: Anda lebih menyukai tugas yang memerlukan solusi kreatif daripada mengikuti prosedur yang sudah ditentukan.

q16: Anda menikmati berpartisipasi dalam kegiatan berbasis tim.

q19: Anda secara aktif mencari pengalaman dan area pengetahuan baru untuk dijelajahi.

q42: Anda senang menggali ide dan sudut pandang yang kurang dikenal.

q56: Anda menyelesaikan berbagai hal menurut metode tanpa melewatkan langkah apa pun.

q30: Anda senang berdebat tentang dilema etika.

q59: Anda kesulitan memenuhi tenggat.

q2: Ide-ide kompleks dan baru lebih membuat Anda bersemangat daripada yang sederhana dan terang-terangan.

q28: Anda tidak mudah terpengaruh oleh argumen yang berlandaskan

emosi.

q37: Anda tertarik pada berbagai bentuk ekspresi kreatif, seperti menulis.

q24: Anda lebih suka menyelesaikan tugas-tugas terlebih dahulu sebelum bersantai.

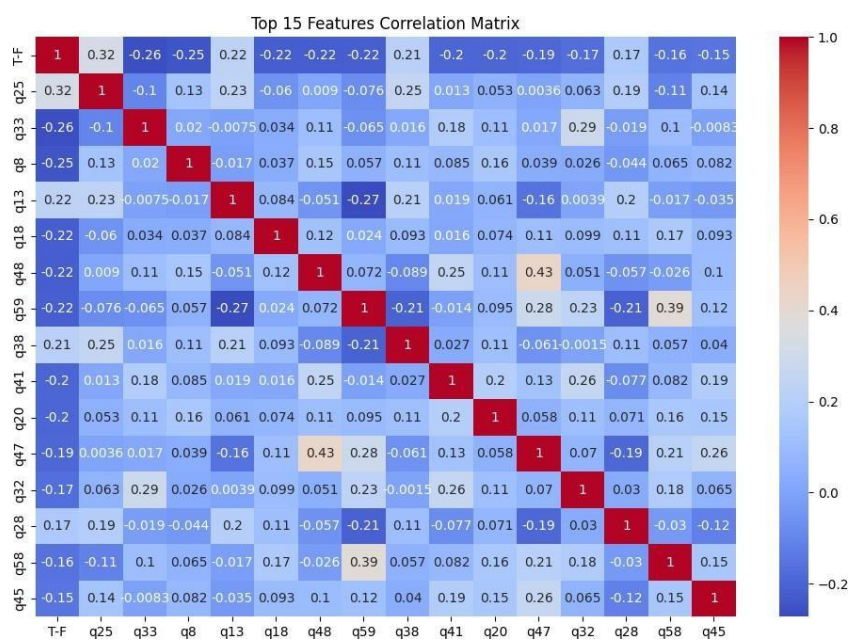
q44: Jika rencana Anda terganggu, prioritas utama Anda adalah kembali ke rencana tersebut secepat mungkin.

q52: Anda percaya bahwa merenungkan pertanyaan filosofis abstrak hanya membuang-buang waktu.

q4: Ruang hidup dan kerja Anda bersih dan terorganisir.

4.7.3 Correlation Value dimensi T-F

Distribusi kepribadian berdasarkan dimensi T-F dapat dilihat dalam gambar 4.23.



Gambar 4. 23 Visualisasi Kepribadian T-F

Pada dimensi T-F pertanyaan yang ditanyakan mewakili bagaimana seseorang membuat keputusan dan mengatasi emosi. *Thinking* merupakan kepribadian yang lebih cenderung untuk melangkah keluar dari situasi untuk menganalisisnya tanpa melibatkan perasaan, dan lebih suka membuat keputusan berdasarkan logika objektif. Sedangkan untuk *Feeling* merupakan kepribadian yang memilih melangkah ke dalam situasi untuk mempertimbangkan nilai berdasarkan perasaannya sendiri dan perasaan orang lain.

Korelasi yang memiliki nilai positif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah *Thinking*. Sebaliknya korelasi yang memiliki nilai negatif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah *Feeling*. Dan di bawah ini merupakan 15 pertanyaan ter atas yang memiliki korelasi paling besar untuk dimensi T-F:

q25: Saat terjadi perselisihan, Anda lebih mengutamakan membuktikan pendapat Anda daripada menjaga perasaan orang lain.

q33: Saat ada pertentangan antara kenyataan dan perasaan, Anda biasanya lebih memilih untuk mengikuti kata hati.

q8: Kisah dan penghayatan emosi seseorang mempunyai dampak lebih signifikan bagi Anda daripada angka atau data.

q13: Anda mengutamakan fakta daripada perasaan orang ketika menentukan langkah tindakan.

q18: Anda lebih mengutamakan untuk bersikap peka daripada sepenuhnya jujur.

q48: Saat mengambil keputusan, Anda lebih mempertimbangkan perasaan orang yang terkena dampak keputusan itu daripada pertimbangan yang paling logis atau efisien.

q59: Anda kesulitan memenuhi tenggat.

q38: Anda biasanya membuat pilihan berdasarkan fakta objektif daripada kesan emosional.

q41: Anda menghindari melakukan panggilan telepon.

q20: Anda mudah merasa khawatir bahwa segala hal akan berubah menjadi lebih buruk.

q47: Emosi lebih dominan mengendalikan diri Anda.

q32: Anda merasa bosan atau kehilangan minat saat diskusi menjadi terlalu

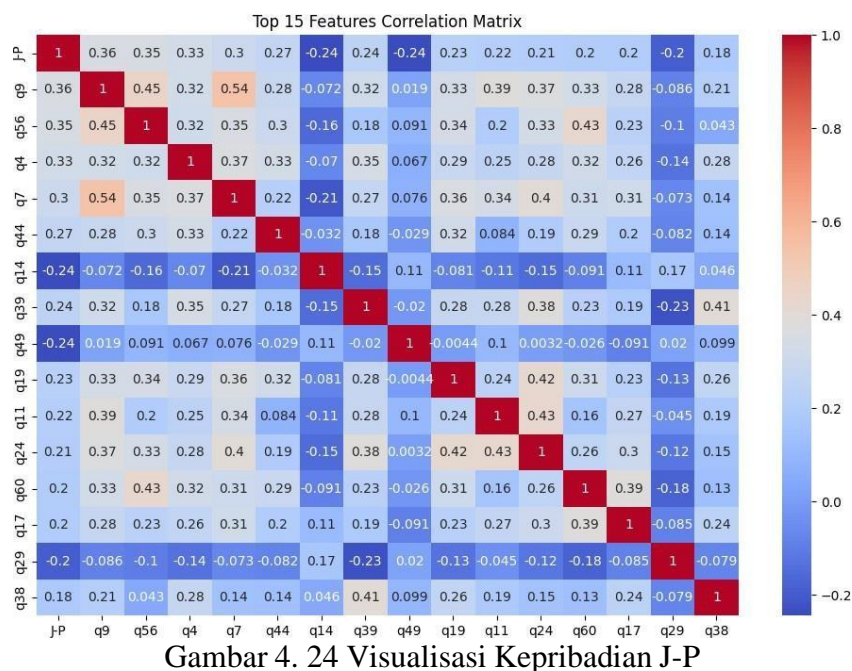
teoretis. q28: Anda tidak mudah terpengaruh oleh argumen yang berlandaskan emosi.

q58: Saat membuat pilihan, Anda lebih cenderung mengandalkan intuisi emosional daripada penalaran logis.

q45: Anda masih merasa terganggu dengan kesalahan yang Anda lakukan di masa lalu.

4.7.4 Correlation Value dimensi J-P

Distribusi kepribadian berdasarkan dimensi J-P dapat dilihat dalam gambar 4.24.



Gambar 4. 24 Visualisasi Kepribadian J-P

Pada dimensi T-F pertanyaan yang ditanyakan mewakili bagaimana pendekatan seseorang terhadap pekerjaan, perencanaan, dan pengambilan keputusan.

Judging merupakan kepribadian yang lebih suka menjalani kehidupan dengan cara yang terencana dan terorganisir, serta senang mencapai penyelesaian dan membuat keputusan. Sedangkan *Perceiving* merupakan kepribadian yang lebih suka menjalani kehidupan dengan berimprovikasi dan menemukan peluang, mereka cenderung fleksibel dan santai.

Korelasi yang memiliki nilai positif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah *Judging*. Sebaliknya korelasi yang memiliki nilai negatif menunjukkan semakin banyak responden yang setuju dengan pertanyaan, maka semakin mereka cenderung ke arah *Perceiving*. Dan di bawah ini merupakan 15 pertanyaan ter atas yang memiliki korelasi paling besar untuk dimensi J-P:

q9: Anda suka menggunakan alat pengorganisasian seperti jadwal dan daftar tugas.

q56: Anda menyelesaikan berbagai hal menurut metode tanpa melewatkan langkah

apa pun.

q4: Ruang hidup dan kerja Anda bersih dan terorganisir.

q7: Anda memprioritaskan dan merencanakan tugas dengan efektif, seringkali menyelesaikannya jauh sebelum tenggat waktu.

q44: Jika rencana Anda terganggu, prioritas utama Anda adalah kembali ke rencana tersebut secepat mungkin.

q14: Anda sering membiarkan hari berlalu tanpa jadwal sama sekali.

q39: Anda suka menyusun daftar tugas sehari-hari.

q49: Gaya kerja pribadi Anda lebih mendekati letupan energi spontan daripada upaya yang teratur dan konsisten.

q19: Anda secara aktif mencari pengalaman dan area pengetahuan baru untuk dijelajahi.

q11: Anda merasa nyaman untuk mendekati seseorang yang menurut Anda menarik dan memulai percakapan.

q24: Anda lebih suka menyelesaikan tugas-tugas terlebih dahulu sebelum bersantai.

q60: Anda yakin bahwa segalanya akan berjalan dengan baik.

q17: Anda menikmati bereksperimen dengan pendekatan baru dan belum teruji.

q29: Anda sering menunda pekerjaan hingga detik-detik terakhir.

q38: Anda biasanya membuat pilihan berdasarkan fakta objektif daripada kesan emosional.

4.8 Analisis Hasil Another Correlation

Pada tahap ini dilakukan analisis terhadap pertanyaan yang memiliki korelasi yang besar untuk setiap dimensi dan pertanyaan yang memiliki sedikit korelasi untuk setiap dimensi. Pertanyaan yang memiliki korelasi tinggi dengan tipe kepribadian disajikan dalam tabel 4.5.

Tabel 4. 5 Nilai Pertanyaan yang Memiliki Korelasi Tinggi

	q59	q32
E-I	-0.133797	-0.105590
S-N	0.224520	0.143905
T-F	-0.215205	-0.172063
J-P	-0.115518	-0.106335

Dari tabel di atas dapat di ketahui korelasi q59 untuk dimensi E-I adalah -0.133797 dimana nilai negatif menunjukkan kecenderungan responden terhadap kepribadian *Introvert*. Korelasi untuk dimensi S-N memiliki nilai 0.224520, dimana nilai positif menunjukkan bahwa kecenderungan responden terhadap kepribadian *Sensing*. Kemudian korelasi untuk dimensi T-F memiliki nilai -0.215205, dimana nilai negatif menunjukkan kecenderungan reponden terhadap kepribadian *Feeling*. Dan terakhir korelasi untuk dimensi J-P memiliki nilai -0.115518, dimana nilai negatif menunjukkan kecenderungan responden terhadap kepribadian *Perceiving*.

Selain itu korelasi korelasi q32 untuk dimensi E-I adalah -0.105590 dimana nilai negatif menunjukkan kecenderungan responden terhadap kepribadian *Introvert*. Korelasi untuk dimensi S-N memiliki nilai 0.143905, dimana nilai positif menunjukkan bahwa kecenderungan responden terhadap kepribadian *Sensing*. Kemudian korelasi untuk dimensi T-F memiliki nilai -0.172063, dimana nilai negatif menunjukkan kecenderungan reponden terhadap kepribadian *Feeling*. Dan terakhir korelasi untuk dimensi J-P memiliki nilai -0.106335, dimana nilai negatif menunjukkan kecenderungan responden terhadap kepribadian *Perceiving*.

Kemudian hasil analisis pertanyaan yang memiliki sedikit korelasi untuk setiap dimensi. Pada tabel 4.6 menampilkan pertanyaan yang memiliki sedikit korelasi dengan kepribadian MBTI.

Tabel 4. 6 Nilai Pertanyaan yang Memiliki Sedikit Korelasi

	q26	q40	q50	q54
E-I	-0.052935	-0.063383	0.063872	0.068835
S-N	0.081449	-0.046663	-0.090514	0.031618
T-F	0.004790	-0.002268	-0.021245	-0.090755
J-P	0.027527	0.083200	-0.065395	0.065573

Pertanyaan yang mendekati nilai 0 menunjukkan bahwa tidak adanya hubungan yang signifikan antara pertanyaan dengan dimensi.

Pada dimensi E-I untuk q26 dan q40 dengan nilai -0.052935 dan -0.063383 menunjukkan sedikit kecenderungan responden terhadap kepribadian *Introvert*. Untuk q50 dan q54 dengan nilai 0.063872 dan 0.068835 menunjukkan sedikit kecenderungan responden terhadap kepribadian *Extrovert*.

Kemudian pada dimensi S-N untuk q40 dan q50 dengan nilai -0.046663 dan -0.090514 menunjukkan sedikit kecenderungan terhadap kepribadian *iNtuition*. Untuk q26 dan q54 dengan nilai 0.081449 dan 0.031618 menunjukkan sedikit kecenderungan terhadap kepribadian *Sensing*.

Kemudian pada dimensi T-F untuk q40, q50, dan q54 dengan nilai -0.002268, -0.021245, dan -0.090755 memiliki sedikit kecenderungan terhadap kepribadian *Feeling*. Untuk q26 yang memiliki nilai 0.004790 menunjukkan sedikit kecenderungan terhadap kepribadian *Thinking*.

Dan terakhir pada dimensi J-P untuk q50 dengan nilai -0.065395 memiliki sedikit kecenderungan terhadap kepribadian *Perceiving*. Sedangkan untuk q40, q50, dan q54 dengan nilai 0.027527, 0.083200, dan 0.065573 memiliki sedikit kecenderungan terhadap kepribadian *Judging*.

4.9 Analisis Hasil Modelling

Pada tahap ini dataset yang ada di bagi menjadi 2 yaitu *train set* dan *test set* dengan

jumlah 139 untuk train set dan 35 untuk test set. Dibawah ini merupakan hasil dari *Modelling* yang telah dilakukan.

4.9.1 Analisis Hasil *Modelling* dimensi E-I

Hasil pemodelan untuk dimenasi E-I ditampilkan dalam tabel 4.7.

Tabel 4. 7 Hasil *Modelling* Dimensi E-I

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.614286	0.774194	0.545455	0.64
RandomForestClassifier	0.757143	0.829268	0.772727	0.8
ExtraTreesClassifier	0.771429	0.833333	0.795455	0.813953
GradientBoostingClassifier	0.771429	0.85	0.772727	0.809524
AdaBoostClassifier	0.757143	0.813953	0.795455	0.804598
xgb_classifier	0.742857	0.795455	0.795455	0.795455

Dimensi E-I dengan model *Decision Tree*

Model *Decision Tree* memiliki nilai akurasi 0.614286 yang merupakan nilai paling rendah dari model-model yang lain. Meskipun termasuk model yang sederhana *Decision Tree* rentan terhadap *overfitting*. Untuk *precision* model *Decision Tree* memiliki nilai 0.774194 atau 77% dimana nilai ini digunakan untuk mengukur seberapa banyak dari prediksi yang benar dan nilai positif yang dihasilkan. Tetapi nilai ini juga merupakan nilai yang paling rendah daripada model- model yang lain,

Untuk *Recall* model ini memiliki nilai 0.545455. *Recall* digunakan untuk mengukur seberapa banyak dari total nilai positif yang berhasil diidentifikasi oleh model. Dalam hal ini model *Decision Tree* kurang efektif untuk mengidentifikasi dari dimensi E-I. Untuk f1-score, model ini memiliki nilai 0.64. Nilai *F1-score* merupakan nilai yang digunakan untuk mengukur kinerja model, dan nilai ini juga menggambarkan nilai rata-rata untuk *precision* dan *Recall*. Dari nilai tersebut model *Decision Tree* belum menunjukkan kinerja yang baik untuk dimensi E-I.

Dimensi E-I dengan model *Random Forest*

Model *Random Forest* untuk dimensi E-I memiliki nilai akurasi 0.757143, nilai ini menunjukkan bahwa model *Random Forest* memiliki performa yang baik secara

keseluruhan. Untuk *precision* model *Random Forest* mendapat kan nilai 0.829268 atau 82%. Nilai ini menunjukkan bahwa model *Random Forest* baik dalam menghindari *false positives*.

Untuk *Recall* model ini mendapatkan nilai 0.772727 atau 77%. Dimana nilai ini menunjukkan nilai yang baik, model ini dapat menangkap sebagian besar nilai positif yang sebenarnya. Dan nilai ini merupakan nilai *Recall* yang paling tinggi daripada model-model yang lain. Untuk *F1-score* model *Random Forest* mendapatkan nilai 0.8 yang menunjukkan bahwa model *Random Forest* dapat memberikan keseimbangan antara *precision* dan *Recall*.

Dimensi E-I dengan model *Extra Trees*

Model *Extra Trees* memiliki nilai akurasi 0.771429. Nilai ini menunjukkan bahwa model ini memiliki performa yang baik untuk dimensi E-I, dan model ini juga memiliki nilai akurasi paling tinggi daripada model-model lainnya. Untuk *precision* model ini memiliki nilai 0.833333 atau 83%. Dimana nilai ini menunjukkan bahwa model *Extra Trees* akurat dalam mengidentifikasi dimensi E-I.

Kemudian untuk *Recall* model ini memiliki nilai 0.795455 atau 79%. Nilai ini menunjukkan bahwa model ini dapat menangkap 79% dari semua kasus positif pada dimensi E-I, dan tidak melewatkan banyak kasus positif. Untuk nilai F1 model ini memiliki nilai 0.813953. Nilai ini menunjukkan keseimbangan yang baik antara *Precision* dan *Recall* pada dimensi E-I.

Dimensi E-I dengan model *Gradient Boosting*

Model *Gradient Boosting* memiliki nilai akurasi 0.771429. Nilai ini menunjukkan bahwa model ini memiliki sekitar 77% benar dari semua prediksi yang dibuat. Dan memiliki performa yang baik untuk dimensi E-I. Untuk *precision* model ini memiliki nilai 0.85. Nilai ini menunjukkan bahwa model *Gradient Boosting* bisa mengidentifikasi dimensi E-I sebesar 85%, yang mana model ini sangat akurat dalam mengidentifikasi dimensi E-I.

Kemudian untuk *Recall* memiliki nilai 0.772727. Menunjukkan bahwa model *Gradient Boosting* dapat menangkap 77% dari semua kasus positif yang ada pada dimensi E-I, dan merupakan model yang baik untuk dimensi E-I. Untuk F1 memiliki nilai 0.809524. Nilai ini menunjukkan bahwa model ini memiliki keseimbangan yang baik

antara *precision* dan *Recall* pada dimensi E-I.

Dimensi E-I dengan model *Ada Boost*

Model ini memiliki nilai akurasi 0.757143 yang menunjukkan model ini belum mempunyai performa yang baik untuk mengklasifikasi dimensi E-I. Kemudian untuk *precision* model ini mempunyai nilai 0.813953. Dengan nilai 81% menunjukkan bahwa model ini baik dalam mengidentifikasi kelas yang positif, dan masih ada ruang untuk perbaikan.

Untuk *Recall* model ini memiliki nilai 0.795455. Dengan nilai ini menunjukkan bahwa model *AdaBoost* dapat menangkap 79% dari kasus positif yang ada pada dimensi E-I. Untuk F1 model ini memiliki nilai 0.804598. Menunjukkan bahwa model ini menunjukkan nilai yang efektif untuk dimensi E-I.

Dimensi E-I dengan model *XGBoost*

Model ini memiliki nilai akurasi 0.742857 yang menunjukkan model *XGBoost* mempunyai performa yang baik untuk dimensi E-I. Kemudian untuk *precision* model ini mempunyai nilai 0.795455. Dengan nilai 79% menunjukkan bahwa model ini baik dalam mengidentifikasi kelas yang positif.

Untuk *Recall* model ini memiliki nilai 0.795455. Dengan nilai ini menunjukkan bahwa model *XGBoost* dapat menangkap 79% dari kasus positif yang ada pada dimensi E-I. Untuk F1 model ini memiliki nilai 0.795455. Menunjukkan bahwa model ini menunjukkan nilai yang cukup efektif untuk dimensi E-I.

Secara keseluruhan dapat diketahui model yang paling baik untuk mengidentifikasi dimensi E-I adalah model *Extra Trees*. karena model memiliki kemampuan mengatasi *overfitting*, variasi data, ketergantungan atribut, ketergantungan kelas, dan noise yang lebih baik daripada algoritma lainnya. Model ini memiliki akurasi yang paling tinggi dan F1 score yang paling tinggi. Precision dan recall-nya juga tinggi.

4.9.2 Analisis Hasil dimensi S-N

Pada tabel 4.8 menyajikan hasil pemodelan untuk dimensi S-N.

Tabel 4. 8 Hasil Modelling Dimensi S-N

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.714286	0.711238	0.714286	0.710348
RandomForestClassifier	0.614286	0.601819	0.614286	0.584962
ExtraTreesClassifier	0.714286	0.718709	0.714286	0.700052
GradientBoostingClassifier	0.714286	0.714286	0.714286	0.704127
AdaBoostClassifier	0.685714	0.681651	0.685714	0.681383
xgb_classifier	0.7	0.696853	0.7	0.697112

Dimensi S-N dengan model *Decision Tree*

Model *Decision Tree* memiliki nilai akurasi 0.714286 yang menunjukkan bahwa model ini benar 71% dari semua prediksi yang dibuat. Dan masih ada ruang untuk perbaikan lebih lanjut. Untuk *precision* model *Decision Tree* memiliki nilai 0.711238. Nilai ini menunjukkan bahwa model *Decision Tree* baik dalam mengidentifikasi kelas positif.

Untuk *Recall* model ini memiliki nilai 0.714286. Nilai ini menunjukkan model *Decision Tree* mampu menangkap 71% dari semua kasus positif yang ada pada dimensi S-N. Untuk f1- score, model ini memiliki nilai 0.710348. Nilai ini menunjukkan bahwa model ini memberikan performa yang cukup baik, dapat terlihat dari nilai *Precision* dan *Recall* yang tidak terlalu jauh.

Dimensi S-N dengan model *Random Forest*

Model *Random Forest* untuk dimensi S-N memiliki nilai akurasi 0.614286, nilai ini menunjukkan bahwa model *Random Forest* memiliki performa yang kurang baik secara keseluruhan. Model ini juga merupakan model dengan nilai akurasi yang paling. Untuk *precision* model *Random Forest* mendapat kan nilai 0.601819. Nilai ini menunjukkan bahwa model *Random Forest* memiliki masalah dalam mengidentifikasi dimensi S-N dengan banyaknya kesalahan dalam prediksi positif.

Untuk *Recall* model ini mendapatkan nilai 0.614286 Dimana nilai ini menunjukkan bahwa model ini tidak terlalu baik dalam menangkap kasus positif. Untuk *F1-score* model *Random Forest* mendapatkan nilai 0.584962. Nilai F1 yang rendah menunjukkan bahwa model ini tidak memiliki keseimbangan yang baik antara *Precision* dan *Recall*, dan performa secara keseluruhan kurang memuaskan.

Dimensi S-N dengan model *Extra Trees*

Model *Extra Trees* untuk dimensi S-N memiliki nilai akurasi 0.714286, nilai ini menunjukkan bahwa model ini memiliki performa yang baik secara keseluruhan. Model ini juga merupakan model dengan nilai akurasi yang paling tinggi dan memiliki nilai yang sama dengan model *Decision Tree* dan *Gradient Boosting*. Untuk *precision* model ini mendapatkan nilai 0.718709. Nilai ini menunjukkan bahwa model *Extra Tree* baik dalam mengidentifikasi kelas positif.

Untuk *Recall* model ini mendapatkan nilai 0.714286 Dimana nilai ini menunjukkan bahwa model ini baik dalam menangkap kasus positif. Untuk *F1-score* model *Extra Trees* mendapatkan nilai 0.700052. Nilai F1 menunjukkan bahwa model ini memiliki keseimbangan yang baik antara *Precision* dan *Recall*.

Dimensi S-N dengan model *Gradient Boosting*

Model *Gradient Boosting* memiliki nilai akurasi 0.714286. Nilai ini menunjukkan bahwa model ini memiliki sekitar 71% benar dari semua prediksi yang dibuat. Dengan nilai ini model *Gradient Boosting* memiliki nilai akurasi paling tinggi daripada model-model yang lain. Untuk *precision* model ini memiliki nilai 0.714286. Nilai ini menunjukkan bahwa model *Gradient Boosting* bisa mengidentifikasi kasus positif pada dimensi S-N dengan cukup baik.

Kemudian untuk *Recall* memiliki nilai 0.714286. Menunjukkan bahwa model *Gradient Boosting* dapat menangkap 71% dari semua kasus positif yang ada pada dimensi S-N. model ini memiliki nilai yang sama antara *Recall* dan *Precision*nya. Untuk F1 memiliki nilai 0.704127. Nilai ini menunjukkan bahwa model ini memiliki keseimbangan yang baik antara *Precision* dan *Recall* pada dimensi S-N.

Dimensi S-N dengan model *Ada Boost*

Model ini memiliki nilai akurasi 0.685714 yang menunjukkan model ini

mempunyai performa yang cukup baik untuk mengklasifikasi dimensi S-N.

Kemudian untuk *precision* model ini mempunyai nilai 0.681651. Dengan nilai 68% menunjukkan bahwa model ini cukup baik dalam mengidentifikasi kelas yang positif, dan masih ada ruang untuk perbaikan.

Untuk *Recall* model ini memiliki nilai 0.685714. Dengan nilai ini menunjukkan bahwa model *Bagging* dapat menangkap 68% dari kasus positif yang ada pada dimensi S-N. Dan nilai ini masih kurang baik untuk digunakan pada dimensi S-N.

Dan untuk F1 model ini memiliki nilai 0.681383. Menunjukkan bahwa model ini menunjukkan nilai yang cukup efektif untuk dimensi S-N. Dan masih ada ruang untuk perbaikan.

Dimensi S-N dengan model *XGBoost*

Model ini memiliki nilai akurasi 0.7 yang menunjukkan model *XGBoost* mempunyai performa yang baik dalam mengidentifikasi kasus positif pada dimensi S-N. Kemudian untuk *precision* model ini mempunyai nilai 0.696853. Model ini memiliki nilai sedikit lebih tinggi daripada model *AdaBoost*, tetapi model ini juga masih belum cukup baik untuk dimensi S-N.

Untuk *Recall* model ini memiliki nilai 0.7. Dengan nilai ini menunjukkan bahwa model *XGBoost* dapat menangkap 70% dari kasus positif yang ada pada dimensi S-N. Untuk F1 model ini memiliki nilai 0.697112. Menunjukkan bahwa model ini menunjukkan nilai yang kurang efektif untuk dimensi S-N.

Dari hasil di atas, menunjukkan bahwa *Decision Tree* dan *Extra Trees* memiliki kelebihan yang signifikan dalam hal akurasi, *precision*, *recall* dan F1. Namun, jika dilihat secara keseluruhan dapat diketahui model yang paling baik untuk mengidentifikasi dimensi S-N adalah model *Extra Trees* karena memiliki nilai akurasi paling tinggi dan *precision* paling tinggi.

4.9.3 Analisis Hasil dimensi T-F

Hasil pemodelan untuk dimensi T-F dapat dilihat dalam tabel 4.9.

Tabel 4. 9 Hasil *Modelling* Dimensi T-F

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.528571	0.516812	0.528571	0.509453
RandomForestClassifier	0.6	0.607407	0.6	0.568116
ExtraTreesClassifier	0.614286	0.613215	0.614286	0.603147
GradientBoostingClassifier	0.614286	0.615327	0.614286	0.614681
AdaBoostClassifier	0.642857	0.646531	0.642857	0.643514
xgb_classifier	0.628571	0.627143	0.628571	0.627343

Dimensi T-F dengan model *Decision Tree*

Model *Decision Tree* memiliki nilai akurasi 0.528571 yang merupakan nilai paling rendah dari model-model yang lain. Dan menunjukkan bahwa model ini memiliki kemampuan yang sangat terbatas untuk memprediksi kasus positif, dengan lebih dari setengah prediksi yang salah. Untuk *precision* model *Decision Tree* memiliki nilai 0.516812. Dimana nilai ini digunakan untuk mengukur seberapa banyak dari prediksi yang benar dan nilai positif yang dihasilkan. Dan model ini mempunyai nilai yang paling rendah daripada model-model yang lain untuk dimensi T-F.

Untuk *Recall* model ini memiliki nilai 0.528571. *Recall* digunakan untuk mengukur seberapa banyak dari total nilai positif yang berhasil diidentifikasi oleh model. dalam hal ini model *Decision Tree* kurang efektif untuk mengidentifikasi dari dimensi T-F. Untuk f1-score, model ini memiliki nilai 0.509453. Nilai *F1-score* merupakan nilai yang digunakan untuk mengukur kinerja model, dan nilai ini juga menggambarkan nilai rata-rata untuk *precision* dan *Recall*. Dari nilai tersebut model *Decision Tree* belum menunjukkan kinerja yang kurang baik untuk dimensi T-F.

Dimensi T-F dengan model *Random Forest*

Model *Random Forest* untuk dimensi T-F memiliki nilai akurasi 0.6, nilai ini menunjukkan bahwa model *Random Forest* memiliki performa yang kurang baik dalam mengidentifikasi pada dimensi T-F. Untuk *Precision* model *Random Forest* mendapatkan nilai 0.607407 atau 60%. Nilai ini menunjukkan bahwa model *Random Forest* kurang baik untuk mengidentifikasi dimensi T-F.

Untuk *Recall* model ini mendapatkan nilai 0.6 atau 60%. Dimana nilai ini menunjukkan nilai yang kurang baik, Dalam hal ini model *Random Forest* kurang efektif untuk mengidentifikasi dari dimensi T-F. Untuk *F1-score* model *Random Forest* mendapatkan nilai 0.568116 Menunjukkan bahwa model ini tidak cukup baik untuk dimensi T-F.

Dimensi T-F dengan model *Extra Trees*

Model *Extra Trees* memiliki nilai akurasi 0.614286. Nilai ini menunjukkan bahwa model ini memiliki performa yang kurang baik untuk dimensi T-F. Untuk *Precision* model ini memiliki nilai 0.613215 atau 61%. Dimana nilai ini menunjukkan bahwa model *Extra Trees* kurang akurat untuk mengidentifikasi dimensi T-F.

Kemudian untuk *Recall* model ini memiliki nilai 0.614286 atau 61 %. Nilai ini menunjukkan bahwa model ini dapat hanya menangkap 61%, model ini juga belum cukup baik untuk dimensi T-

F. Untuk nilai F1 model ini memiliki nilai 0.603147. Nilai ini menunjukkan bahwa model ini belum cukup baik untuk dimensi T-F.

Dimensi T-F dengan model *Gradient Boosting*

Model *Gradient Boosting* memiliki nilai akurasi 0,614286 yang menunjukkan bahwa model *Gradient Boosting Classifier* kurang baik dalam mengidentifkasi dimensi T-F. Untuk *Precision* model ini memiliki nilai 0.615327. Nilai ini menunjukkan bahwa model *Gradient Boosting* kurang efektif untuk mengidentifikasikan dimensi T-F.

Kemudian untuk *Recall* memiliki nilai 0,614286. Menunjukkan bahwa model *Gradient Boosting* dapat menangkap 61% akan tetapi model ini belum cukup baik untuk mengidentifikasi dimensi T-F. Untuk F1 memiliki nilai 0.614681. Nilai ini menunjukkan bahwa model ini tidak cukup baik untuk dimensi T-F. Dan membutuhkan perbaikan

kedepannya.

Dimensi T-F dengan model *Ada Boost*

Model ini memiliki nilai akurasi 0.642857 nilai ini merupakan nilai akurasi paling tinggi dari antara model lain untuk dimensi T-F. Kemudian untuk *Precision* model ini mempunyai nilai 0.646531. Model ini memiliki nilai 64% meskipun model ini mempunyai nilai paling tinggi dari antara model yang lain, tapi masih belum cukup efektif untuk digunakan pada dimensi T-F.

Untuk *Recall* model ini memiliki nilai 0.643514 Dengan nilai ini menunjukkan bahwa model ini belum cukup efektif digunakan pada dimensi T-F. Untuk F1 model ini memiliki nilai 0.643514. Pada model ini memiliki nilai 64%, model ini masih kurang efektif terhadap dimensi T-F.

Dimensi T-F dengan model *XGBoost*

Model ini memiliki nilai akurasi 0.628571 yang menunjukkan model *XGBoost* mempunyai performa yang kurang baik untuk dimensi T-F. Kemudian untuk *Precision* model ini mempunyai nilai 0.627143. Dengan nilai ini menunjukkan bahwa tidak cukup baik untuk digunakan pada dimensi T-F.

Untuk *Recall* model ini memiliki nilai 0,628571. Dengan nilai ini menunjukkan bahwa model *XGBoost* masih belum cukup efektif untuk dimensi T-F. Untuk F1 model ini memiliki nilai 0.627343 . Menunjukkan bahwa model ini masih belum cukup baik untuk dimensi T-F.

Secara keseluruhan dapat diketahui model yang paling baik untuk mengidentifikasi dimensi T-F adalah model *Ada Boost* dengan nilai akurasi sampai nilai F1 yang paling tinggi di antara model-model lainnya.

4.9.4 Analisis Hasil dimensi J-P

Pada tabel 4.10 menampilkan hasil pemodelan untuk dimensi J-P.

Tabel 4. 10 Hasil Modelling Dimensi J-P

	Accuracy	Precision	Recall	F1
DecisionTreeClassifier	0.7	0.69	0.7	0.691228
RandomForestClassifier	0.657143	0.632157	0.657143	0.623176
ExtraTreesClassifier	0.7	0.692529	0.7	0.666229
GradientBoostingClassifier	0.685714	0.67013	0.685714	0.662143
AdaBoostClassifier	0.657143	0.646259	0.657143	0.649531
xgb_classifier	0.685714	0.670908	0.685714	0.654578

Dimensi J-P dengan model *Decision Tree*

Model *Decision Tree* memiliki nilai akurasi 0.7 yang merupakan nilai paling tinggi dari model-model yang lain. Model ini memiliki nilai akurasi yang sama dengan model *Extra Trees*. Untuk *Precision* model *Decision Tree* memiliki nilai 0.69 atau 69% dimana nilai ini digunakan untuk mengukur seberapa banyak dari prediksi yang benar dan nilai positif yang dihasilkan. Nilai ini cukup efektif digunakan pada dimensi J-P.

Untuk *Recall* model ini memiliki nilai 0.7. *Recall* digunakan untuk mengukur seberapa banyak dari total nilai positif yang berhasil diidentifikasi oleh model. dalam hal ini model *Decision Tree* efektif untuk mengidentifikasi dari dimensi J-P. Untuk f1-score, model ini memiliki nilai 0.623176. Nilai *F1-score* merupakan nilai yang digunakan untuk mengukur kinerja model, dan nilai ini juga menggambarkan nilai rata-rata untuk *Precision* dan *Recall*. Dari nilai tersebut model *Decision Tree* menunjukkan kinerja yang cukup baik untuk dimensi J-P.

Dimensi J-P dengan model *Random Forest*

Model *Random Forest* untuk dimensi J-P memiliki nilai akurasi 0.657143, nilai ini menunjukkan bahwa model *Random Forest* belum memiliki performa yang cukup baik dan nilai ini merupakan nilai akurasi paling rendah diantara model-model lainnya. Untuk *Precision* model *Random Forest* mendapat kan nilai 0,632157 atau 63%. Nilai ini menunjukkan bahwa model *Random Forest* masih kurang baik dalam menghindari false

positives.

Untuk *Recall* model ini mendapatkan nilai 0.657143. Dimana nilai ini menunjukkan nilai yang kurang baik, model ini masih kurang baik terhadap dimensi J-P. Untuk *F1-score* model *Random Forest* mendapatkan nilai 0.623176 yang menunjukkan bahwa model *Random Forest* kurang efektif terhadap dimensi J-P dan masih ada ruang untuk perbaikan.

Dimensi J-P dengan model *Extra Trees*

Model *Extra Trees* memiliki nilai akurasi 0.7 atau 70%. Nilai ini menunjukkan bahwa model ini memiliki performa yang baik untuk dimensi J-P, dan model ini juga memiliki nilai akurasi paling tinggi diantara model-model lainnya. Untuk *Precision* model ini memiliki nilai 0.692529 atau 69%. Dimana nilai ini menunjukkan bahwa model *Extra Trees* cukup akurat dalam mengidentifikasi dimensi J-P.

Kemudian untuk *Recall* model ini memiliki nilai 0.7 atau 70 %. Nilai ini menunjukkan bahwa model ini dapat menangkap 70% dari semua kasus positif pada dimensi J-P, dan tidak melewatkan banyak kasus positif. Untuk nilai F1 model ini memiliki nilai 0.662143. Nilai ini menunjukkan bahwa model ini masih kurang baik terhadap dimensi J-P meskipun model yang lain memberikan hasil positif.

Dimensi J-P dengan model *Gradient Boosting*

Model *Gradient Boosting* memiliki nilai akurasi 0.685714. Nilai ini menunjukkan bahwa model ini memiliki sekitar 68%. Jadi dapat disimpulkan bahwa model ini masih cukup efektif untuk digunakan dimensi J-P. Untuk *Precision* model ini memiliki nilai 0.67013. Nilai ini menunjukkan bahwa model *Gradient Boosting* masih kurang cukup untuk mengidentifikasi dimensi J-P.

Kemudian untuk *Recall* memiliki nilai 0.685714. Menunjukkan bahwa model *Gradient Boosting* dapat menangkap 68% kasus positif yang ada di dimensi J-P. dan sudah cukup baik untuk digunakan pada dimensi J-P. Untuk F1 memiliki nilai 0.662143. Nilai ini menunjukkan bahwa model ini kurang efektif pada dimensi J-P.

Dimensi J-P dengan model *Ada Boost*

Model ini memiliki nilai akurasi 0.657143 yang menunjukkan model ini kurang baik untuk mengklasifikasi dimensi J-P. Kemudian untuk *Precision* model ini mempunyai

nilai 0.646259. Dengan nilai 64% menunjukkan bahwa model ini masih belum cukup baik dalam mengidentifikasi kelas yang positif, dan masih ada ruang untuk perbaikan.

Untuk *Recall* model ini memiliki nilai 0.657143. Dengan nilai ini menunjukkan bahwa model *AdaBoost* dapat menangkap 65% dari kasus positif yang ada pada dimensi J-P. Untuk F1 model ini memiliki nilai 0.654578. Menunjukkan bahwa model ini belum efektif terhadap dimensi J-P.

Dimensi J-P dengan model *XGBoost*

Model ini memiliki nilai akurasi 0.685714 yang menunjukkan model *XGBoost* mempunyai performa yang cukup baik terhadap dimensi J-P. Kemudian untuk *Precision* model ini mempunyai nilai 0.670908. Dengan nilai 67% menunjukkan bahwa model ini cukup baik dalam mengidentifikasi kelas yang positif.

Untuk *Recall* model ini memiliki nilai 0.685714. Dengan nilai ini menunjukkan bahwa model *XGBoost* dikatakan belum cukup efektif untuk dimensi J-P. Untuk F1 model ini memiliki nilai 0.654578. Menunjukkan bahwa model ini menunjukkan nilai yang tidak cukup efektif untuk dimensi J-P.

Secara keseluruhan dapat diketahui model yang paling baik untuk mengidentifikasi dimensi J-P adalah model *Decision Tree* dengan nilai akurasi sampai nilai F1 yang paling tinggi di antara model-model lainnya.

4.10 Analisis Hasil *Testing*

Pada tahap ini dilakukan *Testing* untuk semua dimensi MBTI yang di gabungkan dengan menggunakan *Custom Accuracy Score* yang dapat memberikan seberapa baik suatu model klasifikasi dalam memprediksi dimensi MBTI, dan *Custom Accuracy Score* juga dapat digabungkan dengan matriks lainnya seperti *precision*, *Recall*, dan *F1-score* untuk memberikan gambaran yang lebih komprehensif tentang kinerja model. Rumus *Custom Accuracy Score* adalah sebagai berikut:

$Custom Accuracy Score = (1/N) \times \sum (1/M \times \sum (1 \text{ jika } p_{ij} = t_{ij}, 0 \text{ jika tidak}))$

Dimana:

- N adalah jumlah prediksi yang dilakukan oleh model.
- M adalah jumlah label biner dalam setiap prediksi atau label sebenarnya.
- p_{ij} dan t_{ij} adalah label biner ke-j dalam prediksi ke-i dan label sebenarnya ke-i, masing-masing.
- 1 jika $p_{ij} = t_{ij}$, 0 jika tidak adalah fungsi indikator yang mengembalikan nilai 1 jika prediksi sesuai dengan label sebenarnya dan 0 jika tidak.

Gambar 4. 25 Rumus *Custom Accuracy Score*

Dan hasil yang di dapat adalah sebagai berikut:

```
{'DecisionTreeClassifier': 0.6392857142857142,
 'RandomForestClassifier': 0.6571428571428571,
 'ExtraTreesClassifier': 0.7,
 'GradientBoostingClassifier': 0.6964285714285714,
 'AdaBoostClassifier': 0.6857142857142857,
 'xgb_classifier': 0.6892857142857143}
```

Gambar 4. 26 Hasil Testing untuk Semua Dimensi

Berdasarkan hasil prediksi yang telah dilakukan dapat diketahui bahwa nilai akurasi model yang paling tinggi adalah model *Extra Trees* dengan nilai akurasi 0.70, model ini mendapatkan nilai paling tinggi karena model *Extra Trees* memilih fitur dan melakukan pemisahan secara acak dibandingkan dengan model-model lainnya. Hal ini dapat membantu untuk mengurangi *Overfitting* dan meningkatkan generalisasi model. Kemudian ada model *Gradient Boosting* yang memiliki nilai akurasi yaitu 0.69, model ini mendapatkan nilai yang tinggi dikarenakan mempunyai kombinasi model yang kuat. *Gradient Boosting* efektif dalam memperbaiki kesalahan dan dapat menangkap interaksi pada masing-masing dimensi. Kemudian model *XGBoost* dengan nilai akurasi 0.689, model *AdaBoost* dengan nilai akurasi 0.685, model *Random Forest* dengan nilai akurasi 0.65 dan terakhir ada model *Decision Tree* dengan nilai akurasi 0.63. Dan model *Decision Tree* merupakan model dengan nilai akurasi yang paling rendah, menandakan bahwa model ini kurang efektif dalam memprediksi pola data tertentu dibandingkan dengan model-model sebelumnya.

4.10.1 Analisis hasil High Score Correlation

Pada tahap ini yang akan digunakan adalah pertanyaan-pertanyaan yang memiliki korelasi yang tinggi terhadap masing-masing dimensi. Untuk mengetahui pertanyaan yang memiliki korelasi tertinggi dengan menggunakan perintah di bawah ini:

```
highcorr_ei =
list(target_corr_ei[target_corr_ei>=0.1][1:].index)
highcorr_sn =
list(target_corr_sn[target_corr_sn>=0.1][1:].index)
highcorr_tf =
list(target_corr_tf[target_corr_tf>=0.1][1:].index)
highcorr_jp =
list(target_corr_jp[target_corr_jp>=0.1][1:].index)
```

Gambar 4. 27 Kode perintah menampilkan pertanyaan yang memiliki korelasi pada masing masing dimensi

Dan dibawah ini adalah pertanyaan-pertanyaan yang memiliki korelasi yang tinggi untuk setiap dimensi:

```
Number of High Corr Question of E-I: 29
Number of High Corr Question of S-N: 33
Number of High Corr Question of T-F: 25
Number of High Corr Question of J-P: 31
```

Gambar 4. 28 Hasil Pertanyaan yang Memiliki Korelasi Tinggi Terhadap Setiap Dimensi

Setelah mendapatkan pertanyaan-pertanyaannya maka akan dilakukan predeksi MBTI dengan *test set* yang telah disediakan menggunakan *Custom Accuracy Score* yang sama. Dan hasilnya seperti dibawah ini:

```
{'DecisionTreeClassifier': 0.6035714285714285,
 'RandomForestClassifier': 0.7178571428571429,
 'ExtraTreesClassifier': 0.7285714285714285,
 'GradientBoostingClassifier': 0.7,
 'AdaBoostClassifier': 0.7035714285714286,
 'xgb_classifier': 0.7142857142857143}
```

Gambar 4. 29 Hasil Testing untuk Pertanyaan yang Memiliki Korelasi Tinggi

Dari hasil diatas dapat dilihat model yang memiliki nilai akurasi yang paling tinggi untuk memprediksi tipe kepribadian MBTI berdasarkan pertanyaan-pertanyaan yang

memiliki korelasi yang paling tinggi adalah *Extra Trees* dengan nilai akurasi 0.728, kemudian model tertinggi setelah *Extra Trees* ada *Random Forest* dengan nilai akurasi 0.7178, kemudian model *XGBoost* dengan nilai akurasi 0.7142, kemudian model *AdaBoost* dengan nilai akurasi 0.703, dilanjut dengan model *Gradient Boosting* dengan nilai akurasi 0.70, dan terakhir model yang memiliki nilai paling rendah ada *Decision Tree* dengan nilai 0.603.

Model *Extra Trees* mendapatkan nilai akurasi paling tinggi karena memiliki beberapa kelebihan. Pertama, algoritma ini dapat mengatasi *overfitting* dengan lebih baik, sehingga dapat meningkatkan akurasi. Kedua, *Extra Trees* dapat mengatasi variasi data dan ketergantungan atribut dengan lebih baik, sehingga dapat meningkatkan akurasi. Terakhir, algoritma ini dapat mengatasi noise dan ketergantungan kelas dengan lebih baik, sehingga dapat meningkatkan akurasi.

BAB V

KESIMPULAN DAN SARAN

5.1 Kesimpulan

Berdasarkan hasil penelitian yang telah dilakukan, dapat ditarik beberapa kesimpulan:

- a. Implementasi *Tree Based Model* untuk klasifikasi MBTI telah berhasil dilakukan melalui beberapa tahapan:
 1. Pengumpulan data dari 174 responden melalui kuesioner yang diselenggarakan dengan tes 16Personalities
 2. *Pre-processing* data yang meliputi penanganan *missing value* dan *split label*
 3. Analisis korelasi dan distribusi data untuk setiap dimensi MBTI
 4. Implementasi dan evaluasi enam algoritma *Tree Based Model*
- b. Pada penelitian ini algoritma *Tree Based Model* yang paling akurat adalah *Extra Trees* dengan nilai akurasi 0.728, algoritma lain seperti *Random Forest*, *XGBoost*, *AdaBoost*, dan *Gradient Boosting* juga mendapat nilai akurasi di atas 0.70. menandakan beberapa algoritma turunan *Tree Based Model* ini bisa digunakan untuk kasus klasifikasi tipe kepribadian dengan alat ukur MBTI.

5.2 Saran

Berdasarkan hasil penelitian, beberapa saran untuk pengembangan penelitian selanjutnya:

- a. Pengembangan Dataset:
 1. Memperbesar jumlah sampel data untuk meningkatkan representativitas
 2. Mengumpulkan data dari berbagai kelompok demografis yang berbeda
 3. Mempertimbangkan penggunaan data longitudinal untuk mengamati perubahan kepribadian

- b. Peningkatan Metodologi:
 - 1. Mengintegrasikan teknik *deep learning* dengan *Tree Based Model*
 - 2. Mengimplementasikan teknik *ensemble learning* yang lebih *advanced*
 - 3. Mengoptimalkan *hyperparameter tuning* untuk setiap algoritma
 - 4. Menerapkan teknik *cross-validation* yang lebih *robust*
- c. Pengembangan Aplikasi:
 - 1. Membuat sistem *assessment* MBTI berbasis web yang terintegrasi
 - 2. Merancang *interface* yang *user-friendly* untuk penggunaan model
- d. Eksplorasi Fitur:
 - 1. Menambahkan analisis *sentiment* untuk meningkatkan akurasi klasifikasi
 - 2. Mengintegrasikan data *behavioral* untuk memperkaya fitur
 - 3. Mengembangkan *feature engineering* yang lebih *sophisticated*
- e. Validasi dan Evaluasi:
 - 1. Melakukan validasi silang dengan instrumen kepribadian lainnya
 - 2. Mengembangkan metrik evaluasi yang lebih komprehensif
 - 3. Melakukan studi komparatif dengan model-model *machine learning* lainnya

DAFTAR PUSTAKA

- Angsaweni, A., & Maharani, W. (2022). Identification of Big Five personality on Twitter users using the *AdaBoost* method. *Building of Informatics, Technology and Science (BITS)*, 4(2), 377–383. <https://doi.org/10.47065/bits.v4i2.1853>
- Breiman, L. (2001). *Random forests* (Vol. 45).
- Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory*. Prentice-Hall.
- Bronfenbrenner, U. (1979). *The ecology of human development: Experiments by nature and design*. Harvard University Press.
- Cattell, R. B. (1946). *The description and measurement of personality*. Harcourt, Brace, & World.
- Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic minority over-sampling technique. *Journal of Artificial Intelligence Research*, 16, 321–357.
- Chen, F., Shen, X., Liu, Y., Wang, Y., Zhang, Z., & Wang, M. (2023). Myers-Briggs Type Indicator analysis using decision-tree-related methods. *Applied and Computational Engineering*, 2(1), 899–913. <https://doi.org/10.54254/2755-2721/2/20220564>
- Chen, T., & Guestrin, C. (2016). *XGBoost: A scalable tree boosting system*. <https://doi.org/10.1145/2939672.2939785>
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*(pp. 4171–4186).
- Dzaki, R. (2024). *PREDIKSI PENYAKIT JANTUNG DENGAN MENGGUNAKAN MACHINE LEARNING AUTOGLUON*.
- Eysenck, H. J. (1967). *The biological basis of personality*. Springfield, III: Thomas.
- Faisal Amri, C., Prasetyowati, S., & Sibaroni, Y. (2022). Social media user personality classification based on how user live and make decision. *Jurnal Ilmiah Penelitian dan Pembelajaran Informatika*, 7. <https://www.youtube.com/watch?v=iz7IE1g4XM4>
- Freud, S. (1923). *The ego and the id*. SE, 19, 1–66.
- Freund, Y., & Schapire, R. E. (1997). A decision-theoretic generalization of on-line

- learning and an application to boosting. *Journal of Computer and System Sciences*, 55(1), 119–139. <https://doi.org/10.1006/jcss.1997.1504>
- Friedman, J. H. (2001). Greedy function approximation: A *Gradient Boosting* machine. In *The Annals of Statistics*, 29(5), 1189–1232. <https://doi.org/10.1214/aos/1013203451>
- Gallup. (2024). *State of the global workplace*.
- Garg, S., Kumar, P., & Garg, A. (2021). Comparison of *machine learning* algorithms for content-based personality resolution of tweets. <https://ssrn.com/abstract=3626306>
- Geurts, P., Ernst, D., & Wehenkel, L. (2006). Extremely randomized trees. *Machine learning*, 63(1), 3–42. <https://doi.org/10.1007/s10994-006-6226-1>
- Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: Data mining, inference, and prediction* (2nd ed.). Springer.
- Ikhran, F., & Ramsari, N. (2023). Penggunaan indikator Myers-Briggs Type Indicator (MBTI) untuk. In *SOSIERA* (Vol. II).
- John, O. P., & Srivastava, S. (1999). The Big Five trait taxonomy: History, measurement, and theoretical perspectives. In L. A. Pervin & O. P. John (Eds.), *Handbook of personality: Theory and research* (2nd ed., pp. 102–138). Guilford Press.
- Jung, C. G. (1971). *Psychological types*. Princeton University Press.
- Maslow, A. H. (1943). A theory of human motivation. *Psychological Review*, 50(4), 370–396. <https://doi.org/10.1037/h0054346>
- McCrae, R. R., & Costa, P. T. (2008). Empirical and theoretical status of the Five-Factor Model of personality traits. In *The SAGE Handbook of Personality Theory and Assessment* (Vol. 1, pp. 273–294).
- McCrae, R. R., & John, O. P. (1992). An introduction to the Five-Factor Model and its applications. *Journal of Personality*, 60(2), 175–215. <https://doi.org/10.1111/j.1467-6494.1992.tb00970.x>
- Mushtaq, Z., Ashraf, S., & Sabahat, N. (2020, November 5). Predicting MBTI personality type with K-means clustering and *Gradient Boosting*. In *Proceedings - 2020 23rd IEEE International Multi-Topic Conference, INMIC 2020*. <https://doi.org/10.1109/INMIC50486.2020.9318078>
- Myers, I. B., & Myers, P. B. (1995). *Gifts differing: Understanding personality type*. Davies- Black Publishing.
- Nisha, K. A., Kulsum, U., Rahman, S., Hossain, Md. F., Chakraborty, P., & Choudhury, T. (2022). A comparative analysis of *machine learning* approaches in personality prediction using MBTI. In *Advances in Intelligent Systems and*

- Computing* (pp. 13–23). https://doi.org/10.1007/978-981-16-2543-5_2
- Quinlan, J. R. (1993). *C4.5: Programs for machine learning*. Morgan Kaufmann Publishers.
- Quinlan, J. R. (1986). Induction of *Decision Trees*. In *Machine learning* (Vol. 1, pp. 81–106).
- Riso, D. R., & Hudson, R. (2000). *The wisdom of the Enneagram: The complete guide to psychological and spiritual growth for the nine personality types*. Bantam.
- Rogers, C. R. (1959). A theory of therapy, personality and interpersonal relationships. *Psychotherapy: Theory, Research & Practice*, 1(1), 1–10. <https://doi.org/10.1037/h0084080>
- Russell, S., & Norvig, P. (2003). *Artificial intelligence: A modern approach* (2nd ed.). Prentice Hall.
- Ryan, G., Katarina, P., & Suhartono, D. (2023). MBTI personality prediction using *machine learning* and SMOTE for balancing data based on statement sentences. *Information*, 14(4), 217. <https://doi.org/10.3390/info14040217>
- Sathiya, S., Selvi, S. T., & Gubbiyappa, K. S. (2020). Personality trait prediction using *machine learning* techniques. In *Proceedings of the Third International Conference on Computational Intelligence and Informatics* (pp. 61–70). Springer, Singapore.
- Skinner, B. F. (1953). *Science and human behavior*. The Free Press.
- Stelmack, R. M., & Stalikas, A. (1991). Galen and the humour theory of temperament. *Personality and Individual Differences*, 12(3), 255–263. [https://doi.org/10.1016/0191-8869\(91\)90102-7](https://doi.org/10.1016/0191-8869(91)90102-7)
- World Health Organization. (2023). *World health statistics 2023: Monitoring health for the SDGs, sustainable development goals*.
- Zhang, H. (2023). MBTI personality prediction based on BERT classification. In *Highlights in Science, Engineering and Technology CSIC* (Vol. 2022).