

**PREDIKSI FAKTOR PENDUKUNG KESUKSESAN *FUNDRAISING* DIGITAL
BAZNAS RI MENGGUNAKAN ALGORITMA C4.5**

SKRIPSI

Diajukan Untuk Melengkapi Salah Satu Syarat

Memperoleh Gelar Sarjana Komputer



OLEH:

Yuliana

NIM. 2026016

**PROGRAM STUDI TEKNIK INFORMATIKA
FAKULTAS TEKNIK DAN ILMU KOMPUTER
UNIVERSITAS NAHDLATUL ULAMA INDONESIA
JAKARTA**

2024

LEMBAR PERNYATAAN

Yang bertanda tangan di bawah ini:

Nama : Yuliana
NIM : 2026016
Program Studi : Teknik Informatika
Fakultas : Teknik dan Ilmu Komputer
Jenjang Pendidikan : Strata 1

Menyatakan bahwa skripsi yang saya buat dengan judul:

**PREDIKSI FAKTOR PENDUKUNG KESUKSESAN *FUNDRAISING* DIGITAL
BAZNAS RI MENGGUNAKAN ALGORITMA C4.5**

1. Merupakan hasil karya tulis ilmiah sendiri, bukan merupakan karya yang pernah diajukan untuk memperoleh gelar akademik oleh pihak lain, dan bukan merupakan hasil plagiat.
2. Saya izinkan untuk dikelola oleh UNIVERSITAS NAHDLATUL ULAMA INDONESIA sesuai dengan norma hukum dan etika yang berlaku.

Pernyataan ini saya buat dengan penuh tanggung jawab dan saya bersedia menerima konsekuensi apapun sesuai aturan yang berlaku apabila di Kemudian hari pernyataan ini tidak benar.

Jakarta, 25 Maret 2024

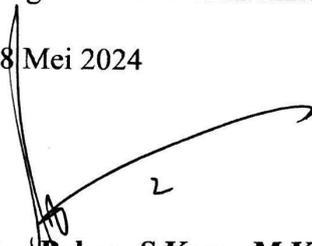

Yuliana

LEMBAR PENGESAHAN

NIM : 2026016
Nama : Yuliana
Program Studi : Teknik Informatika
Fakultas : Teknik dan Ilmu Komputer
Jenjang Pendidikan : Strata 1
Judul Skripsi : PREDIKSI FAKTOR PENDUKUNG KESUKSESAN
FUNDRAISING DIGITAL BAZNAS RI
MENGUNAKAN ALGORITMA C4.5

Skripsi ini telah dipertahankan di hadapan dewan penguji ujian skripsi Fakultas Teknik, program studi Teknik Informatika dan dinyatakan LULUS.

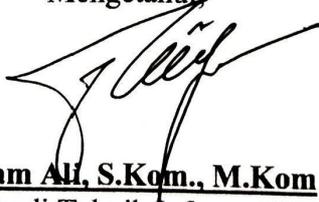
Jakarta, 08 Mei 2024


Sry Dhina Pohan, S.Kom., M.Kom
Penguji I


Andi Aljabar, S.Kom., M.T.I
Penguji II


Binti Mamluatul Karomah, S.Kom., M.Kom
Pembimbing

Mengetahui


Ircham Ali, S.Kom., M.Kom
KaProdi Teknik Informatika

KATA PENGANTAR

Puji syukur Alhamdulillah kehadirat Allah SWT yang telah melimpahkan segala rahmat dan karunia-Nya, sehingga penulis dapat menyelesaikan laporan skripsi yang merupakan salah satu persyaratan untuk menyelesaikan program studi strata satu (S1) pada program studi Teknik Informatika di Universitas Nahdlatul Ulama Indonesia.

Penulis menyadari skripsi ini masih jauh dari sempurna. Karena itu, kritik dan saran akan senantiasa penulis terima dengan senang hati. Dengan segala keterbatasan, penulis menyadari pula bahwa skripsi ini takkan terwujud tanpa bantuan, bimbingan, dan dorongan dari berbagai pihak. Untuk itu, dengan segala kerendahan hati, penulis menyampaikan ucapan terima kasih kepada:

1. BAZNAS RI, yang telah memberikan beasiswa riset kategori umum tahun 2023, sehingga penulis dapat menyelesaikan skripsi dengan baik.
2. Bapak Andrian, selaku Direktur Keamanan Informasi, Data, dan Layanan Digital BAZNAS RI. Telah memberikan kesempatan untuk mendapatkan serta mengolah data muzaki yang diperlukan dalam penelitian ini.
3. Bapak H. Juri Ardiantoro, M.Si., Ph.D., selaku Rektor di Universitas Nahdlatul Ulama Indonesia.
4. Ibu Adrinoviarini, M.Sc., selaku Dekan Fakultas Teknik dan Ilmu Komputer di Universitas Nahdlatul Ulama Indonesia.
5. Bapak Ircham Ali, S.Kom., M.Kom., selaku ketua program studi Teknik Informatika di Universitas Nahdlatul Ulama Indonesia.
6. Ibu Binti Mamluatul Karomah, S.Kom., M.Kom, selaku pembimbing skripsi pada program studi Teknik Informatika di Universitas Nahdlatul Ulama Indonesia.
7. Ibu kandung dan ketiga saudara perempuan tercinta yang telah mendukung penulis baik spirit maupun materi.

8. Sahabat seperjuangan saat mengikuti program kampus Merdeka MSIB Batch 3 yang telah memberikan dukungan moral untuk terus menyelesaikan skripsi ini.
9. Saudara dan sahabat-sahabatku, terutama kawan-kawan angkatan 2020 yang telah memberikan dukungan moral untuk terus menyelesaikan skripsi ini.

Semoga Allah SWT membalas kebaikan dan selalu mencurahkan hidayah serta taufik-Nya, Amin.

Jakarta, 25 Maret 2024

Penulis

ABSTRACT

Zakat fundraising is an essential activity for the National Amil Zakat Agency of the Republic of Indonesia (BAZNAS RI) in supporting various charitable and social programs. This study aims to apply the C4.5 algorithm as a predictive method in increasing the effectiveness of zakat fundraising. The research method used adopts CRISP-DM (Cross-Industry Standard Process for Data Mining), which involves the data preparation stage to process raw data into data that can be processed. Next, the C4.5 algorithm is implemented to build a decision tree model. Muzaki data which includes information on age, gender, occupation group, transaction period, nominal, and nominal category, were used as data input in this study. The results showed that the prediction model developed had an accuracy rate of 92% in predicting the success of zakat fundraising from prospective muzaki. The conclusion of this study is that the application of the C4.5 algorithm as a prediction method in zakat fundraising can help BAZNAS RI in increasing the efficiency and effectiveness of fundraising strategies. With the right prediction model, BAZNAS RI can optimize resource allocation and plan fundraising campaigns more efficiently, so that it can have a greater positive impact in supporting charity and social programs implemented by BAZNAS RI.

Keywords: fundraising, CRISP-DM, C4.5 algorithm, prediction, BAZNAS RI.

ABSTRAK

Fundraising zakat merupakan kegiatan yang esensial bagi Badan Amil Zakat Nasional Republik Indonesia (BAZNAS RI) dalam mendukung berbagai program amal dan sosial. Penelitian ini bertujuan untuk menerapkan algoritma C4.5 sebagai metode prediksi dalam meningkatkan efektivitas *fundraising* zakat. Metode penelitian yang digunakan mengadopsi CRISP-DM (*Cross-Industry Standard Process for Data Mining*), yang melibatkan tahap *data preparation* untuk mengolah data mentah menjadi data yang dapat diproses. Selanjutnya, algoritma C4.5 diimplementasikan untuk membangun model pohon keputusan. Data muzaki yang mencakup informasi tentang umur, gender, occupation group, periode transaksi, nominal, dan kategori nominal, digunakan sebagai input data dalam penelitian ini. Hasil penelitian menunjukkan bahwa model prediksi yang dikembangkan memiliki tingkat akurasi sebesar 92% dalam memprediksi kesuksesan *fundraising* zakat dari calon muzaki. Kesimpulan dari penelitian ini adalah bahwa penerapan algoritma C4.5 sebagai metode prediksi dalam *fundraising* zakat dapat membantu BAZNAS RI dalam meningkatkan efisiensi dan efektivitas strategi *fundraising*. Dengan model prediksi yang tepat, BAZNAS RI dapat mengoptimalkan alokasi sumber daya dan merencanakan kampanye *fundraising* dengan lebih efisien, sehingga dapat memberikan dampak positif yang lebih besar dalam mendukung program amal dan sosial yang dilaksanakan oleh BAZNAS RI.

Kata Kunci: *Fundraising*, CRISP-DM, Algoritma C4.5, Prediksi, BAZNAS RI.

DAFTAR ISI

LEMBAR PERNYATAAN	i
LEMBAR PENGESAHAN	ii
KATA PENGANTAR.....	iii
ABSTRACT	v
ABSTRAK	vi
DAFTAR ISI.....	vii
DAFTAR TABEL.....	x
DAFTAR GAMBAR.....	xi
DAFTAR RUMUS	xii
DAFTAR LAMPIRAN	xiii
BAB I PENDAHULUAN.....	1
1.1 Latar Belakang	1
1.2 Rumusan Masalah	3
1.3 Batasan Masalah	3
1.4 Tujuan Penelitian.....	4
1.5 Manfaat Penelitian.....	4
1.6 Metode Penelitian	5
1.7 Sistematika Penulisan	8
BAB II LANDASAN TEORI	10

2.1	Penelitian Terdahulu	10
2.2	Landasan Teori.....	13
	2.2.2 Zakat.....	15
	2.2.3 Jupyter Notebook.....	16
	2.2.4 <i>Data Mining</i>	17
	2.2.5 Algoritma C4.5	20
	2.2.6 <i>Confusion Matrix</i>	21
	2.2.7 <i>Visual Studio Code</i>	22
	2.2.8 Streamlit.....	23
BAB III METODOLOGI PENELITIAN		25
3.1.	Kerangka Berpikir	25
3.2.	Metodologi Penelitian.....	27
	3.1.1 <i>Business Understanding</i> (Pemahaman Bisnis).....	28
	3.1.2 <i>Data Understanding</i> (Pemahaman Data).....	28
	3.1.3 <i>Data Preparation</i> (Persiapan Data)	30
	3.1.4 <i>Modeling</i> (Pemodelan).....	31
	3.1.5 <i>Evaluation</i> (Pengujian).....	34
	3.1.6 <i>Deployment</i> (Penyebaran)	34
3.3	<i>Problem Solving</i>	36
	3.3.1 Pemahaman yang Lebih Mendalam	36
	3.3.2 Prediksi yang Akurat	36
	3.3.3 Segmentasi Audiens yang Tepat.....	36
	3.3.4 Optimasi Penggunaan Sumber Daya	37
	3.3.5 Perbaikan Berkelanjutan.....	37

BAB IV IMPLEMENTASI DAN EKSPERIMEN.....	38
4.1 Persiapan Data.....	38
4.1.1 Pengumpulan Data	38
4.1.2 Pengolahan Data.....	38
4.2 Implementasi Model	44
4.3 Hasil dan Analisis.....	48
4.3.1 Evaluasi Kinerja Model	48
4.3.2 Hasil Eksperimen.....	52
BAB V PENUTUP.....	57
5.1 Kesimpulan	57
5.2 Saran	57
DAFTAR PUSTAKA	59
LAMPIRAN.....	64

DAFTAR TABEL

Tabel 2. 1 Pemetaan Penelitian Terdahulu.....	10
Tabel 3. 1 Sampling Data.....	32
Tabel 3. 2 Contoh perhitungan <i>entropy</i> dan <i>information gain</i>	33
Tabel 4. 1 Data Mentah.....	39
Tabel 4. 2 Data setelah <i>cleaning</i>	40
Tabel 4. 3 Data setelah <i>transformation</i>	42
Tabel 4. 4 <i>Data Training</i>	43
Tabel 4. 5 <i>Data Testing</i>	43
Tabel 4. 6 Nilai <i>entropy</i> setiap variabel	45
Tabel 4. 7 Nilai <i>information gain</i> setiap variabel	47
Tabel 4. 8 Dataset akhir untuk membangun model.....	47
Tabel 4. 9 Tabel <i>confusion matrix</i>	49
Tabel 4. 10 <i>Classification report</i>	51
Tabel 4. 11 Data uji coba prediksi	52
Tabel 4. 12 Hasil data setelah dilakukan prediksi.....	53

DAFTAR GAMBAR

Gambar 1. 1 Flowchart Metode Penelitian	6
Gambar 3. 1 Kerangka Berpikir	25
Gambar 3. 2 Metodologi CRISP-DM	27
Gambar 3. 3 Flowchart Algoritma C4.5.....	31
Gambar 4. 1 <i>Source code encoder</i>	40
Gambar 4. 2 <i>Source code</i> normalisasi kolom nominal	41
Gambar 4. 3 <i>Source code</i> metode elbow.....	41
Gambar 4. 4 Visualisasi metode elbow	42
Gambar 4. 5 <i>Pipeline machine learning</i> algoritma C4.5	44
Gambar 4. 6 <i>Source code</i> perhitungan entropy setiap variabel.....	45
Gambar 4. 7 <i>Source code</i> perhitungan <i>weighted entropy</i>	46
Gambar 4. 8 <i>Source code</i> perhitungan <i>information gain</i> setiap variabel.....	46
Gambar 4. 9 <i>Source code confusion matrix</i>	49
Gambar 4. 10 Hasil eksperimen data kelompok	54
Gambar 4. 11 Visualisasi hasil eksperimen	54
Gambar 4. 12 Hasil Eksperimen Data Tunggal	55

DAFTAR RUMUS

Rumus <i>Gain</i> (1).....	20
Rumus <i>Entropy</i> (2).....	21
Rumus <i>Accuracy</i> (3).....	22
Rumus <i>Precision</i> (4)	22
Rumus <i>Recall</i> (5).....	22

DAFTAR LAMPIRAN

Lampiran 1. Data mentah sebelum dilakukan nya prediksi.....	64
Lampiran 2. Data hasil prediksi.....	65

BAB I PENDAHULUAN

1.1 Latar Belakang

Penggunaan teknologi dalam *fundraising* dana zakat memiliki potensi besar untuk meningkatkan kesadaran dan partisipasi umat muslim di seluruh dunia. *Fundraising* adalah suatu aktivitas yang bertujuan untuk mengumpulkan dana atau sumber daya lainnya dari masyarakat, baik individu maupun organisasi, guna mendukung pembiayaan program dan kegiatan operasional organisasi atau lembaga dengan tujuan mencapai suatu tujuan tertentu (Perdana & Zen, 2020). Beberapa lembaga sudah melakukan kerjasama dengan *e-commerce*, aplikasi *online* dan sejenisnya untuk memperoleh manfaat dari *fintech* (Febiana et al., 2022).

Zakat merupakan salah satu rukun Islam yang harus dilaksanakan oleh umat muslim. Menurut (*World Giving Index 2022 A Global View of Giving Trends, 2022*), Indonesia merupakan negara paling dermawan di dunia dan memiliki skor *World Giving Index* tertinggi dalam lima tahun berturut-turut. Indonesia mempertahankan peringkat teratasnya dengan Indeks 68%, membantu orang asing (58%), memiliki tingkat donasi tertinggi (84%) dan menjadi sukarelawan (63%) di dunia. Dana ZIS yang diperoleh melalui *fundraising* memiliki dampak yang signifikan terhadap keberadaan sebuah lembaga pengelola ZIS. Hal ini dikarenakan dana ZIS tersebut dapat berkontribusi dalam meningkatkan kesejahteraan masyarakat serta membantu mengurangi tingkat kemiskinan yang ada di negara ini (Alfian & Widodo, 2022).

Badan Amil Zakat Nasional (BAZNAS) merupakan badan resmi dan satu-satunya yang dibentuk oleh pemerintah berdasarkan Keputusan Presiden RI No. 8 Tahun 2001 (Keputusan Presiden Republik Indonesia Nomor 8, 2001) yang memiliki tugas dan fungsi menghimpun dan menyalurkan zakat, infak, dan sedekah (ZIS) pada tingkat nasional. Lahirnya Undang-Undang Nomor 23 Tahun

2011 (Undang-Undang Nomor 23, 2011) tentang Pengelolaan zakat semakin mengukuhkan peran BAZNAS sebagai lembaga yang berwenang melakukan pengelolaan zakat secara nasional. Dalam UU tersebut, BAZNAS dinyatakan sebagai lembaga pemerintah nonstruktural yang bersifat mandiri dan bertanggung jawab kepada Presiden melalui Menteri Agama (BAZNAS, 2019). Salah satu tujuan BAZNAS adalah terwujudnya penyaluran zakat, infaq dan sedekah (ZIS) serta dana sosial keagamaan lainnya (DSKL) yang efektif dalam pengentasan kemiskinan, peningkatan kesejahteraan umat, dan pengurangan kesenjangan sosial. Berdasarkan laporan target *fundraising* dana zakat pada tahun depan yakni 33 Miliar. Oleh karena itu, sangat penting untuk memprediksi dan memahami faktor-faktor yang dapat mempengaruhi kesuksesan *fundraising*, terutama pada Lembaga Negara yang memiliki fokus pada proyek sosial dan kemanusiaan seperti Badan Amil Zakat Nasional.

Menurut *Gartner Group* dalam Larose, data mining adalah suatu proses menemukan hubungan yang berarti, pola, dan kecenderungan dengan memeriksa dalam sekumpulan besar data yang tersimpan dalam penyimpanan dengan menggunakan teknik pengenalan pola seperti teknik statistik dan matematika (Dewi, 2020). Algoritma C4.5 adalah algoritma *data mining* yang digunakan untuk melakukan klasifikasi dan estimasi. Algoritma C4.5 termasuk dalam algoritma induksi *decision tree* yang dikembangkan oleh J. Ross Quinlan yang dikenal dengan nama ID3 (*Iterative Dichotomiser 3*) (Suweleh et al., 2020). Algoritma ini dapat digunakan untuk memprediksi apakah suatu proyek *fundraising* akan berhasil mencapai target *fundraising* dana atau tidak. Sebelum diimplementasikan model Algoritma dilakukan evaluasi menggunakan *confusion matrix*. *Confusion matrix* dapat menghitung berbagai matrik evaluasi, termasuk *accuracy*, *precision*, *recall*, dan lainnya. Matrik-matrik ini memberikan wawasan tentang kinerja model dalam mengklasifikasikan data dan membantu memahami jenis kesalahan yang dibuat oleh model (Paskalis et al., 2019).

Dengan mengimplementasikan algoritma C4.5 pada data proyek-proyek *fundraising* dapat diidentifikasi faktor-faktor apa saja yang berpengaruh pada kesuksesan *fundraising digital* dan bagaimana faktor-faktor tersebut dapat dioptimalkan untuk meningkatkan kesuksesan *fundraising digital* Badan Amil

Zakat Nasional. Oleh karena itu, penelitian ini akan membahas mengenai implementasi algoritma C4.5 berbasis prediksi sebagai faktor pendukung kesuksesan *fundraising digital* dengan studi kasus BAZNAS RI. Penelitian ini diharapkan dapat membantu BAZNAS dalam *fundraising* dana zakat, infak, dan sedekah. Sesuai Visi BAZNAS menjadi lembaga utama mensejahterakan umat. Dalam penelitian ini *fundraising* yang digunakan adalah untuk kasus filantropi atau *fundraising* zakat, infak, dan sedekah (ZIS).

1.2 Rumusan Masalah

Berdasarkan latar belakang yang telah dipaparkan, berikut adalah rumusan masalah pada penelitian ini:

1. Bagaimana membuat model yang bisa membantu BAZNAS agar kampanye *fundraising digital* yang diselenggarakan mengalami keberhasilan?
2. Apa saja faktor yang dapat mempengaruhi kesuksesan *fundraising digital* BAZNAS?

1.3 Batasan Masalah

Berikut adalah batasan masalah yang digunakan agar penelitian lebih fokus dan terarah:

1. Penelitian hanya dilakukan pada *platform* BAZNAS, dengan data yang diambil dari *database* berupa jumlah kampanye baik yang telah selesai maupun kampanye yang sedang berjalan. Data tersebut akan digunakan untuk *data mining* dengan algoritma C4.5.
2. Metode *data mining* yang digunakan adalah *Cross-Industry Standard Process for Data Mining* (CRISP-DM)
3. Data diambil hanya sekali pada waktu pengambilan data dan tidak ada *update* data kembali pada hari setelahnya.

4. Data yang diambil *fundraising* ZIS-DSKL Muzaki di BAZNAS Pusat dengan jumlah data sebanyak 7896.
5. Atribut dan variabel mencakup umur, gender, *occupation group*, periode transaksi, dan kategori nominal yang digunakan sebagai input data yang dapat mempengaruhi partisipasi dan jumlah donasi.
6. Pengukuran evaluasi dari algoritma C4.5 menggunakan *Confusion Matrix* untuk melihat *accuracy*, *precision*, dan *recall*.

1.4 Tujuan Penelitian

Adapun tujuan yang hendak dicapai dalam penelitian ini adalah sebagai berikut :

1. Membuat model prediksi agar kampanye *fundraising digital* yang diselenggarakan oleh BAZNAS RI mengalami keberhasilan.
2. BAZNAS dapat mengidentifikasi faktor-faktor yang mempengaruhi kesuksesan *fundraising digital* melalui *e-commerce* dan aplikasi *online*, maka BAZNAS memiliki kesempatan untuk meningkatkan kinerjanya berdasarkan faktor-faktor tersebut.

1.5 Manfaat Penelitian

Manfaat yang diharapkan dari penelitian ini antara lain sebagai berikut :

1. Bagi Perguruan Tinggi

Penelitian ini diharapkan dapat menjadi bahan referensi tambahan, terutama dalam hal pemanfaatan dan pengolahan data untuk menghasilkan informasi yang bermanfaat bagi civitas akademika.

2. Bagi Badan Amil Zakat Nasional

Penelitian ini bertujuan untuk menyediakan informasi mengenai faktor-faktor yang mempengaruhi kesuksesan *fundraising digital*.

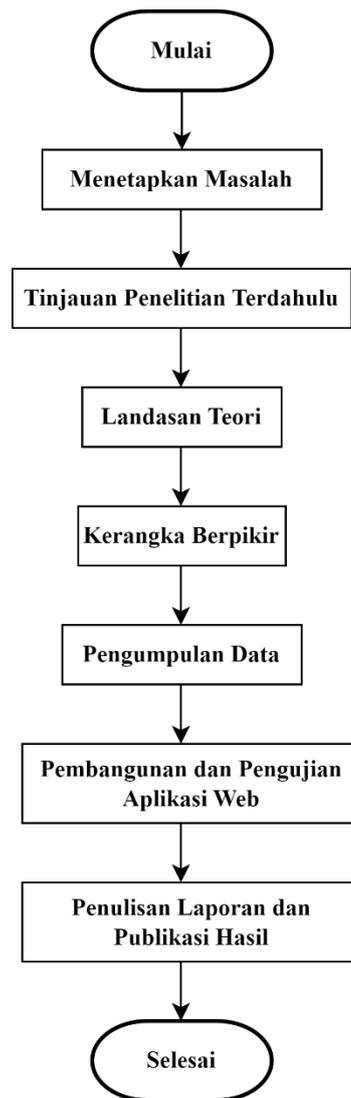
Informasi ini dapat digunakan untuk mengoptimalkan faktor-faktor tersebut guna meningkatkan efektivitas *fundraising* zakat.

3. Bagi Penulis

Penelitian ini memenuhi salah satu syarat kelulusan dari program Strata 1 di Program Studi Teknik Informatika, Fakultas Teknik dan Ilmu Komputer, Universitas Nahdlatul Ulama Indonesia. Selain itu, penelitian ini juga berkontribusi dalam meningkatkan wawasan dan keterampilan penulis, sekaligus memperkuat kompetensi intelektual.

1.6 Metode Penelitian

Penelitian ini dilakukan melalui beberapa tahapan. Adapun metode yang digunakan dalam tiap-tiap tahapan antara lain yaitu menetapkan masalah, tinjauan penelitian terdahulu, landasan teori, kerangka berpikir, pengumpulan data, pembangunan dan pengujian aplikasi web, serta penulisan laporan dan melakukan publikasi hasil.



Gambar 1. 1 Flowchart Metode Penelitian

Berikut penjelasan dari *flowchart* metode penelitian sebagai pedoman serangkaian pelaksanaan penelitian hingga selesai.

1. Menetapkan Masalah

Penelitian ini bertumpu pada tantangan yang dihadapi oleh Badan Amil Zakat Nasional (BAZNAS) dalam mengoptimalkan kampanye *fundraising digital*, menemukan solusi praktis dan strategis yang dapat diimplementasikan untuk memaksimalkan potensi *fundraising* zakat secara digital.

2. Tinjauan Penelitian Terdahulu

Melakukan tinjauan penelitian terdahulu yang bertujuan untuk mengumpulkan dan mengevaluasi literatur terkait yang sudah ada untuk membangun fondasi teoritis penelitian. Dalam tahap ini, peneliti menelusuri berbagai sumber seperti jurnal ilmiah, buku, artikel, dan dokumen lain yang relevan dengan topik penelitian. Tujuannya adalah untuk mengidentifikasi, menganalisis, dan mendiskusikan temuan-temuan sebelumnya, metodologi yang digunakan, serta kesenjangan yang masih ada dalam studi yang telah dilakukan. Tinjauan ini membantu peneliti dalam memahami tentang topik yang diteliti, menentukan metodologi yang tepat, dan merumuskan hipotesis yang kuat berdasarkan kerangka teoritis yang telah ditetapkan dan kekosongan penelitian yang ada.

3. Landasan Teori

Landasan teori dalam penelitian merupakan kerangka konseptual yang digunakan untuk mendukung proses penelitian dan menjelaskan fenomena yang sedang diteliti. Ini melibatkan pemilihan teori-teori relevan yang telah diakui dalam literatur akademik yang menyediakan dasar pemahaman dan orientasi bagi penelitian.

4. Kerangka Berpikir

Kerangka berpikir ini bertujuan untuk memberikan panduan dan struktur dalam mengimplementasi algoritma C4.5 sebagai faktor pendukung kesuksesan fundraising digital. Dimulai dari masalah, eksperimen, pengujian dan validasi hasil, serta hasil dari eksperimen.

5. Pengumpulan Data

Mengumpulkan data menggunakan teknik pengumpulan data dokumentasi. Dalam melaksanakan pengumpulan data, peneliti memastikan bahwa data yang diperoleh adalah valid serta mendapatkan persetujuan dari *stakeholder* untuk menjaga kerahasiaan data, dan memastikan transparansi dalam proses pengumpulan data. Data dokumentasi ini dari data muzaki BAZNAS RI dilingkup pusat yang disimpan pada penyimpanan *Cloud Storage Google*. Terdapat total 7896 data primer dari transaksi muzaki yang diambil dari tahun 2022 dan 2023.

6. Pembangunan dan Pengujian Aplikasi Web

Pembangunan dan pengujian aplikasi web untuk prediksi fundraising digital menggunakan metode CRISP-DM (*Cross-Industry Standard Process for Data Mining*) meliputi beberapa tahapan yaitu *Business Understanding, Data Understanding, Data Preparation, Modeling, Evaluation, dan Deployment*.

7. Penulisan Laporan dan Publikasi Hasil

Tahap penulisan laporan dan publikasi hasil merupakan bagian akhir dari proses penelitian, yang melibatkan dokumentasi menyeluruh dari seluruh proses dan hasil penelitian. Setelah laporan selesai dan disempurnakan melalui proses revisi, langkah selanjutnya adalah publikasi hasil. Publikasi dilakukan dengan melakukan pengajuan artikel ke jurnal ilmiah Tujuan utama dari publikasi adalah untuk berkontribusi pada pengetahuan yang ada, membagikan wawasan baru dengan komunitas ilmiah dan praktisi, serta mendapatkan pengakuan atas pekerjaan penelitian yang telah dilakukan.

1.7 Sistematika Penulisan

Sistem penulisan dalam tugas akhir ini disusun dalam bentuk karya ilmiah dengan struktur penulisan sebagai berikut:

BAB 1 PENDAHULUAN

Bab ini berisi uraian tentang latar belakang masalah yang mendasari pentingnya diadakan penelitian, perumusan dan pembatasan masalah penelitian, tujuan penelitian, metode penelitian, dan sistematika penulisan.

BAB 2 LANDASAN TEORI

Bab ini berisi tinjauan penelitian terdahulu dan landasan teori yang mendeskripsikan pengertian *fundraising*, zakat, jupyter notebook, *data mining*, algoritma c4.5, *confusion matrix*, visual studio code, dan streamlit.

BAB 3 METODOLOGI PENELITIAN

Bab ini berisi tentang kerangka berpikir, metodologi penelitian dengan menerapkan CRISP-DM, dan *problem solving*.

BAB 4 IMPLEMENTASI DAN EKSPERIMEN

Bab ini berisi tentang penjelasan prosedur testing dan proses pengujian sistem yang telah dibuat.

BAB 5 PENUTUP

Bab ini berisi pokok-pokok kesimpulan dan saran-saran yang perlu disampaikan kepada pihak-pihak yang berkepentingan dengan hasil penelitian.

DAFTAR PUSTAKA

Daftar pustaka berisi daftar referensi yang digunakan dalam penelitian.

LAMPIRAN

Lampiran berisi tautan seluruh source code pembangunan aplikasi web, data mentah, dan data setelah dibersihkan (*cleaning*).

BAB II

LANDASAN TEORI

2.1 Penelitian Terdahulu

Dalam penulisan skripsi ini peneliti menggali informasi dari penelitian-penelitian sebelumnya yang akan dijabarkan pada Tabel 2.1 sebagai bahan perbandingan, baik mengenai kekurangan atau kelebihan yang sudah ada.

Tabel 2. 1 Pemetaan Penelitian Terdahulu

No	Judul, Penulis, Tahun	Objek	Metode	Kekurangan
1.	Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5 (Gaol, 2020)	Pendidikan	Algoritma C4.5	Tidak ada nya <i>testing model</i> , <i>evaluation</i> dan <i>deployment</i> .
2.	Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 PT. Murni Indah Sentosa (Fikri & Verina, 2020)	Kesehatan	Algoritma C4.5	Kurang nya jumlah atribut atau variabel hingga tidak ada batasan nya. Program aplikasi belum bisa di <i>update</i> secara <i>online</i> . Tidak ada <i>evaluation</i> sehingga tidak diketahui tingkat akurasi nya.
3.	Analisis Pengembangan Model Prediksi Kesuksesan Kickstarter Menggunakan Algoritma Backpropagation dan Random Forest (Agustina et al., 2022)	<i>Crowdfunding</i> (Kickstarter)	Backpropagation ANN dan Random Forest	Tidak ada nya <i>testing model</i> dan <i>deployment</i> .
4.	<i>Using Machine Learning Approach Towards Successful Crowdfunding Prediction</i> (Rafflesia et al., 2023)	<i>Crowdfunding</i> (Kickstarter)	Machine Learning; algoritma <i>logistic regression</i> , <i>random forest</i> , dan <i>extreme gradient boosting</i> (XGBoost)	Tidak menyebutkan variabel atau <i>feature</i> yang digunakan.
5.	Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma <i>Decision tree C4.5</i> (Paskalis et al., 2019)	Pendidikan	Algoritma C4.5	Jumlah data uji yang digunakan masih terbatas.

6.	Penerapan Algoritma C4.5 Untuk Prediksi Churn Rate Pengguna Jasa Telekomunikasi. (Utami et al., 2020)	Jasa	Algoritma C4.5	Terbatasnya jumlah data, variabel yang tidak terpakai, dan tidak adanya analisis lebih lanjut tentang faktor-faktor penyebab perpindahan pelanggan.
----	---	------	----------------	---

Penelitian pertama oleh (Gaol, 2020), dengan judul “Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5”. Penelitian yang dilakukan adalah untuk memprediksi mahasiswa yang berpotensi menjadi non-aktif pada Sekolah Tinggi Manajemen Informatika dan Komputer (STMIK) Triguna Dharma Medan. Didapatkan hasil gain tertinggi dengan atribut nilai gagal sebesar 0,503. Sehingga menghasilkan kriteria-kriteria yang tepat dalam menganalisa mahasiswa yang berpotensi non aktif.

Penelitian kedua oleh (Fikri & Verina, 2020) “Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 PT. Murni Indah Sentosa”. Penelitian ini berkaitan dengan penggunaan algoritma data mining C4.5 yang dapat membantu perusahaan dalam memprediksi penjualan dan mengelola persediaan dengan efektif. Didapatkan hasil gain terbesar adalah Gain (Jumlah Terjual). Maka Jumlah Terjual menjadi node akar (*root node*).

Penelitian ketiga oleh (Agustina et al., 2022), dengan judul “Analisis Pengembangan Model Prediksi Kesuksesan Kickstarter Menggunakan Algoritma Backpropagation dan Random Forest”. Penelitian ini berkaitan dengan penggalangan dana melalui platform Kickstarter dan bertujuan untuk membuat model prediksi yang dapat membantu para *fundraiser* dan *funder* dalam memprediksi apakah proyek Kickstarter yang akan dikerjakan bisa sukses atau tidak. Pembangunan model machine learning dengan menggunakan 2 algoritma. Model pertama menggunakan algoritma backpropagation ANN dan random forest memberikan akurasi tertinggi masing-masing sebesar (89%, 98%) Sedangkan model kedua memberikan akurasi tertinggi masing-masing sebesar (69%, 65.7%). Model yang akan dipakai untuk memprediksi kesuksesan kickstarter adalah model

pertama dengan menggunakan algoritma random forest yang merupakan model terbaik dengan nilai akurasi 98% dan nilai f1-score nya 98%.

Penelitian keempat oleh (Raflesia et al., 2023), dengan judul “Using Machine Learning Approach Towards Successful Crowdfunding Prediction”. Penelitian ini bertujuan untuk membuat model prediksi kesuksesan proyek Kickstarter yang melibatkan penggunaan *machine learning* untuk memprediksi keberhasilan proyek crowdfunding di platform Kickstarter antara 2020-2022. Data yang sudah dikumpulkan dan disiapkan akan digunakan untuk membuat model machine learning dengan menggunakan algoritma *logistic regression*, *random forest*, dan *extreme gradient boosting* (XGBoost). Didapatkan hasil tingkat akurasi pada Logistic regression 84%, Logistic regression + PCA 84%, Random forest + PCA 82%, XGBoost + PCA 83%, logistic regression dengan log-transformed data 86%, Random forest dengan log-transformed data 86%, dan XGBoost dengan log-transformed data 86%. Hasil eksperimen menunjukkan bahwa algoritma XGBoost memiliki kinerja terbaik di antara yang lain.

Penelitian kelima oleh (Paskalis et al., 2019), dengan judul Model Prediksi Awal Masa Studi Mahasiswa Menggunakan Algoritma *Decision tree C4.5*. Penelitian ini bertujuan untuk membangun sebuah model untuk prediksi awal masa studi mahasiswa pada Program Studi Informatika Universitas Tanjungpura. Hasil penelitian ini menunjukkan bahwa model yang dibangun menggunakan algoritma Decision Tree C4.5 efektif dalam memprediksi masa studi mahasiswa. Karena nilai precision, recall, dan accuracy yang dihasilkan menunjukkan angka rata-rata diatas 50%. Sedangkan nilai error rate rata-rata pengujian dibawah 20% menunjukkan bahwa kesalahan prediksi yang dilakukan rendah.

Penelitian keenam oleh (Utami et al., 2020), dengan judul “Penerapan Algoritma C4.5 Untuk Prediksi Churn Rate Pengguna Jasa Telekomunikasi”. Penelitian ini bertujuan untuk menerapkan teknik klasifikasi untuk mengukur nilai akurasi dalam memprediksi penyebab perpindahan pengguna jasa telekomunikasi. Hasil penelitian menunjukkan bahwa penggunaan algoritma klasifikasi C4.5 dalam memprediksi penyebab perpindahan pelanggan di industri telekomunikasi

berhasil. Model yang dikembangkan mencapai tingkat akurasi sebesar 87%, precision sebesar 87,5%, dan recall sebesar 96%. Hasil evaluasi juga menunjukkan bahwa atribut harga merupakan faktor yang paling menarik dalam prediksi perpindahan pelanggan.

Berdasarkan analisis dari penelitian-penelitian sebelumnya, diusulkan untuk melakukan penelitian dengan menggunakan algoritma C4.5 untuk melakukan prediksi kesuksesan pada platform *fundraising* BAZNAS. Dalam memilih metodologi, dibandingkan beberapa metode yang digunakan dan diputuskan untuk menggunakan metodologi data mining CRISP-DM. Penggunaan CRISP-DM merujuk pada Standar Kompetensi Kerja Nasional: KepMen Ketenagakerjaan No 299 tahun 2020 (Pambudi & Abidin, 2023). Dalam keenam penelitian sebelumnya, tidak terdapat proses *deployment*, oleh karena itu, dipilih penggunaan metodologi terbaik yaitu CRISP-DM. Proses *deployment* akan dilakukan dengan menggunakan Streamlit, sebuah *framework open-source* yang menyediakan fitur *deployment* yang mudah, sehingga memungkinkan pengguna untuk dengan cepat membagikan aplikasi web yang telah dibuat kepada pengguna lain. Dalam penelitian sebelumnya, diperlukan penambahan jumlah atribut atau variabel agar ada batasan, serta penambahan jumlah data sebagai variabel pembanding. Oleh karena itu, akan digunakan dataset yang jumlahnya cukup besar agar penelitian ini dapat dilakukan dengan sebaik-baiknya.

2.2 Landasan Teori

Melalui penerapan algoritma C4.5 dalam *fundraising digital*, organisasi zakat dapat memanfaatkan *machine learning* untuk memprediksi dan mengoptimalkan strategi *fundraising* dana secara lebih efektif, meningkatkan partisipasi muzaki, serta meningkatkan dampak sosial dari dana zakat yang terkumpul.

2.2.1 *Fundraising*

Fenomena *fintech* telah membawa dampak yang sangat signifikan pada sektor keuangan, merubah paradigma tradisional, dan menciptakan peluang inovasi bagi lembaga-lembaga terkait, termasuk organisasi zakat (Hudaefi et al., 2020). *Fundraising* digital zakat menjadi salah satu inovasi yang mengambil manfaat dari kemajuan teknologi informasi dan komunikasi dalam mendekatkan proses *fundraising* digital zakat kepada masyarakat. Dengan memanfaatkan platform digital seperti aplikasi, situs web, dan media sosial, *fundraising* digital zakat berhasil membuka pintu bagi lebih banyak potensi muzaki atau pemberi zakat untuk berpartisipasi dalam membantu sesama (Doddy et al., 2022).

Dalam *fundraising* digital zakat, proses pengumpulan dana zakat menjadi lebih mudah dan efisien. Sebelumnya, muzaki harus datang langsung ke lembaga zakat atau mengirimkan dana zakat secara konvensional, namun sekarang mereka dapat berzakat dengan cepat dan nyaman melalui platform digital. Hal ini menarik perhatian khalayak yang lebih luas, terutama generasi muda yang sangat berorientasi pada teknologi. Selain memudahkan muzaki, *fundraising* digital zakat juga membawa manfaat bagi lembaga zakat itu sendiri. Dengan menerapkan platform digital, lembaga zakat dapat mengelola dan memantau dana yang terkumpul dengan lebih transparan dan akurat. Informasi mengenai penggunaan dana zakat juga dapat diakses dengan mudah oleh muzaki, sehingga mereka merasa lebih percaya dan yakin bahwa dana yang mereka sumbangkan digunakan dengan tepat (Alim & Z. Basri, 2020).

Selain itu, *fundraising* digital zakat membuka peluang bagi lembaga zakat untuk menggalang dana secara lebih kreatif. Melalui media sosial, lembaga zakat dapat menyampaikan cerita-cerita inspiratif tentang program-program bantuan yang telah berhasil dilaksanakan. Meskipun *fundraising digital* zakat memiliki potensi besar dalam meningkatkan partisipasi masyarakat dalam berzakat, perlu diingat bahwa pendekatan ini juga perlu mengatasi beberapa tantangan. Misalnya, kekhawatiran mengenai keamanan data dan transaksi *online* perlu ditangani dengan serius agar masyarakat merasa nyaman dan percaya saat berzakat melalui platform digital. Dengan adanya *fundraising* digital zakat, harapan terbuka lebar

untuk menjadikan zakat sebagai instrumen yang lebih kuat dalam mewujudkan keadilan sosial dan membantu masyarakat yang membutuhkan. Kolaborasi antara teknologi dan kepedulian sosial dapat menjadi kunci dalam menciptakan transformasi positif dalam pengumpulan dana zakat, membawa dampak besar bagi masyarakat yang membutuhkan dan memberikan manfaat jangka panjang bagi lembaga zakat itu sendiri (Febiana et al., 2022).

2.2.2 Zakat

Zakat berasal dari kata “zaka”, memiliki makna yang kaya akan nilai dan filosofi. Zakat mengandung berbagai makna di antaranya adalah “al-Numu” yang berarti tumbuh dan berkembang, “al-Ziyadah” yang berarti tambahan, “al-Barakah” yang berarti berkah, dan “al-Thohir” yang berarti penyucian. Makna-makna ini mencerminkan sifat dan tujuan zakat sebagai suatu bentuk kewajiban bagi umat Muslim untuk menyisihkan sebagian dari kekayaan mereka untuk tujuan tertentu. Dalam konteks definisi, Majma Lughah al-Arabiyyah al-Mu'jam al-Wasith mengartikan zakat sebagai bagian tertentu dari kekayaan yang harus diberikan oleh setiap Muslim dengan syarat-syarat tertentu kepada penerima manfaat yang sah. Zakat merupakan salah satu dari lima rukun Islam yang harus dijalankan oleh setiap Muslim yang mampu. Selain sebagai kewajiban ibadah, zakat juga memiliki tujuan sosial dan ekonomi yang lebih luas (Alim & Z. Basri, 2020).

Salah satu tujuan utama membayar zakat adalah untuk mendistribusikan kekayaan dalam rangka mencapai kesetaraan ekonomi dan keadilan sosial. Dengan menyisihkan sebagian dari kekayaan mereka, orang-orang yang lebih beruntung memberikan kontribusi yang signifikan dalam membantu orang-orang yang membutuhkan, sehingga tercipta lingkungan sosial yang lebih adil dan berempati. Pembayaran zakat juga berfungsi sebagai "al-Barakah", yaitu berkah yang membawa keberlimpahan dan keberuntungan bagi yang melaksanakannya. Zakat dapat membersihkan hati dan jiwa seseorang dari keserakahan dan materialisme, dan membuka pintu rezeki dan berkah dari Allah (Karim et al., 2022).

Zakat juga memiliki dimensi "al-Ziyadah", yang berarti tambahan. Selain meningkatkan kualitas hidup penerima zakat, pembayaran zakat juga memberikan manfaat tambahan bagi muzaki (orang yang membayar zakat) dengan memperoleh pahala dan keberkahan dari Allah atas amal kebbaikannya. Dengan memahami makna dan tujuan zakat, umat Muslim diharapkan untuk menjalankan kewajibannya dengan ikhlas dan penuh rasa tanggung jawab. Zakat bukan hanya sekadar pembayaran wajib, tetapi juga merupakan sarana untuk mencapai tujuan mulia dalam membangun masyarakat yang berkeadilan, berempati, dan berkah. Dengan berbagi dari kekayaan yang Allah anugerahkan kepada mereka, umat Muslim dapat berkontribusi secara aktif dalam menciptakan perubahan positif dalam kehidupan sesama dan meraih pahala dari-Nya (Karim et al., 2022).

2.2.3 Jupyter Notebook

Jupyter Notebook adalah sebuah aplikasi web *open-source* yang menawarkan berbagai keunggulan bagi para penggunanya. Dalam Jupyter Notebook, pengguna dapat membuat dan berbagi dokumen yang kaya akan konten, menggabungkan teks naratif, kode program, visualisasi, dan elemen-elemen interaktif lainnya (Mait et al., 2022). Kemampuan untuk menyatukan teks dan kode program dalam satu dokumen membuat eksplorasi data dan prediksi data menjadi lebih interaktif dan efisien. Pengguna dapat menjalankan kode program dan melihat hasilnya dengan cepat, sehingga memungkinkan eksplorasi data secara langsung dan iteratif. Selain itu, Jupyter Notebook juga menyediakan beragam *library* visualisasi yang memudahkan pengguna untuk memvisualisasikan data dengan lebih jelas dan menarik. Jupyter Notebook berjalan di dalam web browser dan dapat diinstal melalui paket Python seperti Anaconda atau dengan menggunakan manajer paket pip. Selain sebagai alat eksplorasi data, Jupyter Notebook juga berguna dalam melakukan prediksi data. Dengan menyediakan lingkungan pemrograman yang lengkap, pengguna dapat menggunakan berbagai algoritma dan *library machine learning* untuk melakukan prediksi terhadap data yang ada. Kelebihan lain dari Jupyter Notebook adalah kemampuannya dalam mendokumentasikan langkah-langkah prediksi dan analisis

data. Pengguna dapat menyertakan teks naratif untuk menjelaskan proses analisis dan interpretasi hasil, sehingga dokumentasi menjadi lebih lengkap dan mudah dimengerti. Selain itu, Jupyter Notebook juga memungkinkan pengguna untuk berbagi kode dan hasil dengan orang lain. Dokumen Jupyter dapat diekspor ke berbagai format, seperti HTML, PDF, atau slide presentasi, sehingga pengguna dapat dengan mudah berbagi hasil penelitian kepada orang lain.

Python sebagai bahasa pemrograman utama yang digunakan dalam Jupyter Notebook juga memiliki banyak kelebihan. Python adalah bahasa pemrograman yang terkenal karena kemampuannya dalam mendukung pemrograman berorientasi objek dan berbagai paradigma pemrograman lainnya. Selain itu, Python dapat dijalankan di berbagai platform sistem operasi seperti PCs, Macintosh, dan UNIX sehingga membuatnya menjadi pilihan yang fleksibel untuk berbagai keperluan (Ginting et al., 2020). Beberapa kelebihan Python antara lain pengembangan program yang lebih cepat dengan kode yang lebih sedikit, dukungan multi platform yang memudahkan portabilitas program, kemudahan dalam mempelajari bahasa Python bagi pemula, sistem pengelolaan memori yang otomatis yang mengurangi beban kerja pengguna, serta dukungan penuh untuk paradigma pemrograman berorientasi objek yang membantu dalam membangun program yang lebih modular dan mudah dipahami.

Kombinasi Jupyter Notebook dan bahasa pemrograman Python menjadi pilihan yang kuat bagi para *data scientist* dan pengembang aplikasi untuk menjalankan eksplorasi data, prediksi data, serta membuat laporan dan presentasi interaktif. Kelebihan-kelebihan dari Jupyter Notebook dan Python menciptakan lingkungan kerja yang efisien, interaktif, dan efektif dalam menghadapi tantangan analisis data dan pengembangan program.

2.2.4 Data Mining

Data mining terbagi atas dua kata, Data yaitu kumpulan fakta yang terekam atau sebuah entitas yang tidak memiliki arti dan selama ini terabaikan. *Mining* yaitu proses penambangan sehingga, *data mining* itu dapat diartikan sebagai proses penambangan data yang menghasilkan sebuah *output* (keluaran)

berupa pengetahuan. *Data mining* merupakan proses semi otomatis yang menggunakan teknik statistik, matematika, kecerdasan buatan, dan *machine learning* untuk mengekstraksi dan mengidentifikasi informasi pengetahuan potensial dan berguna yang tersimpan di dalam database besar (Zai, 2022). *Data mining* adalah bagian dari proses KDD (*Knowledge Discovery in Databases*) yang terdiri dari beberapa tahapan seperti pemilihan data, pra pengolahan, transformasi, data mining, dan evaluasi hasil.

Menurut (Takdirillah, 2020) *Data mining* terbagi menjadi beberapa kelompok berdasarkan tugas yang dapat dilakukan, yaitu *classification*, *association*, *clustering*, *prediction*, dan *estimation*.

a. Klasifikasi (*Classification*)

Suatu teknik dengan melihat pada kelakuan dan atribut dari kelompok yang telah didefinisikan. Teknik ini dapat memberikan klasifikasi pada data baru dengan memanipulasi data yang ada yang telah diklasifikasi dan dengan menggunakan hasilnya untuk memberikan sejumlah aturan. Contoh algoritma *classification* seperti k-Nearest Neighbors (KNN), Naïve Bayes, ID3, C4.5, CART, dan lain-lain.

b. Asosiasi (*Association*)

Menemukan pola yang menggambarkan kekuatan hubungan fitur atau variabel dalam data. Pola yang ditemukan biasanya merepresentasikan bentuk aturan implikasi (subset fitur). Contoh algoritma *association* seperti algoritma FP-Growth, A Priori, Chi Square, Coefficient of Correlation, dan lain-lain.

c. Klasterisasi (*Clustering*)

Melakukan pengelompokan data-data ke dalam sejumlah kelompok (*cluster*) berdasarkan kesamaan karakteristik masing-masing data pada kelompok-kelompok yang ada. Data yang masuk dalam batas kesamaan dengan kelompoknya akan bergabung dalam kelompok tersebut, dan akan terpisah dengan kelompok yang berbeda jika diluar batas kesamaan kelompok tersebut. Contoh algoritma *clustering* seperti algoritma K-

Means, Fuzzy C-Means, K-Medoid, *Self-Organization Map* (SOM), dan lain-lain.

d. Prediksi (*Prediction*)

Model prediksi berkaitan dengan pembuatan model yang dapat melakukan pemetaan dari setiap himpunan variabel ke setiap targetnya, kemudian menggunakan model tersebut untuk memberikan nilai target pada himpunan baru. Contoh algoritma *prediction* seperti C4.5, *Linear Regression*, *Support Vector Machine*, *Neural Network*, dan lain-lain.

e. Estimasi (*Estimation*)

Teknik untuk melakukan estimasi terhadap sebuah data baru yang tidak memiliki keputusan berdasarkan histori data yang telah ada, dimana variabel target estimasi lebih ke arah numerik daripada ke arah kategori. Model dibangun menggunakan record lengkap yang menyediakan nilai dari variabel target sebagai nilai prediksi. Selanjutnya, pada peninjauan berikutnya estimasi nilai dari variabel target dibuat berdasarkan nilai dari variabel prediksi. Contoh algoritma *estimation* sama halnya dengan algoritma *prediction* yaitu *Linear Regression*, *Support Vector Machine*, *Neural Network*, dan lain-lain.

Metode atau algoritma data mining juga sangat bervariasi. Pemilihan metode atau algoritma tersebut bergantung pada tujuan dilakukannya data mining. *Data mining* biasanya mengolah data dari database dengan ukuran yang besar dan dari data tersebut dilakukan pencarian pola atau trend sesuai dengan tujuan dari penerapan data mining tersebut (Liliana et al., 2021). Hasil dari pengolahan *data mining* tersebut selanjutnya dapat digunakan untuk pengambilan keputusan prediksi yang dibutuhkan. Prediksi merupakan sebuah fungsi yang dapat menemukan pola tertentu dari suatu data. Pola-pola tersebut dapat diketahui dari berbagai variabel yang ada pada data. Ketika sudah menemukan pola, maka pola yang didapat tersebut bisa digunakan untuk memprediksi variabel lain yang belum diketahui nilai ataupun jenisnya.

2.2.5 Algoritma C4.5

Algoritma C4.5 merupakan salah satu jenis algoritma pohon keputusan (*decision tree*) yang digunakan dalam *data mining* dan *machine learning*. Struktur pohon keputusan mirip dengan flowchart, di mana setiap simpul internal melakukan pengujian pada atribut, setiap cabang merupakan hasil dari pengujian, dan setiap simpul daun menjadi label kelas. Simpul paling atas dalam pohon adalah simpul akar. Secara umum langkah-langkah algoritma C4.5 untuk membangun pohon keputusan adalah sebagai berikut; pilih atribut sebagai akar, buat cabang untuk tiap-tiap nilai, bagi kasus dalam cabang, dan ulangi proses untuk setiap cabang sampai semua kasus pada cabang memiliki kelas yang sama (Mardi, 2019).

Algoritma pohon keputusan termasuk *supervised learning*, yang berarti memerlukan variabel sasaran atau target sebelumnya sebagai *data training*. Dengan menggunakan algoritma C4.5, dapat mempelajari data dengan jumlah yang besar dan membangun model pembelajaran berupa pohon keputusan yang dapat digunakan untuk memprediksi data yang belum pernah muncul sebelumnya (Febriarini & Astuti, 2019). Pohon keputusan merupakan salah satu alternatif dalam pemecahan masalah yang diambil dari data tersebut. Sekumpulan *data training* harus dipersiapkan untuk membentuk algoritma dengan nilai-nilai variabel target. Dalam metode ini dievaluasi semua atribut menggunakan ukuran statistik seperti *information gain* dan perhitungan *entropy*. *Information gain* merupakan perolehan informasi atau ukuran efektivitas suatu atribut dalam mengklasifikasikan data. Persamaan (1) merupakan rumus *information gain* yaitu (Widaningsih, 2019).

$$Gain(S, A) = Entropy(S) - \sum \frac{|S_v|}{|S|} Entropy(S_v) \quad (1)$$

Keterangan :

A : atribut

|S_v| : jumlah sampel untuk nilai v

$|S|$: jumlah seluruh sampel data

Entropy adalah keberagaman suatu data. Pada persamaan (2) merupakan rumus entropi:

$$Entropy(S) = -\sum p_i \log_2 p_i \quad (2)$$

Keterangan:

p_i = porsi atau rasio antara jumlah sampel kelas i dengan jumlah semua sampel pada himpunan data

2.2.6 *Confusion Matrix*

Confusion matrix adalah alat yang sangat penting dalam bidang klasifikasi untuk mengevaluasi efisiensi prediksi dari model klasifikasi. *Confusion matrix* merupakan suatu tabel yang menggambarkan performa yang dicapai oleh pengklasifikasi pada himpunan data pengujian. Tabel ini terdiri dari baris dan kolom yang masing-masing mewakili kelas aktual dan prediksi dari instans data pengujian. Semua matrik evaluasi lain yang digunakan dalam bidang klasifikasi (*accuracy*, *precision*, dan *recall*) berasal dari *confusion matrix*s. *Confusion matrix* untuk masalah klasifikasi biner berisi empat nilai; *True Positive* (TP), *True Negative* (TN), *False Positive* (FP), dan *False Negative* (FN). TN dan TP menunjukkan frekuensi model yang benar dalam mengklasifikasikan instans data masing-masing sebagai negatif dan positif; sedangkan FN dan FP menunjukkan frekuensi model yang salah dalam mengklasifikasikan instans data sebagai negatif dan positif (Ryoba et al., 2020).

Matrik evaluasi yang berasal dari *confusion matrix* dan digunakan untuk menilai kinerja model meliputi *accuracy*, *recall*, dan *precision* (Ferdyandi et al., 2022). *Accuracy* adalah salah satu matrik yang paling banyak digunakan dalam menilai kinerja pengklasifikasi. Akurasi merupakan rasio dari total prediksi benar (TP dan TN) dengan jumlah total *data testing*. Persamaan untuk menghitung akurasi adalah:

$$Accuracy = \frac{TP+TN}{TP+TN+FP+FN} \quad (3)$$

Precision mengukur sejauh mana prediksi model yang diklasifikasikan sebagai positif adalah benar. Persamaan untuk menghitung precision adalah:

$$Precision = \frac{TP}{TP+FP} \quad (4)$$

Recall mengukur sejauh mana model mampu mengidentifikasi kasus positif dengan benar. Persamaan untuk menghitung recall adalah:

$$Recall = \frac{TP}{TP+FN} \quad (5)$$

Matrik-matrik ini sangat penting dalam mengevaluasi kinerja model klasifikasi, dan dapat membantu dalam memahami tingkat keakuratan dan keandalan model dalam memprediksi kelas tertentu. Dengan memahami *confusion matrix* dan matrik evaluasi yang berasal darinya, dapat mengoptimalkan model klasifikasi untuk mencapai hasil yang lebih baik dalam memprediksi kelas dari data yang belum dikenal.

2.2.7 Visual Studio Code

Visual Studio Code (VS Code) adalah sebuah teks editor ringan dan handal yang dikembangkan oleh Microsoft untuk sistem operasi multiplatform, yang berarti tersedia untuk versi Linux, Mac, dan Windows. Teks editor ini secara langsung mendukung bahasa pemrograman JavaScript, TypeScript, dan Node.js, serta bahasa pemrograman lainnya dengan bantuan plugin yang dapat dipasang melalui marketplace Visual Studio Code (seperti C++, C#, Python, Go, Java, dan lain-lain). Selain itu, VS Code juga bersifat *open source*, yang artinya kode sumbernya dapat dilihat dan kontribusi dapat diberikan untuk pengembangannya. Kode sumber dari VS Code dapat ditemukan di GitHub. Hal ini menjadikan VS Code sebagai pilihan favorit bagi para pengembang aplikasi, karena mereka dapat

berpartisipasi dalam proses pengembangan kedepannya. (Permana & Romadlon, 2019).

2.2.8 Streamlit

Deployment model ke dalam sistem seringkali memerlukan pengodean ulang model agar lebih cepat atau lebih kompatibel dengan lingkungan produksi yang sudah ada. Proses ini dapat memakan waktu dan tenaga jika tidak menggunakan alat yang tepat. Untuk mengatasi tantangan ini, diperlukan sebuah *framework* yang dapat mengubah skrip data menjadi aplikasi website dengan cepat dan mudah. Peran penting Streamlit, sebuah *framework open source* dari Python yang memungkinkan para profesional di bidang data untuk membuat aplikasi *website* menggunakan bahasa Python, baik untuk mengaplikasi model dari *machine learning* atau *data science* (Ferdyandi et al., 2022).

Streamlit menjadi pilihan ideal bagi para *data scientist* karena sejumlah fitur unggulannya. Salah satunya adalah kemampuan untuk *me-refresh* tampilan secara otomatis saat terjadi perubahan pada skrip. Hal ini memungkinkan pengguna untuk melihat perubahan dalam model secara langsung tanpa harus melakukan *refresh* manual. Dengan begitu, proses pengembangan dan iterasi model menjadi lebih cepat dan praktis. Selain itu, Streamlit juga menawarkan berbagai *widget* interaktif yang memudahkan pengguna dalam mengontrol dan memanipulasi data (Dani, 2022). Pengguna dapat dengan mudah menambahkan tombol, slider, dan elemen interaktif lainnya ke dalam aplikasi web mereka untuk meningkatkan interaktivitas dengan pengguna. Kemudahan penggunaan Streamlit juga menjadi daya tarik utama. Para *data scientist* dapat dengan cepat menguasai alat ini tanpa harus memiliki pengetahuan khusus dalam pengembangan web. Dengan bahasa Python sebagai bahasa utama, pengguna dapat fokus pada model tanpa harus terbebani dengan kompleksitas bahasa pemrograman lainnya.

Salah satu keunggulan Streamlit adalah opsi untuk melakukan *deploy* instan. Proses *deploy* menjadi lebih mudah dan cepat, sehingga proyek dapat langsung dipublikasikan dan dibagikan dengan mudah kepada orang lain. Ini membuat Streamlit sangat cocok untuk proyek kolaboratif dan proyek yang perlu

diakses secara *online*. Banyak *data scientist* terkemuka di dunia menggunakan Streamlit untuk proyek-proyek mereka. Selain itu, kompatibilitasnya dengan berbagai *library* Python lainnya membuatnya menjadi pilihan yang ideal untuk mengaplikasikan model *machine learning* atau *data science* dalam konteks yang lebih luas. Dengan semua kelebihan dan kemudahan yang ditawarkan oleh Streamlit, *framework* ini menjadi solusi efisien dan efektif bagi para profesional data yang ingin menciptakan aplikasi website yang interaktif, mudah diakses, dan dapat dengan cepat dipublikasikan untuk keperluan kolaboratif atau tujuan bisnis (Dani, 2022).

BAB III METODOLOGI PENELITIAN

Bab ini menjelaskan tentang metodologi yang digunakan dalam penelitian ini untuk mencapai tujuan penelitian. Metodologi penelitian meliputi kerangka berpikir, metode penelitian, dan *problem solving*. Penjelasan dalam bab ini akan memberikan pemahaman yang jelas tentang bagaimana penelitian ini akan dilakukan.

3.1. Kerangka Berpikir

Kerangka berpikir ini bertujuan untuk memberikan panduan dan struktur dalam mengimplementasi algoritma C4.5 sebagai faktor pendukung kesuksesan *fundraising digital*. Berikut adalah gambar kerangka pemikiran yang akan diimplementasikan pada penelitian ini:

Masalah		
Menerapkan Algoritma C4.5 pada sistem prediksi <i>fundraising digital</i> BAZNAS RI		
Eksperimen		
Tools	Data	Metode
Python Jupyter Notebook VS Code Streamlit	Data Muzaki	Algoritma C4.5 dan CRISP-DM
Pengujian dan Validasi Hasil		
Menguji metode dengan menggunakan <i>data testing</i> , serta menghitung <i>precision</i> dan <i>recall</i> dengan <i>confusion matrix</i> untuk mengetahui akurasi. Untuk sistem yang telah berhasil dibangun dilakukan tahap <i>deployment</i> ke Streamlit.		
Hasil		
Rekomendasi prediksi <i>fundraising digital</i> zakat dari muzaki		

Gambar 3. 1 Kerangka Berpikir

Berikut adalah penjelasan dari kerangka berpikir pada Gambar 3.1:

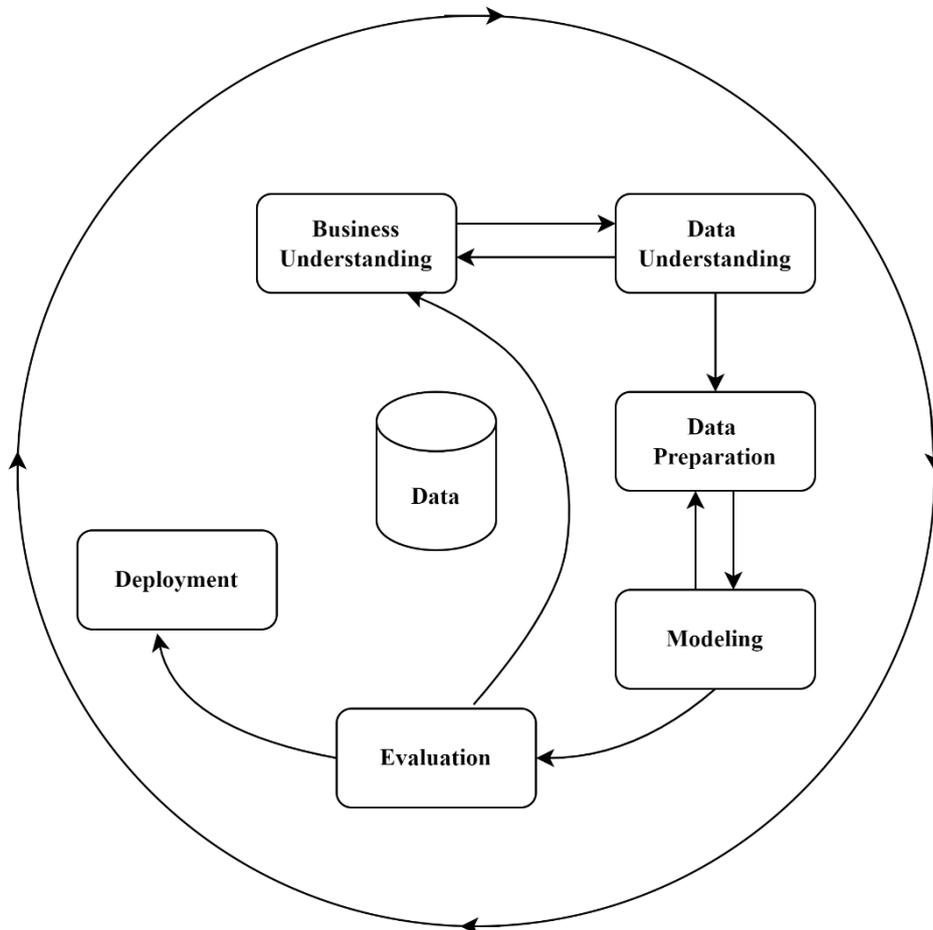
1. Tahap masalah mendefinisikan masalah yang ingin dipecahkan, yaitu membangun sistem prediksi *fundraising digital* BAZNAS menggunakan

algoritma C4.5. Tujuan dari penelitian adalah untuk memprediksi calon muzaki yang berminat berzakat melalui platform digital.

2. Tahap Eksperimen, melakukan pemilihan *tools* yang akan digunakan untuk penelitian. Tools yang dipilih adalah python, jupyter notebook, dan Streamlit. Python digunakan untuk pemrosesan data dan implementasi algoritma C4.5, sementara Jupyter Notebook digunakan untuk eksplorasi data. Selanjutnya, Streamlit digunakan untuk implementasi model prediksi C4.5 ke dalam bentuk aplikasi web yang interaktif. Data muzaki digunakan sebagai *data training* untuk melatih model C4.5 dan menghasilkan model prediksi. Metode CRISP-DM digunakan sebagai pendekatan untuk mengelola seluruh siklus penelitian data, termasuk tahap *data preparation, modeling, evaluation, dan deployment*.
3. Pengujian dan Validasi Hasil, pada tahap ini model prediksi C4.5 yang telah dilatih menggunakan data muzaki akan diuji menggunakan *data testing* yang terpisah. Pengujian ini bertujuan untuk mengevaluasi kinerja model pada data yang belum pernah dilihat sebelumnya dan memastikan bahwa model dapat digeneralisasi dengan baik. Pengujian model juga melibatkan perhitungan *precision* dan *recall* menggunakan *confusion matrix* untuk mengetahui akurasi model dalam memprediksi calon muzaki yang berminat berzakat. Setelah model dianggap berhasil, tahap selanjutnya adalah melakukan *deployment* model ke dalam aplikasi web menggunakan Streamlit. Bertujuan agar sistem prediksi *fundraising digital* BAZNAS dapat diakses oleh pengguna melalui antarmuka web yang sederhana dan ramah pengguna.
4. Hasil akhir dari penelitian ini adalah rekomendasi prediksi *fundraising digital* zakat dari calon muzaki berdasarkan model C4.5 yang telah berhasil dikembangkan. Model ini dapat memberikan informasi tentang kemungkinan calon muzaki berzakat melalui platform digital, sehingga BAZNAS dapat lebih efektif dalam mengajak masyarakat berpartisipasi dalam kegiatan zakat secara digital.

3.2. Metodologi Penelitian

Penelitian ini menerapkan metodologi CRISP-DM (*Cross-Industry Standard Process for Data Mining*) sebagai pendekatan penelitian yang digunakan. CRISP-DM merupakan suatu metode yang menyediakan standar proses pemecahan masalah secara umum dari unit penelitian yang ada. Proses data mining berdasarkan CRISP-DM terdiri dari 6 fase yaitu *Business Understanding* (pemahaman bisnis), *Data Understanding* (pemahaman data), *Data Preparation* (pengolahan data), *Modeling* (pemodelan), *Evaluation* (evaluasi), dan *Deployment* (penyebaran) (Hasanah et al., 2021). Proses metodologi ini terdiri dari 6 fase yang dapat dijelaskan sebagai berikut.



Gambar 3. 2 Metodologi CRISP-DM

3.1.1 *Business Understanding* (Pemahaman Bisnis)

Tahap awal yang penting untuk memahami sepenuhnya konteks bisnis yakni *fundraising* zakat yang bertujuan memaksimalkan pendistribusian dan pendayagunaan ZIS-DSKL untuk mengentaskan kemiskinan, meningkatkan kesejahteraan umat, dan mengurangi kesenjangan sosial. Hasil pemodelan akan digunakan untuk mengambil keputusan bisnis dalam melaksanakan strategi *fundraising* zakat yang efisien. Data yang diperlukan untuk mencapai tujuan penelitian adalah data muzaki dan tidak menggunakan data sensitif agar menjaga kerahasiaan data tersebut. Daftar pemangku kepentingan (*stakeholder*) yang terlibat dalam penelitian yakni:

- a. Direktorat Kajian dan Pengembangan ZIS-DSKL,
- b. Direktorat Layanan, Promosi dan Data Optimasi, dan
- c. Direktorat Keamanan Informasi, Data dan Layanan Digital.

3.1.2 *Data Understanding* (Pemahaman Data)

Tahap selanjutnya setelah *Business Understanding* dalam proyek data mining. Pada tahap ini, melakukan eksplorasi data menggunakan teknik EDA untuk memahami secara mendalam tentang dataset yang akan digunakan dalam penelitian. Data yang digunakan sebanyak 7896 *record* data primer yang didapatkan dari Direktorat Keamanan Informasi Data dan Layanan Digital (DKID). Tujuan dari *data understanding* adalah untuk mendapatkan gambaran tentang karakteristik data, kualitas data, serta mengidentifikasi masalah atau kekurangan yang perlu ditangani sebelum melanjutkan proses prediksi.

- a. Deskripsi Data

Penjelasan singkat tentang dataset yang digunakan, termasuk sumber data, jenis data atau struktur data, dan ukuran dataset.

- b. *Exploratory Data Analysis* (EDA)

Statistik deskriptif untuk setiap kolom dalam dataset, seperti mean, median, standar deviasi, minimum, dan maksimum. Visualisasi data seperti

histogram, scatter plot, atau box plot untuk memahami distribusi dan pola data.

c. Kualitas Data

Evaluasi kualitas data untuk mengidentifikasi nilai yang hilang (*missing values*), duplikat, dan *outliers*. Penanganan data yang hilang atau duplikat untuk memastikan kualitas dataset yang baik.

d. Korelasi Antara Fitur

Korelasi antara fitur (kolom) dalam dataset untuk memahami hubungan dan pengaruh antara fitur-fitur tersebut.

e. Visualisasi dan Segmentasi Data

Visualisasi yang lebih mendalam untuk mendapatkan wawasan tentang kaitan antara fitur-fitur tertentu. Segmentasi data berdasarkan karakteristik tertentu untuk memahami kelompok data yang berbeda dan perilaku mereka.

Dataset *fundraising* ZIS-DSKL Muzaki di BAZNAS Pusat dengan jumlah data sebanyak 7896. Penjelasan atribut dari dataset muzaki adalah sebagai berikut:

- a. Gender, atribut yang menggambarkan jenis kelamin seseorang memiliki dua nilai kategorikal, yaitu pria dan wanita. Atribut ini digunakan untuk mengidentifikasi perbedaan atau pola perilaku berdasarkan jenis kelamin dalam prediksi data.
- b. *Occupation*, atribut yang menunjukkan pekerjaan atau profesi seseorang. Atribut ini dapat memiliki berbagai nilai kategorikal, seperti Amil BAZNAS, Staff, PNS, Wiraswasta, tidak bekerja, dan lainnya. Informasi mengenai pekerjaan bisa memberikan wawasan tentang karakteristik, penghasilan, dan minat dari kelompok individu yang berbeda.
- c. Periode Transaksi, atribut mengacu pada waktu atau tanggal ketika suatu transaksi atau kejadian terjadi. Dalam penelitian ini dibagi menjadi tiga kategori yaitu Reguler, Ramadhan dan Kurban. Atribut ini penting dalam prediksi data untuk melihat tren, musiman, atau pola waktu lainnya yang dapat memberikan wawasan berharga.

- d. Nominal, merupakan atribut jumlah *fundraising* zakat yang dilakukan oleh muzaki. Atribut ini dikelompokkan menjadi dua kategori kecil untuk transaksi dibawah Rp.200.000 dan besar untuk transaksi diatas Rp.200.000. Hal ini dilakukan agar mempermudah dalam proses prediksi data.
- e. Umur, atribut yang mengelompokkan usia individu ke dalam kategori tertentu. Penggunaan kategori umur membantu dalam prediksi demografi dan pemahaman karakteristik kelompok usia yang berbeda. Kategori umur ini dibagi menjadi rentang usia seperti berikut.
 - 1. Masa balita usia 0 – 5 tahun
 - 2. Masa kanak-kanak usia 5 – 11 tahun
 - 3. Masa remaja awal usia 12 – 16 tahun
 - 4. Masa remaja akhir usia 17 – 25 tahun
 - 5. Masa dewasa awal usia 26 – 35 tahun
 - 6. Masa dewasa akhir usia 36 – 45 tahun
 - 7. Masa lansia awal usia 46 – 55 tahun
 - 8. Masa lansia akhir usia 56 – 65 tahun
 - 9. Masa manula usia 65 – ke atas

Sehingga dapat disimpulkan dari temuan utama tentang dataset yang sudah dieksplorasi serta dapat dilakukan identifikasi masalah atau kekurangan yang perlu diatasi dalam tahap *Data Preparation* selanjutnya.

3.1.3 *Data Preparation* (Persiapan Data)

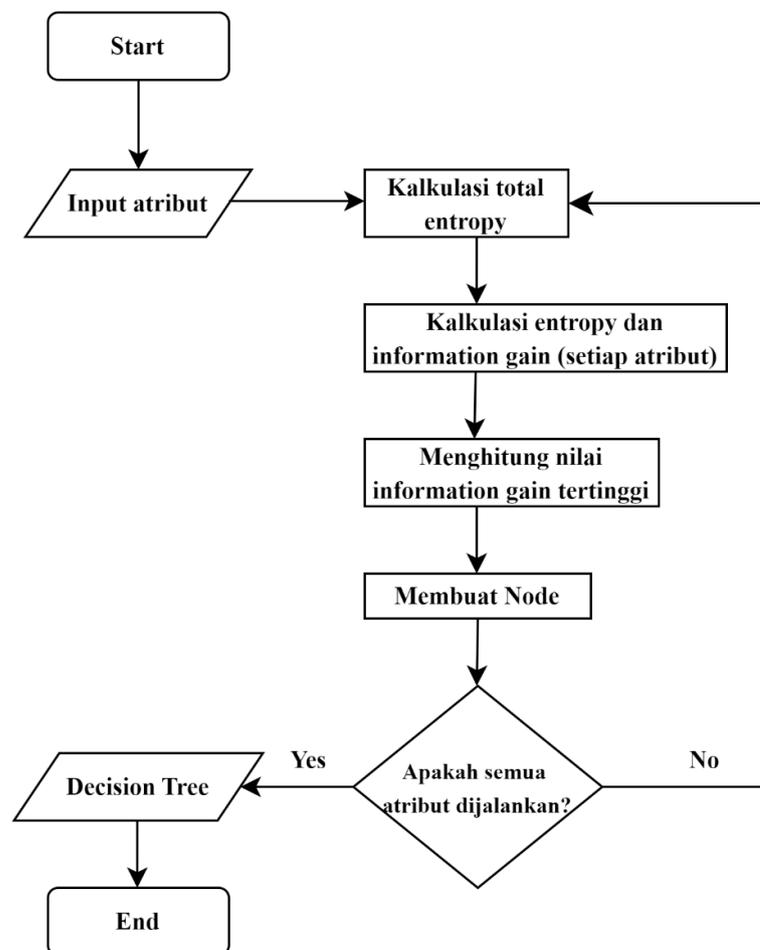
Dalam tahapan ini yaitu membangun dataset akhir dari data mentah menjadi data yang siap untuk diproses. Ada beberapa hal yang akan dilakukan sebagai berikut:

- a. *Data cleaning* (pembersihan data), identifikasi dan menangani *missing values* dalam dataset.
- b. *Data splitting* (pemisahan data), memisahkan data menjadi dua subset yaitu *data training* (data latih) dan *data testing* (data uji).

- c. *Data transformation* (Transformasi data), konversi data kategorikal menjadi bentuk yang sesuai untuk prediksi seperti *one-hot encoding* atau *label encoding* untuk dijadikan masukan dalam tahap pemodelan. Untuk klasterisasi terhadap kolom nominal digunakan *RobustScaler* , *KMeans*, dan *Elbow Method*.

3.1.4 Modeling (Pemodelan)

Pada tahap ini, membangun model prediksi berdasarkan data yang telah dipreparasi sebelumnya. Pengembangan algoritma C4.5 digunakan untuk mengambil keputusan atau membuat prediksi berdasarkan data yang ada. Flowchart dari algoritma C4.5 pada gambar dibawah ini:



Gambar 3. 3 Flowchart Algoritma C4.5

Setelah melakukan input dataset dilakukan kalkulasi *entropy* sesuai dengan rumus (2) pada Bab 2. Selanjutnya menggunakan teknik *random sampling* untuk menampilkan 10 data acak sebagai contoh perhitungan *entropy* dan *information gain*. Di bawah ini adalah tampilan dari 10 data acak yang diambil:

Tabel 3. 1 *Sampling Data*

No.	Gender	Occupation	Periode transaksi	Kategori nominal	Kategori umur
1	pria	Amil BAZNAS	Reguler	Kecil	Dewasa Akhir
2	pria	Staff	Reguler	Besar	Dewasa Awal
3	pria	Amil BAZNAS	Ramadhan	Besar	Dewasa Akhir
4	wanita	Amil BAZNAS	Reguler	Besar	Dewasa Awal
5	pria	Staff	Reguler	Besar	Dewasa Akhir
6	pria	Staff	Reguler	Besar	Dewasa Akhir
7	wanita	Staff	Kurban	Besar	Lansia Awal
8	pria	Staff	Ramadhan	Kecil	Dewasa Awal
9	pria	Staff	Reguler	Kecil	Dewasa Akhir
10	wanita	Staff	Ramadhan	Kecil	Lansia Awal

Dengan menggunakan *sampling* data di atas dilanjutkan dengan menghitung nilai *entropy* total, nilai *entropy* dari nilai setiap atribut kemudian dihitung nilai *information gain* dari setiap atribut. Nilai *entropy* total dan *entropy* dari nilai setiap atribut pada data Tabel 3.2 dihitung menggunakan persamaan (1) pada Bab 2.

Diketahui jumlah kasusnya adalah 10 kasus dengan kelas nominal besar sebanyak 7 dan nominal kecil sebanyak 3. Maka nilai *entropy* total dan *gain* umur adalah:

$$\begin{aligned}
 Entropy(S) &= -\sum p_i \log_2 p_i \\
 Entropy\ Total(S) &= \left(-\frac{S1}{S} \times \log_2 \left(\frac{S1}{S}\right)\right) + \left(-\frac{S2}{S} \times \log_2 \left(\frac{S2}{S}\right)\right) \\
 &= \left(-\frac{7}{10} \times \log_2 \left(\frac{7}{10}\right)\right) + \left(-\frac{3}{10} \times \log_2 \left(\frac{3}{10}\right)\right) \\
 &= 0.360201 + 0.52109 \\
 &= 0.8813
 \end{aligned}$$

$$Gain(\text{umur}) = Entropy(S) - \sum \frac{|Sv|}{|S|} Entropy(Sv)$$

$$\begin{aligned}
&= Entropy\ total - \left(\frac{S_a}{S_{total}} \times Entropy\ a \right) \\
&\quad - \left(\frac{S_b}{S_{total}} \times Entropy\ b \right) \\
&\quad - \left(\frac{S_c}{S_{total}} \times Entropy\ c \right) \\
&= 0.881291 - \left(\frac{3}{10} \times 0.138346 \right) - \left(\frac{5}{10} \times 0.970951 \right) - \left(\frac{2}{10} \times 1 \right) \\
&= 0.881291 - (0.041504) - (0.48548) - (0.2) \\
&= 0.1543
\end{aligned}$$

Keterangan perhitungan pada *entropy* total dan *gain* (umur):

- a = Dewasa Awal
- b = Dewasa Akhir
- c = Lansia Awal
- S1 = Nominal Besar
- S2 = Nominal Kecil
- S = Jumlah Kasus

Dilakukan perhitungan nilai *entropy* sehingga didapatkan nilai *gain* dari semua atribut sehingga dapat dilihat pada Tabel 3.2 berikut.

Tabel 3. 2 Contoh perhitungan *entropy* dan *information gain*

	Keterangan	Jumlah Kasus (S)	Nominal Besar (S1)	Nominal Kecil (S2)	Entropy	Gain
Total		10	7	3	0.8813	
Umur						0.1543
	Dewasa Awal	3	2	1	0.1383	
	Dewasa Akhir	5	3	2	0.9709	
	Lansia Awal	2	1	1	1	
Gender						-0.0839
	Pria	7	4	3	0.9852	
	Wanita	3	2	1	0.9183	
Occupation						-0.0839
	Amil BAZNAS	3	2	1	0.9183	
	Staff	7	4	3	0.9852	

Berdasarkan nilai *gain* yang didapatkan dari semua atribut, nilai *gain* tertinggi adalah atribut umur. Sehingga yang menjadi akar pertama *decision tree* adalah atribut umur. Selanjutnya membuat node dan menjalankan semua atribut

sehingga menghasilkan sebuah model pohon keputusan. Interpretasi hasil dari model prediksi untuk mendapatkan wawasan tentang faktor-faktor yang mempengaruhi hasil serta digunakan untuk mengambil keputusan bisnis atau *fundraising* zakat dan memberikan rekomendasi kepada BAZNAS RI.

3.1.5 Evaluation (Pengujian)

Tahap ini dilakukan dengan melihat tingkat performa dari model yang dihasilkan oleh algoritma. Evaluasi adalah proses kritis untuk memahami sejauh mana model dapat diandalkan dalam mengambil keputusan atau membuat prediksi yang akurat. Parameter yang digunakan untuk evaluasi komparasi algoritma adalah *Confusion Matrix* dengan aturan nilai akurasi, presisi dan recall. Hasil dari model ini merupakan hasil temuan dari pengujian data yang telah disiapkan untuk pembuatan model. Oleh karena itu, evaluasi terhadap model awal yang terbentuk menjadi penting untuk menguji kualitas dan efektivitas model sebelum memutuskan apakah model yang ditemukan layak untuk diimplementasikan (*deploy*). Identifikasi kesalahan yang dilakukan oleh model, seperti jenis kesalahan yang sering terjadi dan penyebab nya. Oleh karena itu, dilakukan upaya perbaikan yang dapat dilakukan berdasarkan identifikasi kesalahan tersebut. Selanjut nya pengujian stabilitas model untuk memastikan bahwa model memberikan hasil yang konsisten sehingga dapat diandalkan pada data yang berbeda atau data masa depan.

3.1.6 Deployment (Penyebaran)

Tahapan terakhir, model prediksi yang telah dikembangkan dan dievaluasi diterapkan dalam lingkungan produksi atau operasional. Model tersebut diimplementasikan dan diintegrasikan ke dalam sistem yang digunakan oleh BAZNAS RI untuk mengambil keputusan atau mendukung proses bisnis (*fundraising* zakat).

Dalam tahapan terbagi menjadi dua yaitu *deploy* secara sederhana dan *deploy* secara menyeluruh. *Deploy* secara menyeluruh merupakan suatu penelitian

intensif secara terus menerus dengan penggunaan data yang lebih besar dan proses *data mining* dilakukan secara paralel. Dalam penelitian ini dilakukan *deploy* secara terus menerus, karena akan dipergunakan oleh BAZNAS RI dalam melakukan prediksi *fundraising digital* dikemudian hari.

Integrasi model C4.5 kedalam aplikasi web menggunakan *framework* Streamlit. Langkah-langkah integrasi model C4.5 dilakukan sebagai berikut:

1. Model yang telah dihasilkan dalam bentuk *file* atau objek yang dapat digunakan untuk melakukan prediksi.
2. Membuat aplikasi streamlit yang akan menampilkan antarmuka web untuk menggunakan model C4.5.
3. Dalam aplikasi streamlit, muat model C4.5 yang telah dilatih sebelumnya ke dalam skrip Python. Model dimuat dari file atau objek yang telah disimpan selama proses pelatihan.
4. Buat antarmuka pengguna (*user interface*) menggunakan komponen-komponen yang disediakan oleh Streamlit, seperti tombol, input teks, dan grafik. Ini akan menjadi tampilan aplikasi web yang memungkinkan pengguna untuk memasukkan data atau memilih fitur untuk dilakukan prediksi.
5. Ketika pengguna memasukkan data atau memilih fitur, aplikasi akan memanfaatkan model C4.5 yang telah dimuat untuk melakukan prediksi berdasarkan input tersebut.
6. Aplikasi Streamlit akan menampilkan hasil prediksi dari model C4.5 kepada pengguna. Hasil ini bisa berupa klasifikasi atau keputusan berdasarkan fitur yang telah diberikan.
7. Menambahkan fungsionalitas tambahan seperti visualisasi data atau penjelasan tentang model yang digunakan. Hal ini akan mempermudah pengguna atau *stakeholder* untuk melihat visualisasi data hasil prediksi.

3.3 *Problem Solving*

Solusi yang ditawarkan dalam implementasi algoritma C4.5 berbasis prediksi sebagai faktor pendukung kesuksesan *fundraising* digital adalah sebagai berikut.

3.3.1 Pemahaman yang Lebih Mendalam

Dengan menggunakan algoritma C4.5 dapat memperoleh wawasan yang lebih mendalam tentang data dan faktor-faktor yang berkontribusi terhadap kesuksesan *fundraising digital*. Ini memungkinkan untuk lebih memahami perilaku donor dan target audiens, serta mengidentifikasi pola dan tren yang dapat digunakan untuk meningkatkan strategi *fundraising*.

3.3.2 Prediksi yang Akurat

Dengan mengimplementasikan algoritma C4.5 dapat membuat prediksi yang lebih akurat tentang hasil *fundraising digital* di masa depan. Algoritma ini mampu menghasilkan pohon keputusan yang mempertimbangkan berbagai faktor dan variabel yang relevan, sehingga dapat memberikan panduan yang lebih baik dalam mengambil keputusan strategis.

3.3.3 Segmentasi Audiens yang Tepat

Algoritma C4.5 dapat membantu dalam segmentasi audiens berdasarkan karakteristik dan preferensi mereka. Dengan demikian, dapat mengarahkan upaya *fundraising* secara lebih efektif dengan menyampaikan pesan yang relevan kepada masing-masing segmen audiens.

3.3.4 Optimasi Penggunaan Sumber Daya

Melalui implementasi algoritma C4.5 dapat mengoptimalkan penggunaan sumber daya yang tersedia, seperti waktu, energi, dan anggaran. Algoritma ini dapat membantu dalam mengidentifikasi faktor-faktor yang paling berpengaruh terhadap kesuksesan *fundraising*, sehingga dapat mengalokasikan sumber daya dengan lebih efisien.

3.3.5 Perbaikan Berkelanjutan

Dengan menggunakan algoritma C4.5 dapat terus mengembangkan dan meningkatkan model prediksi berdasarkan data yang baru terkumpul. Ini memungkinkan untuk selalu mengikuti tren dan perubahan dalam lingkungan *fundraising digital*, sehingga strategi tetap relevan dan efektif.

Dengan adanya metode algoritma C4.5 dapat menerapkan pendekatan yang lebih ilmiah dan berbasis data dalam *fundraising digital*. Hal ini memungkinkan BAZNAS untuk mengoptimalkan strategi *fundraising*, meningkatkan kesuksesan kampanye *fundraising*, dan mencapai tujuan dengan lebih efektif.

BAB IV

IMPLEMENTASI DAN EKSPERIMEN

Dalam bab ini menjelaskan hasil penelitian dan pengujian yang dilakukan terkait penggunaan algoritma C4.5 dalam prediksi hasil *fundraising* zakat. Membagi pembahasan menjadi tiga bagian utama yaitu persiapan data, implementasi model, dan hasil analisis model.

4.1 Persiapan Data

Dalam mempersiapkan data sebelum dilakukan proses modeling Algoritma C4.5 dilakukan pengumpulan data dan pengolahan data. Data didapatkan dari Direktorat Keamanan Informasi Data dan Layanan Digital (DKID) BAZNAS RI. Data primer tersebut dari data transaksi muzaki ditingkat pusat sebanyak 7896 *record*.

4.1.1 Pengumpulan Data

Mengumpulkan data menggunakan teknik pengumpulan data dokumentasi. Dalam melaksanakan pengumpulan data, peneliti memastikan bahwa data yang diperoleh adalah valid serta mendapatkan persetujuan dari *stakeholder* untuk menjaga kerahasiaan data, dan memastikan transparansi dalam proses pengumpulan data. Data dokumentasi (primer) ini dari data muzaki BAZNAS RI dilingkup pusat yang disimpan pada *Cloud Storage Google*. Terdapat total 7896 data transaksi muzaki yang diambil dari tahun 2022 dan 2023 yang disimpan dalam format CSV (*Comma Separated Values*).

4.1.2 Pengolahan Data

Data mentah sebelum dilakukan proses *cleaning* pada tabel 4.1 terdiri dari 7896 baris dan 9 kolom. Kolom terdiri dari tanggal transaksi, umur, gender,

occupation, *occupation group*, *npwz*, *name*, nominal, dan periode transaksi. Pada isi kolom *occupation*, *npwz*, dan *name* di sensor karena termasuk data muzaki yang sensitif.

Tabel 4. 1 Data Mentah

No.	tanggal transaksi	umur	gender	occupation	occupation group	npwz	name	nominal	periode transaksi
1	2022-12-23	47	pria	-	Amil	-	-	280833	Reguler
2	2022-12-23	45	pria	-	Amil	-	-	550000	Reguler
3	2022-12-23	36	wanita	-	Amil	-	-	242654	Reguler
4	2022-12-23	40	pria	-	Amil	-	-	172263	Reguler
5	2022-12-23	52	pria	-	Amil	-	-	189250	Reguler
6	2022-12-23	36	pria	-	Amil	-	-	587500	Reguler
7	2022-12-23	41	pria	Kesehatan	Kesehatan	-	-	500000	Reguler
8	2022-12-23	49	pria	-	Amil	-	-	875000	Reguler
9	2022-12-23	32	pria	-	Amil	-	-	190365	Reguler
10	2022-12-23	42	pria	-	Amil	-	-	375000	Reguler
...
7896	2022-07-27	51	pria	PHPU	Amil	-	-	30750	Reguler

a. *Data cleaning* (Pembersihan data).

Pada dataset terdapat data muzaki yang bersifat sensitif dan kolom yang tidak dipakai dalam proses *modeling*. Oleh karena itu dilakukan penghapusan kolom tanggal transaksi, *occupation*, *npwz*, dan *name*. Kolom tanggal transaksi dihapus karena sudah ada kolom periode transaksi, sehingga kolom tanggal transaksi tidak diperlukan. Sama halnya juga dengan kolom *occupation* tidak diperlukan, karena sudah ada kolom *occupation group*. Kemudian pada kolom *npwz* dan *name* dihapus karena termasuk data sensitif dan sangat privasi, serta pada kolom *occupation* juga terdapat nomor induk amil.

Selanjutnya menyimpan kembali data yang sudah dihapus kolomnya ke dalam file CSV baru. Pembacaan pada data setelah penghapusan kolom dapat dilihat terdapat 7896 *record* dan 5 kolom data.

Tabel 4. 2 Data setelah *cleaning*

No.	umur	gender	occupation group	nominal	periode transaksi
1	32	pria	Lainnya	300000	Reguler
2	33	wanita	Amil	50000	Reguler
3	48	wanita	PNS	212500	Reguler
4	37	wanita	PNS	1000000	Reguler
5	48	wanita	PNS	212500	Reguler
6	48	wanita	PNS	212500	Reguler
7	34	wanita	Konsultan	360000	Reguler
8	36	pria	Amil	152887	Reguler
9	30	wanita	PNS	436828	Reguler
10	30	wanita	PNS	229179	Reguler
...
7896	51	pria	Lainnya	30750	Reguler

b. *Data transformation* (Transformasi data)

Pada tahap ini melakukan serangkaian pengolahan data yaitu melakukan teknik *Exploratory Data Analysis* (EDA) dan melakukan klusterisasi menggunakan metode Elbow.

Pertama menggunakan fungsi `df.info()` untuk melihat struktur dan karakteristik data dalam DataFrame. Informasi yang didapat tidak ada nilai null, tipe data kolom interger dan *object*, serta banyak entri. Kemudian, menggunakan `LabelEncoder` untuk mengkodekan kolom gender dan periode transaksi dalam DataFrame. Hal ini akan menggantikan nilai-nilai dalam kolom tersebut dengan kode numerik yang sesuai. Pada kolom gender angka 0 berarti pria dan angka 1 berarti wanita. Pada kolom periode transaksi angka 0 berarti Kurban, angka 1 berarti Ramadhan, dan angka 2 berarti Reguler.

```
from sklearn.preprocessing import LabelEncoder

encoder = LabelEncoder()
df['gender'] = encoder.fit_transform(df['gender'])
df['periode_transaksi'] = encoder.fit_transform(df['periode_transaksi'])
print(df)
```

Gambar 4. 1 *Source code encoder*

Setelah itu, dilakukan penskalaan data menggunakan *RobustScaler* pada kolom nominal dalam Dataframe. Penskalaan pada kolom nominal bertujuan untuk menormalkan atau menstandarisasi rentang nilai dalam kolom tersebut.

```
scaler = RobustScaler().fit(df_cluster["nominal"].to_numpy().reshape(-1, 1))
nominal = scaler.transform(df_cluster["nominal"].to_numpy().reshape(-1, 1))
df_cluster = pd.DataFrame(nominal.ravel(), columns=["nominal"])
df_cluster
```

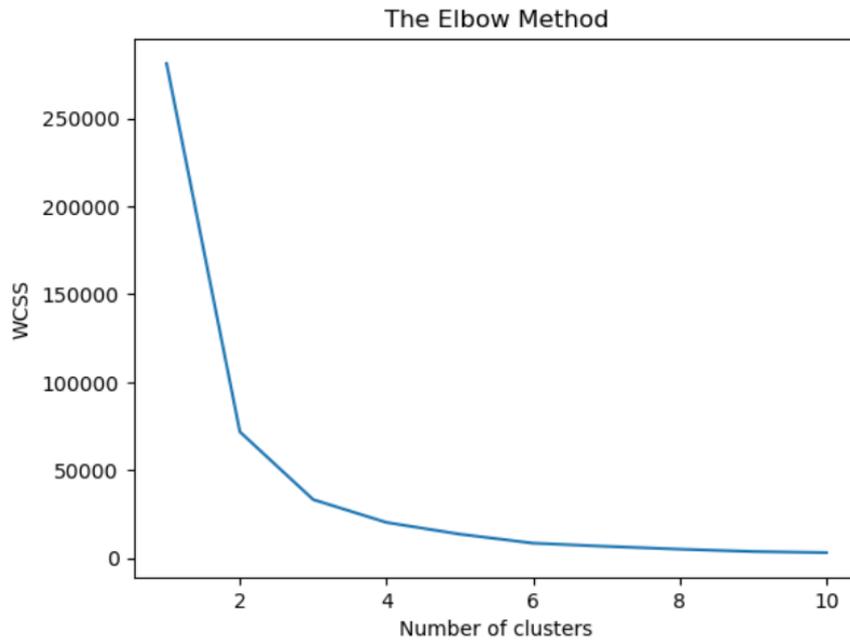
Gambar 4. 2 *Source code* normalisasi kolom nominal

Selanjutnya mengelompokkan nilai yang ada pada kolom nominal menjadi beberapa kategori nominal agar lebih mudah dalam melakukan modeling. Pengelompokan kolom nominal menggunakan metode Elbow dan Algoritma KMeans. Dapat dilihat *source code* yang digunakan pada gambar 4.2.

```
wcss = []
for i in range(1, 11):
    kmeans = KMeans(n_clusters = i, init = 'k-means++', random_state = 42)
    kmeans.fit(df_cluster)
    wcss.append(kmeans.inertia_)
plt.plot(range(1, 11), wcss)
plt.title('The Elbow Method')
plt.xlabel('Number of clusters')
plt.ylabel('WCSS')
plt.show()
```

Gambar 4. 3 *Source code* metode elbow

Hal pertama membuat sebuah *list* kosong 'wcss' untuk menyimpan nilai *Within Cluster Sum of Squares* (WCSS) dari setiap percobaan klasterisasi. Kemudian melakukan iterasi dari 1 hingga 10 yang mewakili jumlah klaster yang mungkin untuk dicoba. Selanjutnya membuat objek Kmeans dengan jumlah klaster yang berbeda dan melatih model Kmeans pada data klaster. Fungsi 'wcss.append(kmeans.inertia_)' digunakan untuk menambahkan nilai WCSS dari hasil klasterisasi ke dalam *list* 'wcss'. Sehingga dapat diketahui hasil nilai klasterisasi dari kolom nominal pada gambar 4.3 dibawah ini.



Gambar 4. 4 Visualisasi metode elbow

Kemudian menggunakan *Silhouette Score* untuk mengevaluasi kualitas kluster yang dihasilkan oleh Kmeans. Nilai *Silhouette Score* yang didapatkan yaitu 83% yang artinya hasil klasterisasi didapatkan sebanyak 4 kategori nominal sudah baik. Kategori nominal terdiri dari angka 0 berarti nominal rendah, 3 berarti nominal sedang, 2 berarti nominal diatas rata-rata, dan 1 berarti nominal tinggi. Berikut ditampilkan beberapa baris data dari DataFrame yang telah diperbarui.

Tabel 4. 3 Data setelah *transformation*

No	umur	gender	occupation group	periode transaksi	kategori nominal
1	32	0	10	2	1
2	33	1	1	2	1
3	48	1	12	2	1
4	37	1	12	2	1
5	48	1	12	2	1
6	48	1	12	2	1
7	34	1	9	2	1
8	36	0	1	2	1
9	30	1	12	2	1
10	30	1	12	2	1
...
7896	51	0	10	2	1

Dilanjutkan dengan mengubah nilai-nilai dalam DataFrame yang memiliki tipe data *object* atau kategorikal seperti kolom *occupation group* dan kategori nominal menjadi nilai numerik menggunakan LabelEncoder.

c. *Data Splitting* (Pemisahan data)

Setelah itu data dibagi menjadi X sebagai fitur dan y sebagai target. X terdiri dari kolom umur, gender, *occupation group*, dan periode transaksi. Sedangkan y yaitu kolom kategori nominal. Data ini dibagi menjadi *data training* (data latih) sebanyak 80% dengan jumlah 6316 record dan *data testing* (data uji) sebanyak 20% dengan jumlah 1580 record dengan total keseluruhan dataset 7896 *record* data. Berikut ditampilkan beberapa baris *data training* yang digunakan untuk melatih model C4.5.

Tabel 4. 4 *Data Training*

No	umur	gender	occupation group	periode transaksi	kategori nominal
1	35	0	7	2	1
2	31	0	9	2	1
3	32	0	1	2	2
4	47	0	4	2	1
5	33	0	6	2	1
6	28	0	6	2	1
7	39	0	6	2	1
8	60	0	8	2	0
9	39	0	1	2	1
10	31	0	1	0	1
...
6316	50	1	12	2	1

Berikut ini juga ditampilkan beberapa baris *data testing* pada tabel 4.5 untuk menguji hasil dari model C4.5.

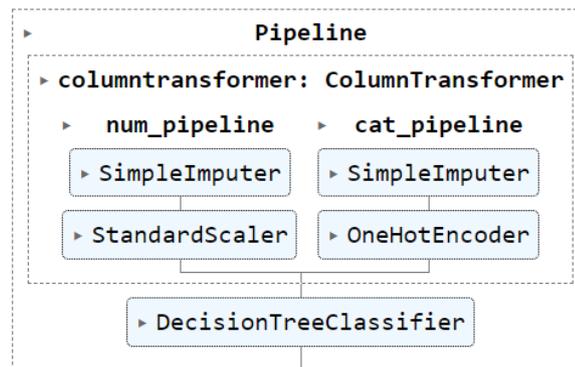
Tabel 4. 5 *Data Testing*

No	umur	gender	occupation group	periode transaksi	kategori nominal
6317	42	0	6	2	1
6318	42	0	19	0	1
6319	40	1	19	1	2
6320	40	1	17	2	1
6321	38	0	19	2	1
6322	32	1	14	2	1
6323	30	0	16	1	1
6324	28	0	14	2	1
6325	34	0	14	2	1
6326	42	0	17	2	1

...
7896	40	1	9	2	1

d. Pipeline

Sebelum mengimplementasikan algoritma C4.5 dilakukan pembuatan *pipeline machine learning* yang bertujuan agar dapat membantu dalam mengorganisir alur kerja pemrosesan data dan model sehingga memudahkan dalam pengembangan dan penyesuaian model.



Gambar 4. 5 Pipeline machine learning algoritma C4.5

Pipeline ini akan memroses semua fitur dalam dataset menggunakan *ColumnTransformer* yang telah didefinisikan sebelumnya. *ColumnTransformer* akan memutuskan bagaimana setiap fitur harus diubah berdasarkan *pipeline* yang telah ditentukan untuk fitur numerik dan kategorikal. Setelah pemrosesan fitur, data akan diberikan ke model *DecisionTreeClassifier* untuk pelatihan dan prediksi. *Pipeline* disimpan dan di-load kembali menggunakan *joblib*. Dari pembuatan *pipeline machine learning* didapat skor akurasi 0.92, yang artinya model memiliki kinerja yang baik.

4.2 Implementasi Model

Dilakukan pembangunan model prediksi berdasarkan data yang telah disiapkan sebelumnya. Pertama, membuat dan melatih model *Decision Tree Classifier* dengan menggunakan kriteria pemisahan *entropy*.

```
def calculate_entropy(series):
    counts = series.value_counts()
    probabilities = counts / counts.sum()
    entropy = -np.sum(probabilities * np.log2(probabilities))
    return entropy
```

Gambar 4. 6 *Source code* perhitungan *entropy* setiap variabel

Dilakukan penghitungan *entropy* dari setiap variabel dalam dataset dimulai dari menghitung jumlah total sampel dalam data menggunakan fungsi ‘len’, kemudian menghitung jumlah kemunculan setiap nilai dalam variabel kategorikal menggunakan metode ‘value_counts’. Selanjutnya menghitung probabilitas kemunculan setiap nilai dengan membagi jumlah kemunculan setiap nilai dengan jumlah total sampel. *Probabilities* adalah porsi atau rasio antara jumlah sampel kelas *i* dengan jumlah semua sampel pada himpunan data. Kemudian menghitung *entropy* dari setiap variabel dalam dataset digunakan rumus pada persamaan (2) yang telah dijelaskan dalam Bab 2. Didapatkan hasil penghitungan *entropy* setiap variabel seperti tabel 4.6 berikut ini. Nilai ini menunjukkan bahwa variabel dengan variasi yang lebih besar yaitu umur dan occupation group memiliki nilai *entropy* yang lebih tinggi.

Tabel 4. 6 Nilai *entropy* setiap variabel

No	Variabel	Nilai Entropy
1	Umur	4.90513981626614
2	Gender	0.7960833381430548
3	Occupation group	3.4041416302684167
4	Periode transaksi	1.2569797128181075

Selanjutnya menghitung *weighted entropy* agar dapat menghitung nilai *information gain*, sehingga didapatkan variabel dengan *information gain* tertinggi yang dijadikan sebagai simpul akar. Hal pertama yang dilakukan mengekstrak nilai unik dari kolom pada DataFrame dan menginisialisasi ‘weighted_entropy_sum’ untuk menyimpan jumlah *entropy* terbobot. Kemudian iterasi melalui nilai-nilai unik dan menghitung *entropy* subset menggunakan fungsi ‘calculate_entropy’. Lalu pembobotan dan akumulasi dengan mengalikan hasil dari ‘calculate_entropy’ dengan proporsi subset terhadap keseluruhan dataset (jumlah sampel dalam subset dibagi total sampel) dan ditambahkan ke

'weighted_entropy_sum'. Kemudian mengembalikan nilai total dari 'weighted_entropy_sum'.

```
def weighted_entropy(df, column, target):
    unique_values = df[column].unique()
    weighted_entropy_sum = 0
    total_samples = len(df)

    for value in unique_values:
        subset = df[df[column] == value]
        subset_entropy = calculate_entropy(subset[target])
        weighted_entropy_sum += (len(subset) / total_samples) * subset_entropy

    return weighted_entropy_sum
```

Gambar 4. 7 Source code perhitungan *weighted entropy*

Kemudian dilakukan perhitungan *information gain* dari setiap variabel dalam dataset menggunakan rumus pada persamaan (1) yang telah dijelaskan dalam Bab 2. Hal pertama yang dilakukan menghitung total *entropy* dari seluruh kolom target dan *weighted entropy*. Selanjutnya mengiterasi setiap kolom dan kode menghitung *entropy* dari kolom tersebut dan *information gain* berdasarkan umur.

```
def information_gain(df, column, target):
    total_entropy = calculate_entropy(df[target])
    w_entropy = weighted_entropy(df, column, target)
    return total_entropy - w_entropy

# Hitung entropy untuk setiap variabel kategorikal dan Information Gain
columns_to_analyze = ['umur', 'gender', 'occupation_group',
                     'periode_transaksi']

for column in columns_to_analyze:
    entropy_value = calculate_entropy(df[column])
    ig = information_gain(df, column, 'umur')
    print(f"Entropy for {column}: {entropy_value}")
    print(f"Information Gain for {column}: {ig}")
```

Gambar 4. 8 Source code perhitungan *information gain* setiap variabel

Didapatkan hasil perhitungan *information gain* setiap variabel seperti tabel 4.7 berikut ini. Dapat terlihat bahwa gain tertinggi oleh variabel umur yang akan dijadikan sebagai simpul akar.

Tabel 4. 7 Nilai *information gain* setiap variabel

No	Variabel	Nilai Information Gain
1	Umur	4.905139816266144
2	Gender	0.10551167788908788
3	Occupation group	0.8829641831609703
4	Periode transaksi	0.033769493514653526

Dataset akhir yang digunakan dalam eksperimen ini untuk membangun model Algoritma C4.5 terdiri dari informasi tentang muzaki yang mencakup berbagai variabel seperti umur, gender, *occupation group*, periode transaksi, dan kategori nominal. Berikut adalah penjelasan untuk setiap variabel.

Tabel 4. 8 Dataset akhir untuk membangun model

No.	Variabel	Tipe Data	Keterangan
1	Umur	Numerik	Usia muzaki dalam tahun
2	Gender	Kategorik	Jenis kelamin muzaki
3	<i>Occupation Group</i>	Kategorik	Jenis pekerjaan muzaki
4	Periode Transaksi	Kategorik	Waktu transaksi
5	Kategori Nominal	Kategorik	Kategori nominal zakat

1. Umur: Variabel numerik yang menggambarkan usia muzaki dalam tahun. Data ini memberikan informasi tentang sebaran usia muzaki yang berpartisipasi dalam proses *fundraising* zakat.
2. Gender: Variabel kategorikal yang menunjukkan jenis kelamin muzaki yaitu pria dan wanita. Informasi ini dapat membantu dalam memahami preferensi dan perilaku muzaki.
3. *Occupation Group*: Variabel kategorikal yang mengidentifikasi jenis pekerjaan muzaki. Data ini membantu dalam melihat pola donasi berdasarkan profesi atau bidang pekerjaan muzaki. Seperti Amil, PNS, Teknisi, TNI/POLRI, dan lain-lain. Dalam dataset yang sudah jadi terdapat 20 jenis pekerjaan dari para muzaki.
4. Periode Transaksi: Variabel kategorikal yang menunjukkan waktu transaksi dalam kategori tertentu. Periode transaksi dibagi menjadi tiga yaitu, Reguler, Ramadhan, dan Kurban. Hal ini dapat memberikan wawasan tentang tren dan pola donasi zakat dari waktu ke waktu.
5. Kategori Nominal: Variabel kategorikal yang mengelompokkan nominal zakat ke dalam empat kategori, yakni kategori nominal tinggi, nominal

diatas rata-rata, nominal sedang, dan nominal rendah. Ini membantu dalam memahami sebaran nominal donasi dan kecenderungan dalam jumlah zakat yang diberikan.

Dengan pemahaman tentang variabel-variabel tersebut, analisis prediksi dapat dilakukan untuk memahami faktor-faktor yang mempengaruhi hasil *fundraising* zakat.

4.3 Hasil dan Analisis

Setelah mengimplementasikan model Algoritma C4.5, dilanjutkan mengevaluasi kinerja model menampilkan hasil dari eksperimen yang dilakukan pada dua jenis data, yaitu data kelompok dan data tunggal. Eksperimen pada data kelompok bertujuan untuk mengidentifikasi pola atau tren secara keseluruhan di antara kelompok data yang lebih besar, sementara eksperimen pada data tunggal berfokus pada prediksi yang lebih mendetail terhadap individu dalam kelompok tersebut.

4.3.1 Evaluasi Kinerja Model

Dalam bagian ini dilakukan evaluasi dan validasi hasil dari model prediksi yang telah dibangun menggunakan algoritma C4.5. Evaluasi dilakukan dengan memeriksa seberapa baik model dapat memprediksi hasil *fundraising* zakat berdasarkan variabel-variabel yang telah ditentukan.

Melakukan prediksi menggunakan model C4.5 pada data pengujian ('X_test'). Hasil prediksinya akan disimpan dalam variabel 'y_pred'. Kemudian menghitung *accuracy*, *precision*, dan *recall* sesuai dengan persamaan (3), (4), dan (5) pada Bab 2 yang sudah dijelaskan. Untuk *source code* dapat dilihat pada gambar 4.9 berikut.

```

y_pred = model_c45.predict(X_test)
accuracy = accuracy_score(y_test, y_pred)
precision = precision_score(y_test, y_pred, average="weighted")
recall = recall_score(y_test, y_pred, average="weighted")
conf_matrix = confusion_matrix(y_test, y_pred)

print("Akurasi:", accuracy)
print("Precision:", precision)
print("Recall:", recall)
print("Confusion Matrix:")
print(conf_matrix)

```

Gambar 4. 9 Source code confusion matrix

Hasil dari evaluasi model menggunakan *confusion matrix* dapat dilihat pada tabel 4.9. Sebelumnya hasil klasterisasi didapatkan sebanyak 4 kategori nominal yang terdiri dari kelas 0 berarti nominal rendah, 3 berarti nominal sedang, 2 berarti nominal diatas rata-rata, dan 1 berarti nominal tinggi.

Tabel 4. 9 Tabel *confusion matrix*

	Prediksi Kelas 0	Prediksi Kelas 1	Prediksi Kelas 2	Prediksi Kelas 3
Aktual Kelas 0	4	7	3	4
Aktual Kelas 1	3	1415	32	0
Aktual Kelas 2	2	69	22	3
Aktual Kelas 3	2	5	0	9

Kelas aktual diperoleh dari *data training* dan digunakan untuk melatih model atau juga disebut sebagai kelas sebenarnya, sedangkan kelas prediksi dihasilkan oleh model ketika diberikan input baru di luar *data testing*. Akurasi yang dihasilkan pada perhitungan menggunakan *source code* pemrograman di *jupyter notebook* dihasilkan adalah:

a. *True Positives* (TP) = $4 + 1415 + 22 + 9 = 1450$

b. *False Positives* (FP)

FP untuk Kelas 0 = $3 + 2 + 2 = 7$

FP untuk Kelas 1 = $7 + 69 + 5 = 81$

FP untuk Kelas 2 = $3 + 32 + 0 = 35$

FP untuk Kelas 3 = $4 + 0 + 3 = 7$

Total FP = $7 + 81 + 35 + 7 = 130$

c. *False Negatives* (FN)

$$\text{FN untuk Kelas 0} = 7 + 3 + 4 = 14$$

$$\text{FN untuk Kelas 1} = 3 + 32 + 0 = 35$$

$$\text{FN untuk Kelas 2} = 2 + 69 + 3 = 74$$

$$\text{FN untuk Kelas 3} = 2 + 5 + 0 = 7$$

$$\text{Total FN} = 14 + 35 + 74 + 7 = 130$$

d. *True Negatives* (TN)

$$\text{TN untuk Kelas 0} = 1415 + 32 + 0 + 69 + 22 + 3 + 5 + 0 + 9 = 1555$$

$$\text{TN untuk Kelas 1} = 4 + 3 + 4 + 2 + 22 + 3 + 2 + 0 + 9 = 49$$

$$\text{TN untuk Kelas 2} = 4 + 7 + 4 + 3 + 1415 + 0 + 2 + 5 + 9 = 1449$$

$$\text{TN untuk Kelas 3} = 4 + 7 + 3 + 3 + 1415 + 32 + 2 + 69 + 22 = 1557$$

Untuk menghitung total TN secara praktis dari total nilai dalam matrix yaitu 1580 yang kemudian kurangi jumlah semua TP, FP, dan FN. Sehingga total TN adalah 840. Selanjutnya mencari nilai *accuracy*, *precision*, dan *recall* dari model yang telah dibangun.

$$\begin{aligned} \text{Accuracy} &= \frac{TP + TN}{TP + TN + FP + FN} \\ &= \frac{1450 + 840}{1450 + 840 + 130 + 130} \\ &= \frac{2290}{2550} \\ &= 0.898 \end{aligned}$$

$$\begin{aligned} \text{Precision} &= \frac{TP}{TP + FP} \\ &= \frac{1450}{1450 + 130} \\ &= \frac{1450}{1580} \\ &= 0.917 \end{aligned}$$

$$\begin{aligned}
Recall &= \frac{TP}{TP + FN} \\
&= \frac{1450}{1450 + 130} \\
&= \frac{1450}{1580} \\
&= 0.917
\end{aligned}$$

Hasil evaluasi model menunjukkan performa yang cukup baik dalam memprediksi kelas-kelas yang berbeda. Pertama, akurasi model mencapai sekitar 89,8%, menunjukkan bahwa sebagian besar prediksi yang dibuat oleh model sesuai dengan data aktual. *Precision* sekitar 91.7% mengindikasikan bahwa sebagian kecil dari prediksi positif yang dibuat oleh model mungkin merupakan kesalahan. *Recall* memiliki nilai sekitar 91.7%, menunjukkan bahwa model mampu mengidentifikasi sebagian besar dari label positif yang ada. Namun, evaluasi lebih lanjut dengan menggunakan *confusion matrix* menunjukkan adanya beberapa kesalahan dalam memprediksi kelas, terutama pada kelas-kelas dengan jumlah sampel yang lebih rendah.

Tabel 4. 10 *Classification report*

	precision	recall	F1-score	support
0	0.36	0.22	0.28	18
1	0.95	0.98	0.96	1450
2	0.39	0.23	0.29	96
3	0.56	0.56	0.56	16
Accuracy			0.92	1580
Macro average	0.56	0.50	0.52	1580
Weighted average	0.90	0.92	0.91	1580

Berdasarkan hasil validasi, performa model dalam mengklasifikasikan data ke dalam empat kelas yang berbeda dapat dilihat dari berbagai metrik evaluasi. Untuk kelas 0, nilai *precision* tercatat sebesar 0.36, sedangkan *recall*-nya adalah 0.22, dan *F1-score* mencapai 0.28. Untuk kelas 1, *precision* yang tinggi sebesar 0.95 menunjukkan tingkat keakuratan prediksi yang tinggi, didukung oleh *recall* sebesar 0.98 dan *F1-score* sebesar 0.96. Kelas 2 menunjukkan *precision* 0.39,

recall 0.23, dan *F1-score* 0.29. Sementara itu, untuk kelas 3, *precision* mencapai 0.56, *recall* juga 0.56, dan *F1-score* 0.56. Akurasi keseluruhan model sekitar 92%. Dengan menggunakan rata-rata dari matriks evaluasi, seperti *macro average* sekitar 0.52 dan *weighted average* sekitar 0.91, dapat memperoleh gambaran tentang kinerja keseluruhan model dalam memprediksi kelas-kelas yang berbeda.

4.3.2 Hasil Eksperimen

Pada bagian ini akan ditampilkan hasil dari eksperimen yang dilakukan pada dua jenis data, yaitu data kelompok dan data tunggal. Eksperimen pada data kelompok bertujuan untuk mengidentifikasi pola atau tren secara keseluruhan di antara kelompok data yang lebih besar, sementara eksperimen pada data tunggal berfokus pada prediksi yang lebih mendetail terhadap individu dalam kelompok tersebut.

a. Eksperimen Data Kelompok

Sebelum melakukan eksperimen pada data kelompok, sebuah file input data dalam format CSV yang terdiri dari 1580 *record* data disiapkan untuk prediksi. Berikut adalah beberapa baris dari data mentah yang ditampilkan dalam tabel 4.11 di bawah ini.

Tabel 4. 11 Data uji coba prediksi

No	umur	gender	occupation group	periode transaksi
1	42	pria	IT	Reguler
2	42	pria	Wiraswasta	Kurban
3	40	wanita	Wiraswasta	Ramadhan
4	40	wanita	Tidak Bekerja	Reguler
5	38	pria	Wiraswasta	Reguler
6	32	wanita	Staff	Reguler
7	30	pria	Teknisi	Ramadhan
8	28	pria	Staff	Reguler
9	34	pria	Staff	Reguler
10	42	pria	Tidak Bekerja	Reguler

1580	40	wanita	Konsultan	Reguler

Pada data uji belum terdapat kolom *predictions*, hanya terdapat kolom umur, gender, occupation group dan periode transaksi. Untuk perbandingan hasil

setelah dilakukannya prediksi, ditampilkan beberapa *record* data pada tabel 4.12 dibawah ini.

Tabel 4. 12 Hasil data setelah dilakukan prediksi

No	umur	gender	occupation group	periode transaksi	predictions
1	42	pria	IT	Reguler	Nominal Tinggi
2	42	pria	Wiraswasta	Kurban	Nominal Tinggi
3	40	wanita	Wiraswasta	Ramadhan	Nominal Diatas Rata-rata
4	40	wanita	Tidak Bekerja	Reguler	Nominal Tinggi
5	38	pria	Wiraswasta	Reguler	Nominal Tinggi
6	32	wanita	Staff	Reguler	Nominal Tinggi
7	30	pria	Teknisi	Ramadhan	Nominal Tinggi
8	28	pria	Staff	Reguler	Nominal Tinggi
9	34	pria	Staff	Reguler	Nominal Tinggi
10	42	pria	Tidak Bekerja	Reguler	Nominal Tinggi
...	Nominal Tinggi
1580	40	wanita	Konsultan	Reguler	Nominal Tinggi

Dalam eksperimen menggunakan data kelompok dapat diakses melalui tautan <https://prediksifundraisingdigital.streamlit.app/>. Setelah mengakses tautan tersebut, pengguna dapat melakukan prediksi dengan mengunggah file inputan. Hasil dari prediksi tersebut akan ditampilkan bersama dengan visualisasi yang mendukung interpretasi hasil. Untuk memberikan gambaran lebih jelas, penulis juga menyertakan tampilan dari aplikasi web streamlit pada gambar 4.10 dan 4.11 berikut ini.

Informasi Tambahan

Aplikasi ini memprediksi hasil fundraising menggunakan model yang telah dilatih dan berdasarkan input yang Anda berikan. Model yang digunakan dalam aplikasi ini adalah model yang telah di-train sebelumnya menggunakan data muzaki.

Hubungi kami di yuliana111099@gmail.com untuk informasi lebih lanjut.

[Kunjungi situs web BAZNAS RI](#)

Jika ingin melakukan prediksi terhadap satu data, klik link berikut.

[Prediksi Satu Data](#)

BAZNAS RI Digital Fundraising Prediction Application

Upload a CSV file

Drag and drop file here
Limit 200MB per file • CSV

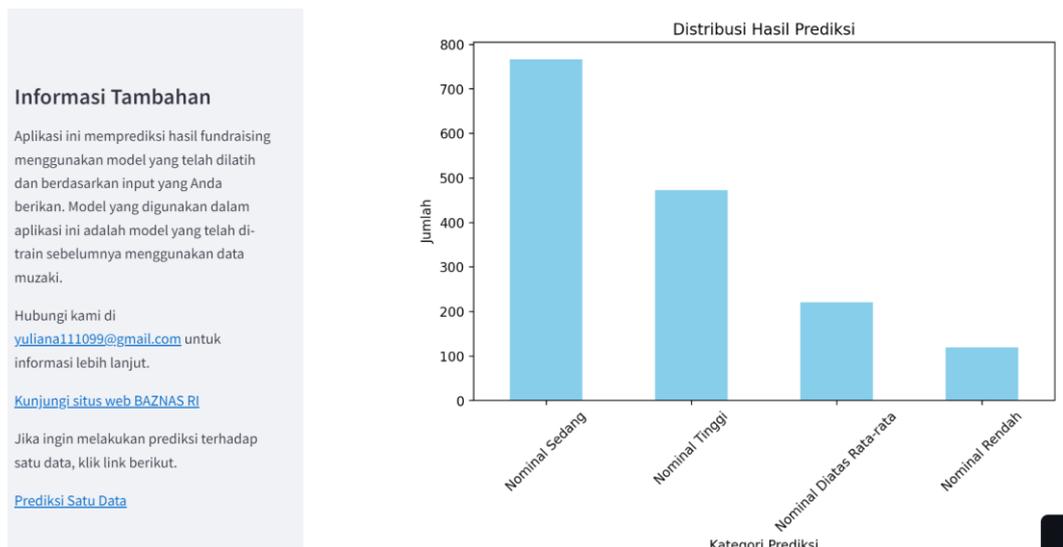
[Browse files](#)

test_data2.csv 39.7KB

	umur	gender	occupation_group	periode_transaksi	Predictions
0	42	pria	IT	Reguler	Nominal Sedang
1	42	pria	Wiraswasta	Kurban	Nominal Diatas Rata-rata
2	40	wanita	Wiraswasta	Ramadhan	Nominal Diatas Rata-rata
3	40	wanita	Tidak Bekerja	Reguler	Nominal Sedang
4	38	pria	Wiraswasta	Reguler	Nominal Sedang
5	32	wanita	Staff	Reguler	Nominal Sedang

Gambar 4. 10 Hasil eksperimen data kelompok

Pada gambar 4.10 dapat dilihat tampilan dari aplikasi prediksi fundraising digital dengan fitur untuk melakukan *upload* file dengan format CSV. Kemudian dibagian *sidebar* juga terdapat informasi tambahan, kontak peneliti, situs web BAZNAS RI, dan link untuk melakukan prediksi data tunggal atau satu data.

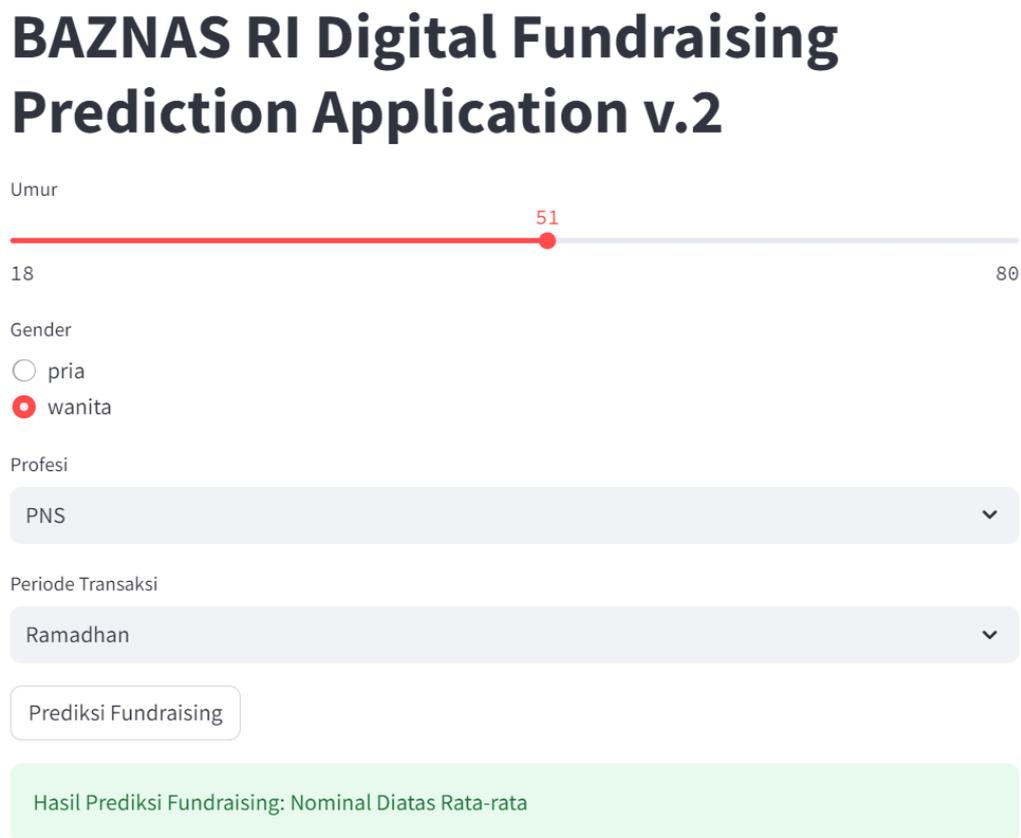


Gambar 4. 11 Visualisasi hasil eksperimen

Dapat dilihat bahwa hasil dari prediksi yang sudah dilakukan yaitu paling banyak diprediksi muzaki akan bertransaksi dengan kategori nominal sedang, sehingga dapat diprediksi bahwa pengumpulan dana zakat akan mencapai target.

b. Eksperimen Data Tunggal

Dalam eksperimen menggunakan data tunggal melalui tautan berikut <https://prediksifundraisingdigitalv2.streamlit.app/>. Setelah mengakses *platform* tersebut, pengguna diminta untuk mengatur parameter dengan menggeser slide umur ke angka 51 tahun, memilih gender sebagai wanita, profesi sebagai PNS, dan periode transaksi sebagai Ramadhan. Dengan mengatur parameter-parameter tersebut, pengguna dapat memperoleh hasil prediksi yaitu jumlah nominal diatas rata-rata yang mungkin terkumpul dari kampanye *fundraising*. Dapat dilihat tampilan dari aplikasi web streamlit pada gambar berikut.



BAZNAS RI Digital Fundraising Prediction Application v.2

Umur: 18 — 51 — 80

Gender:
 pria
 wanita

Profesi: PNS

Periode Transaksi: Ramadhan

Prediksi Fundraising

Hasil Prediksi Fundraising: Nominal Diatas Rata-rata

Gambar 4. 12 Hasil Eksperimen Data Tunggal

Keterangan rentang data nominal dari setiap kategori didapat dari klasterisasi nominal yang sudah dilakukan. Dengan keterangan kategori nominal sebagai berikut (Rupiah):

1. Nominal Rendah: 10.000 s/d 2.146.000
2. Nominal Sedang: 2.175.089 s/d 8.000.000

3. Nominal Diatas rata-rata: 8.058.000 s/d 18.306.000
4. Nominal Tinggi: 19.165.165 s/d 37.936.944

Hasil dari pembuatan model prediksi *fundraising* menunjukkan tingkat akurasi sebesar 92% dengan menggunakan algoritma C4.5. Dengan proses pembangunan model yang cermat, model ini berhasil memprediksi hasil *fundraising* dengan tingkat akurasi yang memuaskan, berdasarkan data yang telah dipersiapkan sebelumnya.

Analisis yang dilakukan juga mengungkap faktor-faktor penting yang mempengaruhi hasil *fundraising*, seperti umur, gender, periode transaksi, dan kategori nominal, yang tercermin dalam kontribusi variabel-variabel tersebut dalam proses pengambilan keputusan model. Misalnya, dapat melihat bahwa usia muzaki, jenis kelamin, dan kategori nominal donasi memiliki dampak yang signifikan terhadap jumlah dana yang berhasil dikumpulkan dalam sebuah periode waktu tertentu. Dengan memahami variabel-variabel kunci yang mempengaruhi hasil *fundraising*, BAZNAS RI dapat mengalokasikan sumber daya dengan lebih efisien, mengoptimalkan strategi dan efektivitas pengumpulan zakat. Hasil ini memberikan wawasan bagi BAZNAS untuk meningkatkan kinerjanya berdasarkan informasi yang diberikan oleh model prediksi. Dengan demikian, pembuatan model prediksi *fundraising* memberikan kontribusi yang signifikan dalam mendukung pencapaian tujuan BAZNAS dalam meningkatkan efektivitas kampanye *fundraising* digital, serta memungkinkan untuk mengoptimalkan strategi dan upaya *fundraising* berdasarkan analisis data yang lebih mendalam.

BAB V

PENUTUP

Dalam penelitian ini, telah berhasil melakukan analisis terhadap data yang dikumpulkan dan mengimplementasikan model prediksi yang efektif.

5.1 Kesimpulan

Dari hasil penelitian yang telah dilakukan, dapat ditarik kesimpulan mengenai temuan utama dan kontribusi yang berhasil diperoleh sebagai berikut.

1. Hasil dari pembuatan model prediksi *fundraising* menunjukkan tingkat akurasi sebesar 92% dengan menggunakan algoritma C4.5.
2. Faktor-faktor penting yang mempengaruhi hasil *fundraising* yaitu umur, gender, periode transaksi, dan kategori nominal, yang tercermin dalam kontribusi variabel-variabel tersebut dalam proses pengambilan keputusan model. Misalnya, dapat melihat bahwa usia muzaki, jenis kelamin, dan kategori nominal donasi memiliki dampak yang signifikan terhadap jumlah zakat yang berhasil dikumpulkan dalam sebuah periode waktu tertentu.
3. Dengan memahami variabel-variabel yang mempengaruhi hasil *fundraising*, BAZNAS RI dapat mengalokasikan sumber daya dengan lebih efisien, mengoptimalkan strategi dan efektivitas pengumpulan zakat. Hasil ini memberikan wawasan bagi BAZNAS untuk meningkatkan kinerjanya berdasarkan informasi yang diberikan oleh model prediksi.

5.2 Saran

Berdasarkan hasil penelitian ini, beberapa saran dapat diberikan untuk pengembangan selanjutnya.

1. Disarankan BAZNAS untuk terus meningkatkan kualitas dan kelengkapan data yang digunakan dalam model prediksi *fundraising*. Hal ini dapat dilakukan dengan meningkatkan aplikasi sistem informasi terkhusus bagian pengumpulan dana zakat, sosialisasi terhadap penggunaan aplikasi sistem informasi BAZNAS agar tidak terjadi salah input data baik kepada petugas konter maupun yang lainnya. Hal ini akan berpengaruh terhadap tingginya kualitas data yang dikumpulkan dengan memperhatikan aspek-aspek seperti keberadaan *missing values* atau *outlier* dalam data. Jika sudah memiliki data dengan kualitas yang tinggi, maka akan mudah dilakukan pengolahan terhadap data tersebut.
2. BAZNAS perlu terus memonitor dan mengevaluasi performa model prediksi ini secara berkala, serta melakukan *tuning* dan penyesuaian yang diperlukan untuk meningkatkan akurasi dan validitasnya.
3. BAZNAS juga dapat mempertimbangkan untuk mengintegrasikan model prediksi ini ke dalam sistem pengelolaan, sehingga dapat digunakan secara langsung dalam pengambilan keputusan dan perencanaan strategi *fundraising digital*.
4. Disarankan juga untuk menyesuaikan data mentah yang diperoleh dari sistem informasi BAZNAS RI menjadi data uji yang siap digunakan untuk melakukan prediksi *fundraising digital* sebelum melakukan pengunggahan. Hal ini memastikan bahwa data yang digunakan untuk prediksi sesuai dengan format yang diharapkan oleh model yang telah dibangun.

DAFTAR PUSTAKA

- Agustina, I., Mulyani, Y., Septiana, T., & Mardiana, M. (2022). Analisis Pengembangan Model Prediksi Kesuksesan Kickstarter menggunakan Algoritma Backpropagation dan Random Forest. *Jurnal Informatika Dan Teknik Elektro Terapan*, 10(3), 173–182. <https://doi.org/10.23960/jitet.v10i3.2742>
- Alfian, M., & Widodo, N. M. (2022). Faktor Penentu Keberhasilan Penggunaan Fintech dalam Pembayaran Zakat , Infaq , Shodaqoh (ZIS). *Jurnal Ilmiah Ekonomi Islam*, 8(03), 2651–2656.
- Alim, M. N., & Z. Basri, M. T. (2020). Financial Determinants In Zakat Institution Management Effecting muzakky loyalty In Indonesia. *International Journal of Business and Management*, 9(2), 35–47.
- BAZNAS. (2019). *Badan Amil Zakat Nasional*. <https://baznas.go.id/profil>
- Dani, H. (2022). Review on Frameworks Used for Deployment of Machine Learning Model. *International Journal for Research in Applied Science and Engineering Technology*, 10(2), 211–215. <https://doi.org/10.22214/ijraset.2022.40222>
- Dewi, A. O. P. (2020). Big Data di Perpustakaan dengan Memanfaatkan Data Mining. *Anuva: Jurnal Kajian Budaya, Perpustakaan, Dan Informasi*, 4(2), 223–230. <https://doi.org/10.14710/anuva.4.2.223-230>
- Doddy, M., Ali, J., Hindardjo, A., & Sani, A. (2022). Boost Zakat Fundraising Through E-Customer Relationship Management in Digital Era. *Technium Business and Management*, 2(2), 61–70. <https://doi.org/10.47577/business.v2i2.6762>
- Febiana, N., Tanjung, H., & Hakiem, H. (2022). Pengaruh literasi zakat, infaq, shadaqah (ZIS), kepercayaan, dan brand awareness terhadap keputusan menyalurkan zakat dan donasi melalui Tokopedia: studi

pada mahasiswa FAI Universitas Ibn Khaldun Bogor pada angkatan 2017-2018. *El-Mal Jurnal Kajian Ekonomi & Islam*, 5(2), 291–313.

Febriarini, A. S., & Astuti, E. Z. (2019). Penerapan Algoritma C4.5 untuk Prediksi Kepuasan Penumpang Bus Rapid Transit (BRT) Trans Semarang. *Eksplora Informatika*, 8(2), 95–103. <https://doi.org/10.30864/eksplora.v8i2.156>

Ferdyandi, M., Setiawan, N. Y., & Abdurrachman Bachtiar, F. (2022). *Prediksi Potensi Penjualan Makanan Beku berdasarkan Ulasan Pengguna Shopee menggunakan Metode Decision Tree Algoritma C4.5 dan Random Forest (Studi Kasus Dapur Lilis)*. 6(2), 588–596.

Fikri, A., & Verina, W. (2020). Penerapan Data Mining Untuk Prediksi Penjualan Alat Medis Menggunakan Algoritma C4.5 Pt. Murni Indah Sentosa. *Infosys (Information System) Journal*, 5(1), 70–82. <https://doi.org/10.22303/infosys.5.1.2020.70-83>

Gaol, N. Y. L. (2020). Prediksi Mahasiswa Berpotensi Non Aktif Menggunakan Data Mining dalam Decision Tree dan Algoritma C4.5. *Jurnal Informasi & Teknologi*, 2(1), 23–29. <https://doi.org/10.37034/jidt.v2i1.22>

Ginting, V. S., Kusriani, K., & Taufiq, E. (2020). Implementasi Algoritma C4.5 untuk Memprediksi Keterlambatan Pembayaran Sumbangan Pembangunan Pendidikan Sekolah Menggunakan Python. *Inspiration: Jurnal Teknologi Informasi Dan Komunikasi*, 10(1), 36–44. <https://doi.org/10.35585/inspir.v10i1.2535>

Hasanah, M. A., Soim, S., & Handayani, A. S. (2021). Implementasi CRISP-DM Model Menggunakan Metode Decision Tree dengan Algoritma CART untuk Prediksi Curah Hujan Berpotensi Banjir. In *Journal of Applied Informatics and Computing (JAIC)* (Vol. 5, Issue 2).

Hudaefi, F. A., Beik, I. S., Zaenal, M. H., Choiri, M., Farchatunnisa, H., & Junari, U. L. (2020). How Does Zakat Institution Respond To Fintech?

- Evidence From Baznas, Indonesia. *International Journal of Zakat and Islamic Philanthropy*, 2(1), 32–40.
- Karim, A., Mufakhidin, A., Kusuma, H. H., Adeni, A., & Fitri, F. (2022). Mitigating Poverty: the Clustering of Potential Zakat in Indonesia. *Analisa: Journal of Social Science and Religion*, 7(1), 109–126. <https://doi.org/10.18784/analisa.v7i1.1641>
- Keputusan Presiden Republik Indonesia Nomor 8. (2001). <https://peraturan.bpk.go.id/Details/57030/keppres-no-8-tahun-2001>
- Liliana, D. Y., Maulana, H., & Setiawan, A. (2021). Data Mining untuk Prediksi Status Pasien Covid-19 dengan Pengklasifikasi Naïve Bayes. *Multinetics*, 7(1), 48–53. <https://doi.org/10.32722/multinetics.v7i1.3786>
- Mait, C. D., Watuseke, J. A., Saerang, P. D. G., & Joshua, S. R. (2022). Sistem Pendukung Keputusan Menggunakan Fuzzy Logic Tahani Untuk Penentuan Golongan Obat Sesuai Dengan Penyakit Diabetes. *Jurnal Media Infotama*, 18(2), 344.
- Mardi, Y. (2019). Data Mining : Klasifikasi Menggunakan Algoritma C4 . 5 Data mining merupakan bagian dari tahapan proses Knowledge Discovery in Database (KDD) . *Jurnal Edik Informatika. Jurnal Edik Informatika*, 2(2), 213–219.
- Pambudi, A., & Abidin, Z. (2023). Penerapan CRISP-DM pada Data Saham PT. Telkom Indonesia Tbk (Studi Kasus: Bursa Efek Indonesia Tahun 2015-2022). *JDMSI*, 4(1), 1–14.
- Paskalis, E., Orpa, K., & Ripanti, E. F. (2019). Model Prediksi Awal Masa Studi Mahasiswa. *Jurnal Sistem Dan Teknologi Informasi*, 7(4), 272–278.
- Perdana, B. A., & Zen, M. (2020). Fundraising Dana Infak dan Sedekah dalam Meningkatkan Kepercayaan Jamaah Masjid. *Tadbir: Jurnal Manajemen Dakwah*, 5(2), 137–146. <https://doi.org/10.15575/tadbir.v5i2.2099>

- Permana, A. Y., & Romadlon, P. (2019). Perancangan Sistem Informasi Penjualan Perumahan Menggunakan Metode SDLC Pada PT. Mandiri Land Prosperous Berbasis Mobile. *SIGMA – Jurnal Teknologi Pelita Bangsa*, 10(2), 153–167.
- Raflesia, S. P., Lestarini, D., Kurnia, R. D., & Hardiyanti, D. Y. (2023). Using machine learning approach towards successful crowdfunding prediction. *Bulletin of Electrical Engineering and Informatics*, 12(4), 2438–2445. <https://doi.org/10.11591/eei.v12i4.5238>
- Ryoba, M. J., Qu, S., Ji, Y., & Qu, D. (2020). The right time for crowd communication during campaigns for sustainable success of crowdfunding: Evidence from Kickstarter platform. *Sustainability (Switzerland)*, 12(18). <https://doi.org/10.3390/su12187642>
- Suweleh, A. S., Susilowati, D., & Hairani. (2020). Aplikasi Penentuan Penerima Beasiswa Menggunakan Algoritma C4.5. *Jurnal BITE*, 2(1), 12–21. <https://doi.org/10.30812/bite.v2i1.798>
- Takdirillah, R. (2020). Penerapan Data Mining Menggunakan Algoritma Apriori Terhadap Data Transaksi Sebagai Pendukung Informasi Strategi Penjualan. *Edumatic : Jurnal Pendidikan Informatika*, 4(1), 37–46. <https://doi.org/10.29408/edumatic.v4i1.2081>
- Undang-Undang Nomor 23.* (2011). <https://peraturan.bpk.go.id/Details/39267/uu-no-23-tahun-2011>
- Utami, yohana T., Shofiana, D. A., & Heningtyas, Y. (2020). Penerapan Algoritma C4.5 Untuk Prediksi Churn Rate Pengguna Jasa Telekomunikasi. *Jurnal Komputasi*, 8(2), 69–76. <https://doi.org/10.23960/komputasi.v8i2.2647>
- Widaningsih, S. (2019). Perbandingan Metode Data Mining Untuk Prediksi Nilai Dan Waktu Kelulusan Mahasiswa Prodi Teknik Informatika Dengan Algoritma C4,5, Naïve Bayes, Knn Dan Svm. *Jurnal Tekno Insentif*, 13(1), 16–25. <https://doi.org/10.36787/jti.v13i1.78>
- World Giving Index 2022 A global view of giving trends.* (2022).

Zai, C. (2022). Implementasi Data Mining Sebagai Pengolahan Data.
Portal Data, 2(3), 1–12.

LAMPIRAN

Berikut disertakan lampiran berupa *source code* yang digunakan dalam pembangunan model Algoritma C4.5 untuk memprediksi hasil *fundraising* digital BAZNAS RI. *Source code* ini mencakup langkah-langkah dari tahap persiapan data hingga pembangunan model prediksi menggunakan algoritma C4.5. Selain *source code* juga terdapat hasil model, file data mentah, file data untuk prediksi, model pipeline, dan *requirements*. Semuanya tersimpan pada <https://github.com/yuliana4763/DataScience>. Semoga lampiran ini memberikan gambaran yang lebih jelas tentang implementasi model prediksi yang telah dibahas dalam penelitian ini. Berikut ini juga dilampirkan beberapa record data sebelum dan sesudah dilakukan nya prediksi.

Lampiran 1. Data Mentah sebelum dilakukan nya prediksi

No	umur	gender	occupation group	periode transaksi
1	42	pria	IT	Reguler
2	42	pria	Wiraswasta	Kurban
3	40	wanita	Wiraswasta	Ramadhan
4	40	wanita	Tidak Bekerja	Reguler
5	38	pria	Wiraswasta	Reguler
6	32	wanita	Staff	Reguler
7	30	pria	Teknisi	Ramadhan
8	28	pria	Staff	Reguler
9	34	pria	Staff	Reguler
10	42	pria	Tidak Bekerja	Reguler
11	34	pria	Amil	Reguler
12	45	pria	Amil	Kurban
13	31	pria	Amil	Kurban
14	42	pria	PNS	Ramadhan
15	40	pria	Amil	Reguler
...
1565	45	pria	Amil	Kurban
1566	47	pria	Industri	Reguler
1567	69	pria	Lainnya	Ramadhan
1568	45	pria	Amil	Reguler
1569	32	pria	PNS	Ramadhan
1570	35	pria	Amil	Kurban
1571	40	pria	Staff	Reguler
1572	40	pria	Tidak Bekerja	Reguler
1573	35	pria	Pegawai Swasta	Ramadhan
1574	51	pria	Lainnya	Ramadhan
1575	30	wanita	Pegawai Swasta	Reguler
1576	27	wanita	Amil	Reguler
1577	30	pria	Marketing	Reguler

1578	38	wanita	Lainnya	Reguler
1579	40	wanita	Konsultan	Reguler
1580	45	pria	Amil	Kurban

Lampiran 2. 1 Data hasil prediksi

No	umur	gender	occupation group	periode transaksi	Predictions
1	42	pria	IT	Reguler	Nominal Sedang
2	42	pria	Wiraswasta	Kurban	Nominal Diatas Rata-rata
3	40	wanita	Wiraswasta	Ramadhan	Nominal Diatas Rata-rata
4	40	wanita	Tidak Bekerja	Reguler	Nominal Sedang
5	38	pria	Wiraswasta	Reguler	Nominal Sedang
6	32	wanita	Staff	Reguler	Nominal Sedang
7	30	pria	Teknisi	Ramadhan	Nominal Diatas Rata-rata
8	28	pria	Staff	Reguler	Nominal Sedang
9	34	pria	Staff	Reguler	Nominal Sedang
10	42	pria	Tidak Bekerja	Reguler	Nominal Sedang
11	34	pria	Amil	Reguler	Nominal Tinggi
12	45	pria	Amil	Kurban	Nominal Tinggi
13	31	pria	Amil	Kurban	Nominal Tinggi
14	42	pria	PNS	Ramadhan	Nominal Diatas Rata-rata
15	40	pria	Amil	Reguler	Nominal Tinggi
...
1565	45	pria	Amil	Kurban	Nominal Tinggi
1566	47	pria	Industri	Reguler	Nominal Sedang
1567	69	pria	Lainnya	Ramadhan	Nominal Rendah
1568	45	pria	Amil	Reguler	Nominal Tinggi
1569	32	pria	PNS	Ramadhan	Nominal Diatas Rata-rata
1570	35	pria	Amil	Kurban	Nominal Tinggi

1571	40	pria	Staff	Reguler	Nominal Sedang
1572	40	pria	Tidak Bekerja	Reguler	Nominal Sedang
1573	35	pria	Pegawai Swasta	Ramadhan	Nominal Diatas Rata-rata
1574	51	pria	Lainnya	Ramadhan	Nominal Rendah
1575	30	wanita	Pegawai Swasta	Reguler	Nominal Sedang
1576	27	wanita	Amil	Reguler	Nominal Tinggi
1577	30	pria	Marketing	Reguler	Nominal Sedang
1578	38	wanita	Lainnya	Reguler	Nominal Sedang
1579	40	wanita	Konsultan	Reguler	Nominal Sedang
1580	45	pria	Amil	Kurban	Nominal Tinggi