BAB 3
METODOLOGI PENELITIAN

3.1 Tahapan Penelitian

Tahapan-tahapan penelitian disusun secara sistematis dan terstruktur mulai dari akuisisi data hingga dokumentasi akhir. Rincian tahapan disajikan sebagai berikut.

3.1.1 Pengumpulan Data Awal

Data awal dalam penelitian ini diperoleh dari toko online milik peneliti yang menyediakan berbagai jenis peralatan pancing. Data ini diekspor dari sistem manajemen toko dalam format spreadsheet dan mencakup ratusan entri produk dengan berbagai atribut, seperti kode produk, nama produk, variasi, harga, jumlah stok, dan informasi administratif lainnya.

Namun demikian, struktur data awal belum sepenuhnya siap untuk diproses dalam sistem pembelajaran mesin. Banyak kolom bersifat administratif dan tidak relevan dengan kebutuhan analisis, seperti `SKU`, `Stok`, `Kode Produk`, dan pengaturan batas pembelian. Selain itu, beberapa atribut penting seperti jenis alat, satuan kekuatan (dalam lbs), dan kategori lokasi penggunaan belum tersedia dalam format terstruktur.

Gambar 3.1 berikut menunjukkan cuplikan tampilan awal dari dataset mentah hasil ekspor dari toko online tersebut.

	A	B	C	D	E	F	G	H	I	J	K	L
3	Kode Produk	Nama Produk	Kode Variasi	Nama Variasi	SKU Induk	SKU	Harga	Stok	Min. Jumlah Pembelian	Maks. Jumlah Pembelian	Maks. Jumlah Pembelian - Tanggal Mulai	Maks. Jumlah Pembelian - Jumlah Hari
6							Mohon masukkan 99 sampai 100000000 untuk harga produk. Batas harga produk termahal dibagi harga		Min. jumlah pembelian merupakan isi dari tingkatan produk. Pembeli dapat memesan variasi yang berbeda untuk	[Per Pesanan + Per Periode] Pengaturan ini akan membatasi Maks. jumlah pembelian yang dapat	[Hanya untuk Pengaturan Per Periode] Mohon tentukan tanggal mulai. Tanggal mulai tercepat adalah hari. Mohon	[Hanya untuk Pengaturan Per Periode] Batas Maks. jumlah pembelian akan berakhir (untuk tipe periode "Tidak
7	29853380891	Reel Ryobi ECUSIMA 2000 Power Handle	256059394328				565000	45	0			
8	24657189627	Kail Pancing Mugen UMITANAGO / KEIRYU	232438379853	UMITANAGO,No.10			22000	30	0			
9	24657189627	Kail Pancing Mugen UMITANAGO / KEIRYU	232438379854	UMITANAGO,No.12			22000	30				
10	24657189627	Kail Pancing Mugen UMITANAGO / KEIRYU	232438379849	UMITANAGO,No.6			22000	30				
11	24657189627	Kail Pancing Mugen UMITANAGO / KEIRYU	232438379850	UMITANAGO,No.7			22000	30				
12	24657189627	Kail Pancing Mugen UMITANAGO / KEIRYU	232438379851	UMITANAGO,No.8			22000	30				
13	24657189627	Kail Pancing Mugen UMITANAGO / KEIRYU	232438379852	UMITANAGO,No.9			22000	30				
14	24035293506	Senar Water Line VINOOTO Hijau	16838314644	0.28mm			11000	200	0			
15	24035293506	Senar Water Line VINOOTO Hijau	16838314645	0.30mm			11000	200				
16	24035293506	Senar Water Line VINOOTO Hijau	16838314640	0.18mm			11000	197				
17	24035293506	Senar Water Line VINOOTO Hijau	16838314641	0.20mm			11000	200				
18	24035293506	Senar Water Line VINOOTO Hijau	16838314642	0.23mm			11000	200				
19	24035293506	Senar Water Line VINOOTO Hijau	16838314643	0.25mm			11000	200				
20	23938140888	Reel Getsu RANKERS 3000	886825898671				227000	25	0			
21	23743822446	Reel Shikari TOKYO 3000 Power Handle	49662927801				168000	0	0			
22	23710515353	Senar DAMYL Putih 8lb,10lb,12lb	230475850633	PUTIH,8 Lb/0.30			8200	156	0			
23	23710515353	Senar DAMYL Putih 8lb,10lb,12lb	230475850634	PUTIH,10 Lb/0.32			9400	87				
24	23710515353	Senar DAMYL Putih 8lb,10lb,12lb	230475850635	PUTIH,12 Lb/0.35			11300	158				
25	23570641641	Reel Shikari RAIZE 4000 Power Handle	221698691704				170000	0	0			
26	23569317157	Reel Daido GLAZER 3000	158224976403				107000	0	0			
27	23271154029	Joran Daido ATHENA 3 III Pro Series 180	184370783619	180/10-20 Lbs			310000	0	0			
28	23271154029	Joran Daido ATHENA 3 III Pro Series 180	184370783618	180/8-17 Lbs			310000	0				
29	23245142140	Reel Shikari RAIZE 3000 Power Handle	235304408280				165000	0	0			
30	23171536633	Reel Elito BATTOSAI 2000 Power Handle	137617558062				150000	38	0			
31	23070717242	Senar Duraking PREMIER 150 mtr	165940307533	0.30mm/16Lb			26000	23	0			
32	23070717242	Senar Duraking PREMIER 150 mtr	165940307528	0.18mm/8Lb			26000	24				
33	23070717242	Senar Duraking PREMIER 150 mtr	165940307529	0.25mm/14Lb			26000	26				
34	23070717242	Senar Duraking PREMIER 150 mtr	165940307530	0.20mm/10Lb			26000	17				
35	23070717242	Senar Duraking PREMIER 150 mtr	165940307531	0.22mm/12Lb			26000	10				

Gambar 3.1. Cuplikan tabel dataset mentah hasil ekspor dari toko online

Dari gambar tersebut terlihat bahwa meskipun beberapa kolom seperti `Nama Produk` dan `Harga` dapat dimanfaatkan, sebagian besar informasi masih perlu diproses ulang. Hal ini meliputi penghilangan atribut yang tidak relevan, penyatuan entri produk yang serupa (karena perbedaan variasi warna), serta konversi kekuatan alat dari satuan *pound* (lbs) ke *kilogram* untuk keperluan pemodelan.

Seluruh proses akuisisi, pembersihan, dan penyusunan dataset primer dilakukan sendiri oleh penulis, mulai dari ekspor data toko online, deduplikasi, konversi satuan, hingga perancangan format dataset agar siap dipakai pada sistem pembelajaran mesin.

Langkah-langkah praproses selanjutnya akan dijelaskan pada bagian berikutnya untuk menghasilkan dataset yang bersih dan siap digunakan dalam proses pelatihan model rekomendasi.

3.1.2 Praproses dan Pembersihan Data

Dataset yang diperoleh dari toko online awalnya terdiri dari 1055 entri produk. Setelah dilakukan tinjauan lebih lanjut, ditemukan bahwa banyak produk mengalami duplikasi, terutama karena perbedaan variasi non-fungsional seperti warna, kemasan, atau kode SKU. Karena penelitian ini berfokus pada analisis fungsional peralatan pancing, maka dilakukan proses deduplikasi untuk menghapus produk yang identik dari sisi fungsi namun berbeda secara kosmetik.

Tahapan pembersihan data dilakukan melalui langkah-langkah berikut:

- Menghapus atribut yang tidak relevan untuk analisis seperti `Kode Produk`, `Kode Variasi`, `SKU`, dan informasi stok,
- Mengelompokkan produk ke dalam empat jenis utama berdasarkan deskripsi nama, yaitu `joran`, `reel`, `senar`, dan `kail`,
- Mengonversi satuan kekuatan dari *pound* (lbs) ke *kilogram* (1 lbs $\approx$ 0,45 kg) untuk memastikan keseragaman numerik antar item,
- Melakukan pemformatan ulang harga menjadi bentuk numerik tunggal tanpa pemisah ribuan atau simbol mata uang.

Dataset juga diperkaya dengan tiga atribut tambahan berupa fitur biner: `Empang`, `Sungai`, dan `Laut`, yang menunjukkan kecocokan produk terhadap jenis lokasi memancing. Nilai untuk atribut ini diperoleh melalui survei kepada pengguna yang akan dijelaskan pada tahapan selanjutnya.

Setelah seluruh proses pembersihan dan transformasi selesai dilakukan, diperoleh total 695 produk unik yang siap digunakan untuk pelatihan model. Rincian jumlah per kategori adalah 137 joran, 135 reel, 331 senar, dan 91 kail. Struktur akhir dataset ditunjukkan pada Gambar 3.2 berikut.

	A	B	C	D	E	F	G
1	Nama Barang	Harga	Jenis Barang	Kekuatan (KG)	Empang	Sungai	Laut
2	Joran Daido ATHENA 3 III Pro Series 180 10-20Lbs	310000	Joran	7	0	0	1
3	Joran Daido ATHENA 3 III Pro Series 180 8-17Lbs	310000	Joran	6	1	1	1
4	Joran Getsu ROYAL ATTACK 165 7-15Lbs	175000	Joran	5	1	1	0
5	Joran Getsu ROYAL ATTACK 165 8-17Lbs	175000	Joran	6	1	1	1
6	Joran Shikari STELLARIS 180 7-15Lbs	130000	Joran	5	1	1	0
7	Joran Shikari STELLARIS 180 10-20Lbs	130000	Joran	7	0	0	1
8	Joran Shikari STELLARIS 180 8-17Lbs	130000	Joran	6	1	1	1
9	Joran Daido PANDORA II 165 180 10-20Lbs	265000	Joran	7	0	0	1
10	Joran Daido PANDORA II 165 10-20Lbs	260000	Joran	7	0	0	1
11	Reel Shikari RAIZE 3000 Power Handle	165000	Reel	7	0	0	1
12	Reel Elito BATTOSAI 2000 Power Handle	150000	Reel	5	1	1	0
13	Reel Daido MEDALLION 4000 Power Handle	194000	Reel	8	0	0	1
14	Reel Daido MEDALLION 2000 Power Handle	180000	Reel	5	1	1	0
15	Reel Daido SPYRO 3000 Power Handle	185000	Reel	7	0	0	1
16	Reel Elito BATTOSAI 1000 Power Handle	145000	Reel	4	1	1	0
17	Reel Elito BATTOSAI 3000 Power Handle	151000	Reel	7	0	0	1
18	Reel Fly Fish BANDHIT 3000	192000	Reel	7	0	0	1
19	Reel Fly Fish BANDHIT 1000	177000	Reel	4	1	1	0
20	Reel Daido SIREN 800	141000	Reel	3	1	1	0
21	Senar TORTUE Ultra Grafite 0.40mm	22500	Senar	17	0	0	1
22	Senar TORTUE Ultra Grafite 0.35mm	22500	Senar	13	0	0	1
23	Senar TORTUE Ultra Grafite 0.20mm	22500	Senar	4	1	1	0
24	Senar Duraking FASTLINE 150 mtr 0.28mm	77000	Senar	9	0	0	1
25	Senar Duraking FASTLINE 150 mtr 0.20mm	28000	Senar	4	1	1	0
26	Senar Duraking FASTLINE 150 mtr 0.30mm	28000	Senar	10	0	0	1
27	Senar Sinsung Hi Class 0.14mm	19500	Senar	2	1	1	0
28	Senar Sinsung Hi Class 0.23mm	19500	Senar	6	1	1	1
29	Senar DAMYL Putih 15lb,20lb,25lb 0.45mm	16200	Senar	20	0	0	1
30	Senar Blood Black Trout FC 0.36mm	48000	Senar	14	0	0	1
31	Kail Mugen SURE K.CHINU No.5	24000	Kail	4	1	1	0
32	Kail Mugen SURE K.CHINU No.6	24000	Kail	4	1	1	0
33	Kail Pancing HANBO No 9	3500	Kail	6	1	1	1
34	Kail Stainless Steel Kotakan Gamamis No.4	15000	Kail	4	1	1	0
35	Kail FLY FISH ACXIS Ring No.1	2300	Kail	2	1	1	0
36	Kail Mugen Kantuki Chinu Size 7	23500	Kail	6	1	1	1
37	Kail Gamakatsu Chinu Ring Eye No.3	35000	Kail	4	1	1	0
38	Kail Pancing Super Golden Fish No.4	9000	Kail	4	1	1	0
39	Kail Pancing Bradley No.6	28000	Kail	4	1	1	0

Gambar 3.2. Cuplikan tabel dataset setelah pembersihan dan konversi satuan

Seperti terlihat pada Gambar 3.2, dataset akhir telah tersusun dalam bentuk tabel yang lebih rapi dan siap digunakan dalam proses pelatihan sistem rekomendasi. Kolom-kolom utama meliputi:

- `Nama Barang`: nama lengkap produk alat pancing,
- `Harga`: nilai harga dalam satuan rupiah, diformat sebagai bilangan numerik,
- `Jenis Barang`: jenis produk, yaitu `joran`, `reel`, `senar`, atau `kail`,
- `Kekuatan (kg)`: kekuatan tarik maksimum alat dalam kilogram,
- `Empang`, `Sungai`, dan `Laut`: kolom biner (0 atau 1) untuk menunjukkan kecocokan terhadap masing-masing lokasi memancing.

Nilai 1 pada kolom lokasi berarti produk tersebut dinilai cocok untuk digunakan pada lokasi tersebut, sedangkan nilai 0 menunjukkan ketidakcocokan. Dengan demikian, satu produk bisa memiliki lebih dari satu nilai 1 jika dinilai cocok untuk beberapa lokasi sekaligus. Pendekatan ini memungkinkan sistem merekomendasikan satu produk dalam lebih dari satu konteks lokasi.

Penetapan model biner ini juga didukung oleh pendekatan selektivitas alat pancing, yang menekankan bahwa efektivitas suatu alat lebih ditentukan oleh desain dan kekuatannya, bukan semata-mata oleh jenis lokasi geografis [2]. Selain itu, studi tentang alat tangkap pasif juga menunjukkan bahwa gear seperti joran atau senar dapat digunakan fleksibel di berbagai perairan, selama spesifikasi teknis sesuai dengan target spesies dan kondisi perairan [38].

Dataset yang telah dibersihkan ini kemudian digunakan untuk membentuk jutaan kombinasi alat pancing valid yang menjadi dasar dari sistem rekomendasi berbasis algoritma *K-Nearest Neighbors*.

3.1.3 Pengumpulan Data Pelengkap melalui Survei

Pengumpulan data pelengkap melalui survei dilakukan untuk melengkapi dataset awal toko online yang tidak memuat informasi eksplisit mengenai kecocokan alat pancing dengan lokasi memancing (empang, sungai, atau laut). Survei ini menjadi dasar pengambilan keputusan pada sistem rekomendasi berbasis preferensi nyata pengguna.

A Desain dan Distribusi Survei

Survei dalam penelitian ini dirancang menggunakan pendekatan studi kuantitatif *cross-sectional* dan memanfaatkan instrumen kuesioner tertutup dengan skala nominal dan pola *multiple response*. Berikut penjelasan aspek metodologis yang diterapkan:

1. **Desain Studi *Cross-Sectional***
   Desain *cross-sectional* merupakan metode pengumpulan data di mana seluruh data survei dikumpulkan dalam satu periode waktu tertentu (snapshot) pada populasi yang menjadi target. Pendekatan ini banyak digunakan dalam penelitian sosial, perilaku konsumen, dan evaluasi layanan, karena memungkinkan analisis hubungan variabel pada satu titik waktu tanpa perlu observasi longitudinal [39]. Studi *cross-sectional* efisien dari segi waktu,

biaya, dan dapat menangkap pola agregat persepsi serta preferensi dalam konteks yang relevan pada saat survei dilakukan.

2. **Instrumen *Multiple Response***

   Kuesioner menggunakan skala nominal dengan model *multiple response*, di mana setiap responden diperbolehkan memilih lebih dari satu kategori lokasi (empang, sungai, laut) untuk setiap spesifikasi alat pancing. Model ini mengakomodasi realita di lapangan, di mana satu alat dapat digunakan secara fleksibel pada beberapa lokasi sekaligus, sehingga hasil survei lebih representatif dan tidak bias tunggal [40]. Penggunaan opsi *multiple response* direkomendasikan dalam survei yang mengeksplorasi perilaku, preferensi, atau situasi penggunaan multifungsi pada populasi nyata.

3. **Teknik *Purposive Sampling* dan Model Self-Administered**

   Distribusi survei dilakukan dengan teknik *purposive sampling*, yaitu pemilihan sampel secara sengaja hanya kepada pelanggan aktif toko online peneliti yang dianggap paling relevan dan memiliki pengalaman nyata dengan produk yang diteliti [41]. Implementasi *purposive sampling* meningkatkan validitas survei, mengurangi bias eksternal, dan memastikan setiap responden benar-benar berasal dari populasi target. Setiap paket pesanan diberikan sisipan *QR Code* yang mengarahkan langsung ke kuesioner digital Google Form (Gambar 3.3), sehingga survei dapat diisi secara *self-administered*. Metode self-administered berbasis QR Code terbukti meningkatkan tingkat respons dan kejujuran jawaban, sekaligus mempermudah distribusi instrumen survei secara efisien [42, 43].

UNIVERSITAS MULTIMEDIA NUSANTARA

Gambar 3.3. Paket pesanan dengan sisipan *QR Code* survei

Gambar 3.3 di atas mendokumentasikan contoh nyata dari distribusi QR Code pada paket pesanan yang dikirim ke pelanggan. Dengan pendekatan ini, setiap pelanggan yang menerima produk secara otomatis memperoleh akses eksklusif ke kuesioner survei, sehingga seluruh responden survei benar-benar berasal dari pengguna aktual produk pancing. Strategi distribusi seperti ini memastikan validitas data serta menurunkan potensi bias responden dari luar populasi target, sebagaimana disarankan dalam penelitian-penelitian survei berbasis QR Code dan purposive sampling [42, 43, 41].

B Jumlah Responden dan Kecukupan Statistik

Selama periode survei, sebanyak 383 responden berpartisipasi dan mengisi kuesioner secara lengkap. Jumlah ini diperoleh secara sukarela dari pelanggan aktif yang menerima sisipan *QR Code* pada paket pesanan. Untuk memastikan kecukupan jumlah responden secara statistik, dilakukan perhitungan kebutuhan minimum sampel berdasarkan rumus Slovin.

$$n = \frac{N}{1 + Ne^2} \tag{3.1}$$

(Rumus Slovin untuk Penentuan Jumlah Sampel Minimum)

Pada persamaan (3.1), n adalah jumlah sampel minimum yang diperlukan, N adalah jumlah populasi (dalam penelitian ini diasumsikan sebanyak 1.000 pelanggan aktif), dan e adalah batas toleransi kesalahan atau margin of error yang diinginkan (seringkali digunakan $e = 0{,}05$ atau 5%).

Dengan memasukkan nilai $N = 1.000$ dan $e = 0,05$ ke dalam rumus Slovin, diperoleh:

$$n = \frac{1000}{1 + 1000 \times 0,05^2} = \frac{1000}{1 + 2,5} = \frac{1000}{3,5} \approx 286$$

Artinya, sampel minimum yang dibutuhkan untuk tingkat kepercayaan 95% dan margin of error 5% adalah 286 responden. Dengan jumlah aktual 383 responden, penelitian ini telah melampaui syarat minimum sehingga tingkat representasi, kekuatan analisis, dan reliabilitas data semakin terjamin.

Praktik penentuan kecukupan statistik seperti ini umum digunakan dalam survei sosial dan pemasaran untuk memastikan hasil analisis dapat digeneralisasikan secara valid pada populasi [41, 39]. Dengan demikian, jumlah responden dalam penelitian ini sangat memadai untuk mendukung validitas dan kredibilitas temuan.

C Karakteristik Kuesioner dan Proses Pengisian

Kuesioner terdiri dari pertanyaan tertutup *multiple response*, memungkinkan responden memilih lebih dari satu lokasi (empang, sungai, laut) untuk tiap spesifikasi alat (ukuran reel, panjang joran, kekuatan senar, ukuran kail). Setelah memindai QR Code, pelanggan diarahkan ke Google Form untuk mengisi survei secara mandiri (self-administered). Metode survei daring seperti ini dinilai efektif untuk menjangkau responden secara luas, cepat, dan biaya rendah [43].

D Rekapitulasi Data dan Penjelasan Tabel Survei

Hasil survei yang telah dikumpulkan dari para responden kemudian direkapitulasi dalam bentuk distribusi frekuensi, yang disajikan pada Tabel 3.1. Tabel ini menampilkan hasil penghitungan jumlah pemilihan setiap kategori lokasi memancing (empang, sungai, laut) berdasarkan spesifikasi alat pancing yang

diujikan melalui kuesioner. Setiap baris pada tabel merepresentasikan satu kategori spesifikasi alat pada satu pertanyaan survei, sedangkan kolom-kolom jawaban menunjukkan berapa banyak responden yang memilih lokasi tersebut sebagai tempat yang paling sesuai untuk spesifikasi yang dimaksud.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3.1. Distribusi Pilihan Lokasi Berdasarkan Spesifikasi Alat Pancing

No.	Pertanyaan	Spesifikasi	Jawaban		
			Empang	Sungai	Laut
1	Menurut Anda, untuk ukuran reel berikut, lokasi manakah yang paling sesuai?	1000–2000	261	252	–
		3000–5000	293	289	103
		6000 ke atas	120	138	383
2	Menurut Anda, untuk panjang joran berikut, cocoknya digunakan di lokasi mana?	120–150 cm	285	280	94
		160–180 cm	282	299	289
		di atas 180 cm	181	215	383
3	Menurut pengalaman Anda, senar dengan kekuatan berikut cocok untuk memancing di mana?	2–5 kg	302	284	110
		6–10 kg	282	290	279
		di atas 10 kg	185	195	383
4	Menurut Anda, ukuran kail manakah yang lebih cocok digunakan di lokasi berikut?	1–3	286	270	94
		4–6	295	278	95
		di atas 6	199	210	383

Tabel 3.1 terdiri dari beberapa bagian utama:

- **Kolom "Pertanyaan"** mendeskripsikan pernyataan atau instruksi yang diberikan pada responden terkait spesifikasi alat tertentu.
- **Kolom "Spesifikasi"** menjelaskan kategori atau varian alat pancing yang dinilai (misal: ukuran reel, panjang joran, kekuatan senar, ukuran kail).
- **Kolom "Jawaban" (Empang, Sungai, Laut)** berisi jumlah responden yang memilih masing-masing lokasi sebagai tempat paling cocok untuk spesifikasi alat tersebut.

Perlu dicatat bahwa setiap responden dapat memilih lebih dari satu lokasi untuk satu spesifikasi alat (*multiple response*), sehingga total angka pada baris yang sama dapat melebihi jumlah responden total (383 orang). Pola distribusi angka-angka pada kolom jawaban merefleksikan kecenderungan dan preferensi aktual para pemancing terkait penggunaan alat berdasarkan pengalaman lapangan.

Sebagai contoh, pada spesifikasi reel 6000 ke atas, sebanyak 383 responden memilih laut sebagai lokasi paling sesuai, sementara jumlah yang memilih empang dan sungai jauh lebih sedikit. Sebaliknya, pada kategori reel 3000–5000, preferensi responden tersebar lebih merata di antara ketiga lokasi, walaupun empang dan sungai tetap mendominasi. Demikian pula, pada alat dengan karakter multifungsi seperti senar 6–10 kg, angka pada ketiga lokasi relatif tinggi, menunjukkan fleksibilitas alat dalam berbagai kondisi.

Tabel rekapitulasi ini menjadi **landasan utama** dalam proses penentuan label kecocokan lokasi pada dataset utama sistem rekomendasi. Dengan menggunakan data empiris hasil survei, sistem rekomendasi yang dikembangkan dapat memberikan saran pemilihan alat pancing yang lebih sesuai dengan kebutuhan nyata pengguna di lapangan, sekaligus memastikan bahwa hasil analisis benar-benar berbasis pada preferensi dan pengalaman kolektif komunitas pemancing.

E Integrasi, Validasi, dan Implikasi Survei

Integrasi hasil survei ke dalam dataset utama dilakukan secara bertahap dan terstruktur agar seluruh proses dapat diaudit serta dipertanggungjawabkan secara akademis. Berikut tahapan utama proses integrasi, validasi, dan implementasi hasil survei dalam sistem rekomendasi:

1. **Rekapitulasi Jawaban Responden**
   Seluruh data hasil kuesioner direkap berdasarkan spesifikasi alat (misal: ukuran reel, panjang joran, kekuatan senar, ukuran kail) dan lokasi penggunaannya (empang, sungai, laut). Untuk setiap kombinasi, dihitung jumlah responden (n) yang memilih lokasi tertentu.

2. **Perhitungan Persentase Pilihan**
   Jumlah pemilihan untuk setiap lokasi pada tiap spesifikasi alat kemudian dikonversi ke bentuk persentase terhadap total responden ($N = 383$) menggunakan rumus:

$$\text{Persentase} = \frac{\text{Jumlah Pilihan Responden } (n)}{\text{Jumlah Total Responden } (N)} \times 100\% \tag{3.2}$$

   Penggunaan rekapitulasi dan persentase dalam survei merupakan pendekatan yang umum dalam penelitian kuantitatif dan validasi instrumen [43].

3. **Penyajian Rekapitulasi Kuantitatif**
   Hasil rekapitulasi jumlah dan persentase untuk setiap spesifikasi alat pancing disajikan dalam Tabel 3.2 berikut:

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA

Tabel 3.2. Rekapitulasi Jumlah dan Persentase Pilihan Lokasi Berdasarkan Spesifikasi Alat Pancing

No.	Jenis Alat	Spesifikasi	Empang		Sungai		Laut	
			n	(%)	n	(%)	n	(%)
1	Reel	1000–2000	261	68.1	252	65.8	0	0.0
		3000–5000	293	76.5	289	75.5	103	26.9
		6000 ke atas	120	31.3	138	36.0	383	100.0
2	Joran	120–150 cm	285	74.4	280	73.1	94	24.5
		160–180 cm	282	73.6	299	78.1	289	75.5
		180 cm ke atas	181	47.3	215	56.1	383	100.0
3	Senar	2–5 kg	302	78.8	284	74.2	110	28.7
		6–10 kg	282	73.6	290	75.7	279	72.9
		10 kg ke atas	185	48.3	195	50.9	383	100.0
4	Kail	1–3	286	74.7	270	70.5	94	24.5
		4–6	295	77.0	278	72.6	95	24.8
		6 ke atas	199	52.0	210	54.8	383	100.0

4. **Penetapan Threshold dan Pemberian Label Biner**

   (a) *Threshold Penetapan Label:*
   Suatu spesifikasi alat pancing dinyatakan cocok untuk lokasi tertentu jika persentase pemilihannya mencapai atau melebihi ambang batas (**threshold**) yang telah ditetapkan, yaitu 50%. Penetapan threshold ini mengacu pada prinsip mayoritas yang banyak digunakan dalam analisis survei kuantitatif dan telah diterapkan pada berbagai studi survei dan validasi instrumen [43, 44].

   (b) *Pemberian Label pada Dataset:*
   Jika persentase pilihan pada suatu lokasi untuk spesifikasi alat tertentu $\geq 50\%$, maka pada dataset utama kolom lokasi tersebut diisi `1` (cocok), sedangkan jika kurang dari 50% diisi `0` (tidak cocok). Dengan demikian, satu produk dapat memperoleh lebih dari satu nilai `1` jika dianggap

multifungsi oleh mayoritas responden. Pendekatan ini didukung oleh literatur selektivitas dan multifungsi alat pancing modern [45, 46].

5. **Integrasi Kolom Biner pada Dataset Utama**
   Nilai label hasil perhitungan tadi diimplementasikan ke dataset utama sebagai kolom `Empang`, `Sungai`, dan `Laut` yang bertipe biner. Proses ini dilakukan secara terprogram agar konsisten dan dapat direplikasi. Dataset akhir mencerminkan preferensi nyata pengguna dan memungkinkan sistem rekomendasi bekerja lebih presisi sesuai kondisi di lapangan [44].

6. **Validasi dan Implikasi Integrasi**

   (a) *Validasi Internal:*
   Proses validasi dilakukan dengan meninjau kembali distribusi label biner pada dataset agar selaras dengan pola data survei dan logika selektivitas alat dalam literatur. Jika terdapat anomali atau kasus dengan distribusi pilihan hampir seimbang, dilakukan diskusi tambahan atau justifikasi dari aspek teknis. Validasi data pada penelitian survei direkomendasikan melalui audit, triangulasi, dan pembandingan pola dengan referensi lain [43, 44].

   (b) *Implikasi Akademis dan Praktis:*
   Integrasi berbasis data survei ini memastikan bahwa sistem rekomendasi yang dikembangkan tidak hanya berbasis asumsi, namun berbasis data empiris dan valid. Dengan demikian, rekomendasi alat pancing yang dihasilkan lebih relevan, multifungsi, dan dapat diaplikasikan dalam praktik nyata di lapangan [45, 46, 44].

Dengan penataan ini, proses integrasi dan validasi hasil survei ke dataset utama dapat dijelaskan dan dipertanggungjawabkan secara jelas, runut, serta mengikuti kaidah ilmiah yang diakui dalam penelitian berbasis survei. Pendekatan berbasis mayoritas seperti ini juga terbuka untuk dikembangkan lebih lanjut melalui analisis threshold yang berbeda atau teknik validasi lain jika dibutuhkan pada riset lanjutan.

3.1.4 Perancangan Sistem

Sistem rekomendasi dirancang sebagai aplikasi berbasis web interaktif menggunakan framework *Streamlit*. Antarmuka pengguna terdiri dari tiga input

utama:

1. Lokasi memancing (empang, sungai, laut),
2. Anggaran total pengguna,
3. Target berat ikan (dalam kilogram).

Sistem akan memberikan satu set kombinasi alat pancing yang paling mendekati preferensi pengguna, berdasarkan data kombinasi produk yang telah dibentuk dan dianalisis sebelumnya.

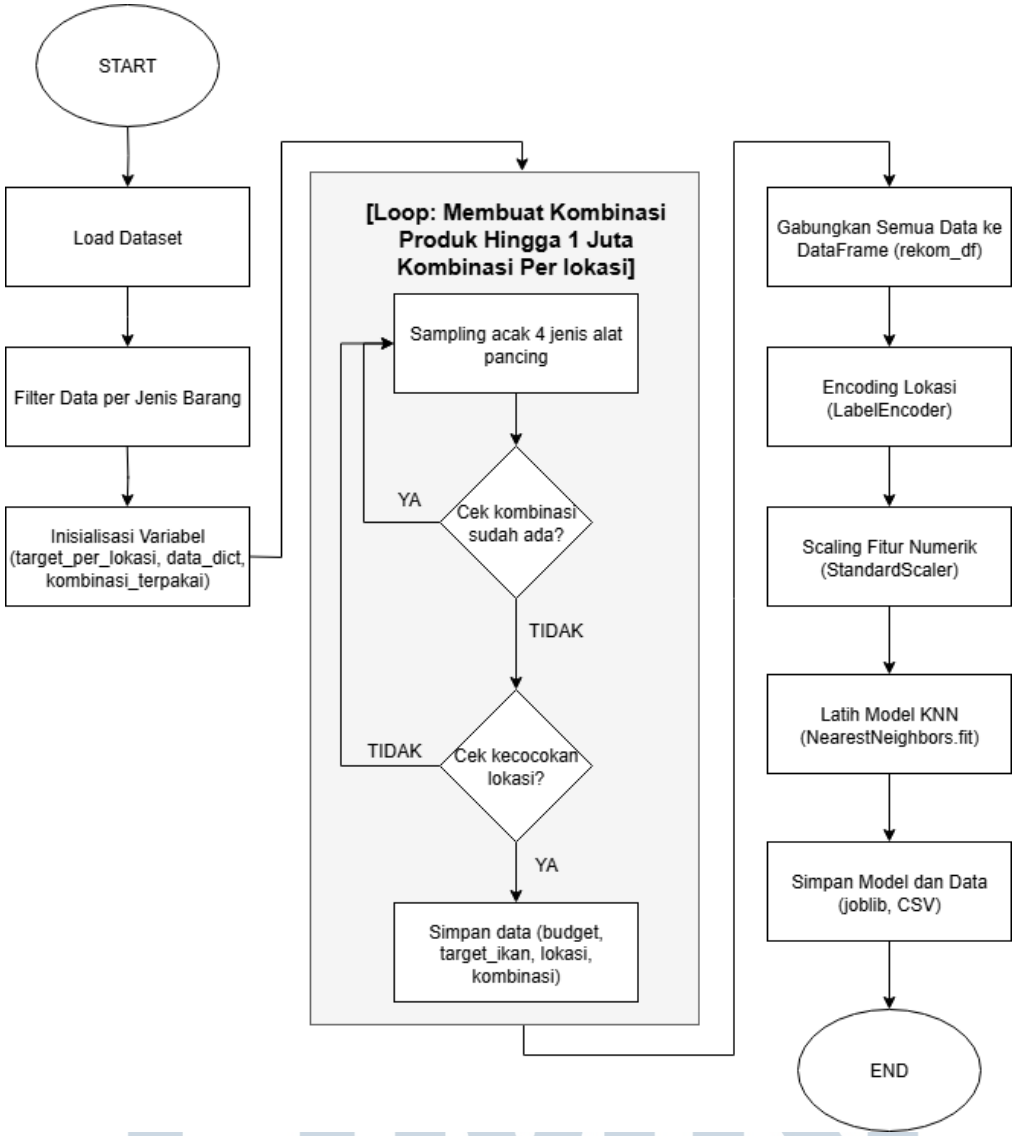
3.1.5 Implementasi Algoritma

Model *K-Nearest Neighbors* diimplementasikan menggunakan pustaka `scikit-learn`, dengan pendekatan `NearestNeighbors` tanpa supervisi. Tiga fitur numerik digunakan dalam pencarian tetangga terdekat:

- Hasil encoding lokasi memancing,
- Total anggaran pengguna,
- Target berat ikan.

Proses pelatihan model KNN dalam penelitian ini dilakukan secara bertahap menggunakan skrip Python. Alur lengkap proses pelatihan dapat dilihat pada Gambar 3.4.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.4. Flowchart Proses Pelatihan Model KNN

Berdasarkan Gambar 3.4, proses dimulai dengan pemuatan dataset dan pemisahan data per jenis barang. Selanjutnya, sistem secara acak membentuk kombinasi alat pancing hingga mencapai target, kemudian melakukan encoding, normalisasi, dan pelatihan model menggunakan algoritma *K-Nearest Neighbors*. Model yang telah dilatih kemudian disimpan untuk digunakan dalam tahap rekomendasi.

Untuk keperluan perbandingan performa, sistem juga diimplementasikan dengan:

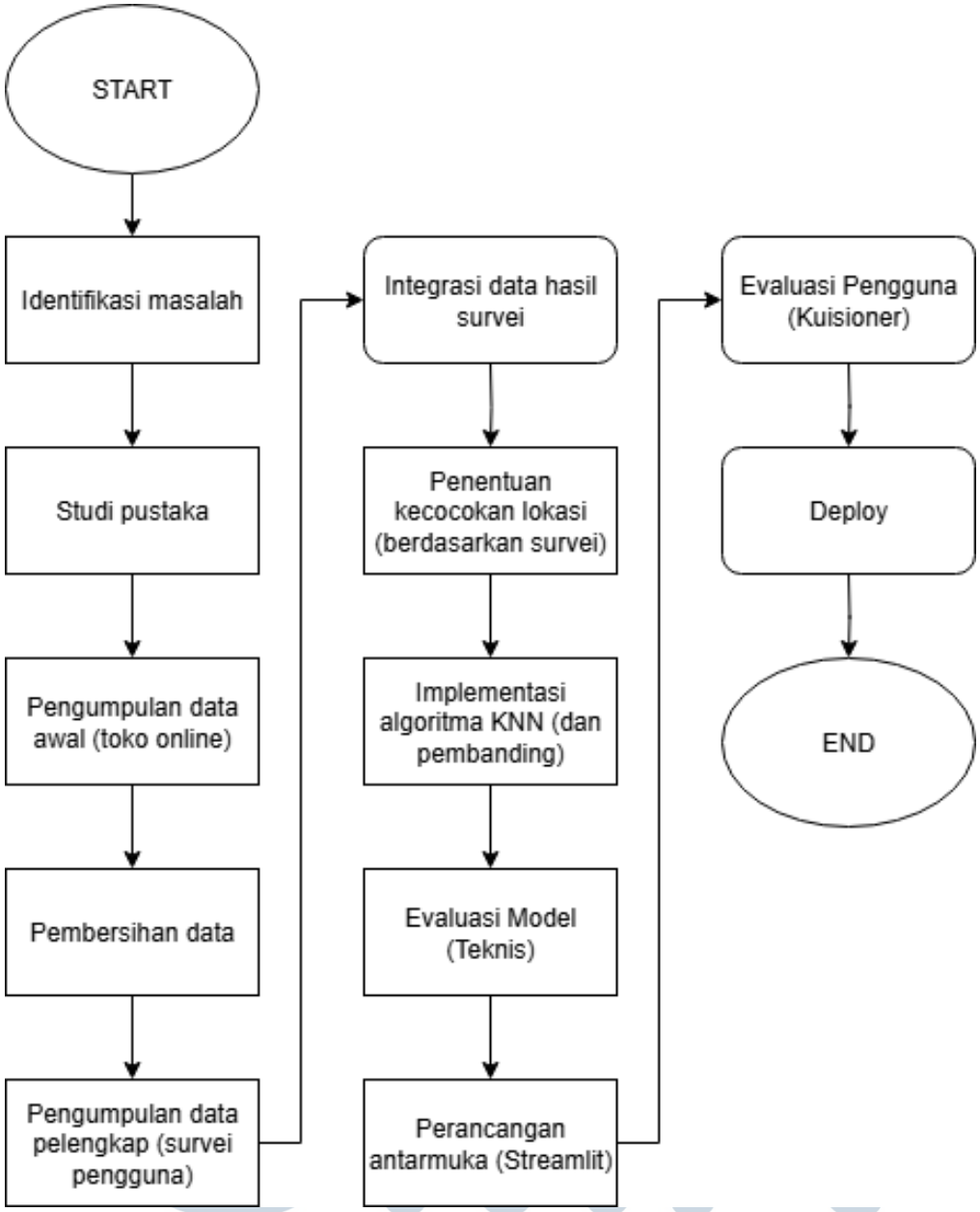
- *K-Means Clustering*, untuk mengelompokkan kombinasi alat,

- *Decision Tree Classifier*, untuk mengklasifikasikan hasil klaster berdasarkan fitur.

3.2 Kerangka Berpikir Penelitian

Kerangka berpikir penelitian ini menjelaskan alur logis yang dilakukan dalam penelitian, mulai dari identifikasi masalah hingga implementasi sistem rekomendasi. Gambar 3.5 menunjukkan kerangka berpikir penelitian yang digunakan.

UMN
UNIVERSITAS
MULTIMEDIA
NUSANTARA



Gambar 3.5. Kerangka Berpikir Penelitian

Berdasarkan Gambar 3.5, penelitian ini diawali dengan tahapan identifikasi masalah dan studi pustaka, dilanjutkan dengan pengumpulan serta pengolahan data, implementasi algoritma KNN, evaluasi teknis, perancangan antarmuka, evaluasi pengguna, hingga tahap akhir berupa deployment sistem.