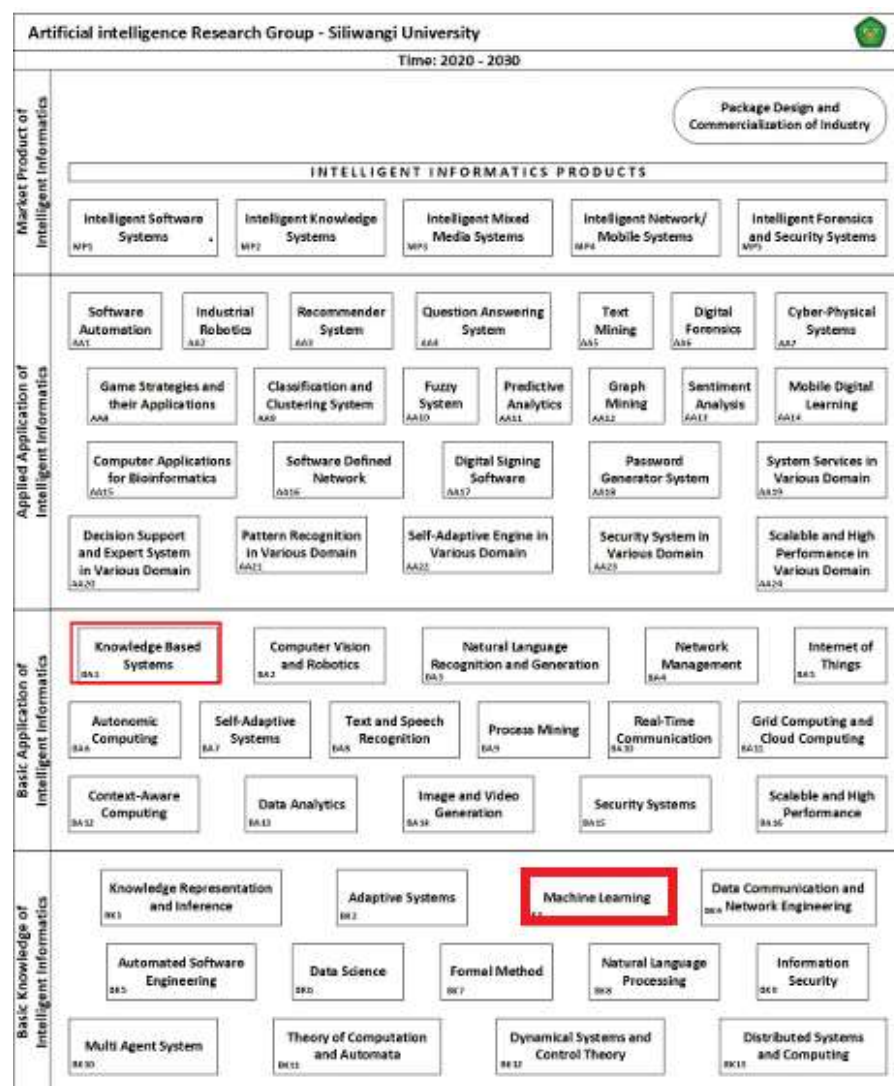


BAB III METODOLOGI PENELITIAN

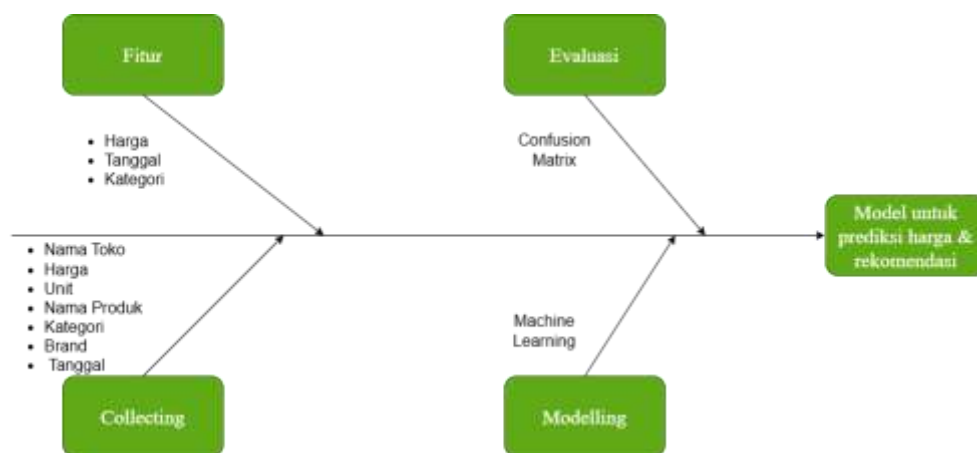
3.1. Roadmap Penelitian

Secara keseluruhan, rencana penelitian ini sejalan dengan *roadmap* di Universitas Siliwangi pada bagian *artificial intelligence*. Topik penelitian yang dipilih yaitu *machine learning* pada ranah *basic knowledge of intelligent informatics*. *Roadmap* penelitian dapat dilihat pada Gambar 3.1.



Gambar 3. 1 Roadmap Penelitian AIS (Universitas Siliwangi)

Gambar 3.1 pemilihan topik *machine learning* pada ranah *basic knowledge of intelligence informatics* sehingga penelitian ini memiliki tujuan pengujian terhadap algoritma yang digunakan untuk menjadi dasar pengetahuan. Penelitian ini, melakukan pemodelan yang berorientasi dalam memprediksi harga produk dan merekomendasikan produk untuk meningkatkan profitabilitas pemasaran. Berikut Gambar 3.2. Merupakan *fishbone* penelitian sebagai dari *roadmap*.



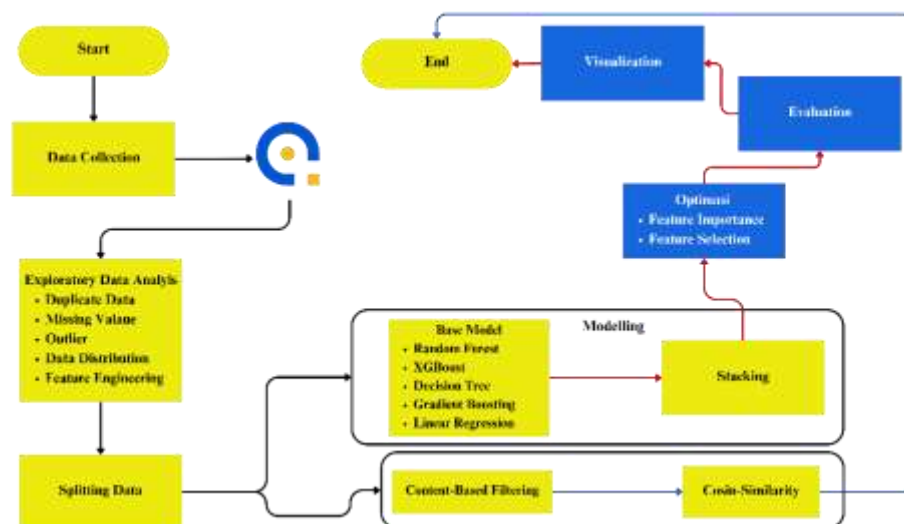
Gambar 3. 2 Fishbone Penelitian

Gambar 3.2, diagram *fishbone* yang mempresentasikan penelitian, dengan tujuan untuk pengujian terhadap model atau algoritma yang digunakan untuk memprediksi harga produk dan merekomendasikan produk. Pengambilan data yang digunakan memiliki informasi mengenai nama toko, harga, unit, nama produk, kategori, *brand*, dan tanggal. Kemudian fitur atau variabel target yang digunakan yaitu harga, dan variabel yang mempengaruhinya yang digunakan kategori dan tanggal. Selanjutnya dalam implementasi modelnya menggunakan pendekatan *machine learning* yang dievaluasi dengan menggunakan *confusion matrix* untuk

mengukur seberapa akurat dari model yang digunakan dalam memprediksi harga dan rekomendasi produk.

3.2. Metode Penelitian

Metode penelitian yang digunakan dalam penelitian "Prediksi Tren Harga dan Rekomendasi Produk Menggunakan Pendekatan *Machine Learning*" adalah CRISP-DM (Cross-Industry Standard Processing for Data Mining). Pada Gambar 3.3. CRISP-DM merupakan metodologi standar untuk mengimplementasikan *machine learning*.



Gambar 3. 3 Metode Penelitian

Penjelasan detail dari setiap tahapannya yaitu sebagai berikut:

1. Data Collection

Tahap ini memfokuskan pada identifikasi, pengumpulan, dan analisis kumpulan data yang dapat membantu mencapai tujuan proyek (Pratama et al., 2023). Pada penelitian ini dataset yang digunakan berasal dari Mitra Startup

Campus. Berikut ilustrasi pada tahapan *data collection*, dapat dilihat pada Gambar

3.4.

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 2367510 entries, 0 to 2367509
Data columns (total 8 columns):
#   Column      Dtype
---  -
0   nama_toko    object
1   harga_pound  float64
2   harga_per_unit float64
3   unit         object
4   nama         object
5   kategori     object
6   brand_sendiri bool
7   tanggal      object
dtypes: bool(1), float64(2), object(5)
memory usage: 128.7+ MB
```

Gambar 3. 4 Dataset

2. Exploratory Data Analysis

Secara umum *Exploratory Data Analysis* (EDA) yaitu mempersiapkan kumpulan data akhir untuk melakukan implementasi model (Rizki & Mulyawan, 2023). Terdapat beberapa tahapan dalam EDA, diantaranya:

- a) *Duplicate data*: Mengecek data yang mempunyai nilai yang sama, dan menangani nilai yang sama tersebut dengan cara melakukan *drop* atau menghapusnya (Cahyani Putri & Arianto, 2024). Gambar 3.5. Merupakan tahapan *duplicate data*, pada dataset yang digunakan mempunyai 445 *duplicate*, Gambar 3.6, menunjukkan dataset yang sudah tidak mempunyai *duplicate data*.

	nama_toko	harga_pound	harga_per_unit	unit	nama	kategori	brand_sendiri	tanggal
0	BritishMart	1.35	1.35	l	Alpro Oat Long Life Drink	drinks	False	2021-01-01
1	BritishMart	2.10	2.10	l	Alpro Almond Chilled Drink	drinks	False	2021-01-01
2	BritishMart	1.00	0.70	l	Volvic Touch of Fruit Strawberry...	drinks	False	2021-01-01
3	BritishMart	2.10	2.10	l	Alpro Soya Chilled Drink	drinks	False	2021-01-01
4	BritishMart	1.00	0.70	l	Volvic Touch of Fruit Lemon & Lim...	drinks	False	2021-01-01
...								
2367505	UKMart	2.70	14.21	kg	Jarlsberg Original 190g	fresh_food	False	2022-04-30
2367506	UKMart	3.00	5.00	kg	Sainsbury's Moroccan Spiced Chicken & Chickpea...	fresh_food	True	2022-05-01
2367507	UKMart	1.25	5.50	kg	Ginsters Cornish Pasty, 227g	fresh_food	False	2022-05-02
2367508	UKMart	3.00	30.00	kg	Sainsbury's British Corned Beef Slices, Taste...	fresh_food	True	2022-05-03
2367509	UKMart	5.50	19.64	kg	New England Seafood Wild Smoked Haddock Fillet...	fresh_food	False	2022-05-04

2367510 rows x 8 columns

Gambar 3. 5 Dataset yang Memiliki Duplicate Data

	nama_toko	harga_pound	harga_per_unit	unit	nama	kategori	brand_sendiri	tanggal
0	BritishMart	1.35	1.35	l	Alpro Oat Long Life Drink	drinks	False	2021-01-01
1	BritishMart	2.10	2.10	l	Alpro Almond Chilled Drink	drinks	False	2021-01-01
2	BritishMart	1.00	0.70	l	Volvic Touch of Fruit Strawberry...	drinks	False	2021-01-01
3	BritishMart	2.10	2.10	l	Alpro Soya Chilled Drink	drinks	False	2021-01-01
4	BritishMart	1.00	0.70	l	Volvic Touch of Fruit Lemon & Lim...	drinks	False	2021-01-01
...								
2367505	UKMart	2.70	14.21	kg	Jarlsberg Original 190g	fresh_food	False	2022-04-30
2367506	UKMart	3.00	5.00	kg	Sainsbury's Moroccan Spiced Chicken & Chickpea...	fresh_food	True	2022-05-01
2367507	UKMart	1.25	5.50	kg	Ginsters Cornish Pasty, 227g	fresh_food	False	2022-05-02
2367508	UKMart	3.00	30.00	kg	Sainsbury's British Corned Beef Slices, Taste...	fresh_food	True	2022-05-03
2367509	UKMart	5.50	19.64	kg	New England Seafood Wild Smoked Haddock Fillet...	fresh_food	False	2022-05-04

2367505 rows x 8 columns

Gambar 3. 6 Dataset yang Sudah Tidak Mempunyai Duplicate Data

- b) *Missing Value*: Menangani nilai yang kosong dalam dataset, dengan cara menghapus data dari nilai yang kosong tersebut, atau imputasi data dengan mengisi nilai yang kosong (Prasetya et al., 2023). Gambar 3.7, menunjukkan dataset yang digunakan terdapat *missing value* sebanyak 129 pada variabel harga_per_unit, dan unit. Pada Gambar 8, dataset yang sudah dilakukan proses penghapusan *missing value*.

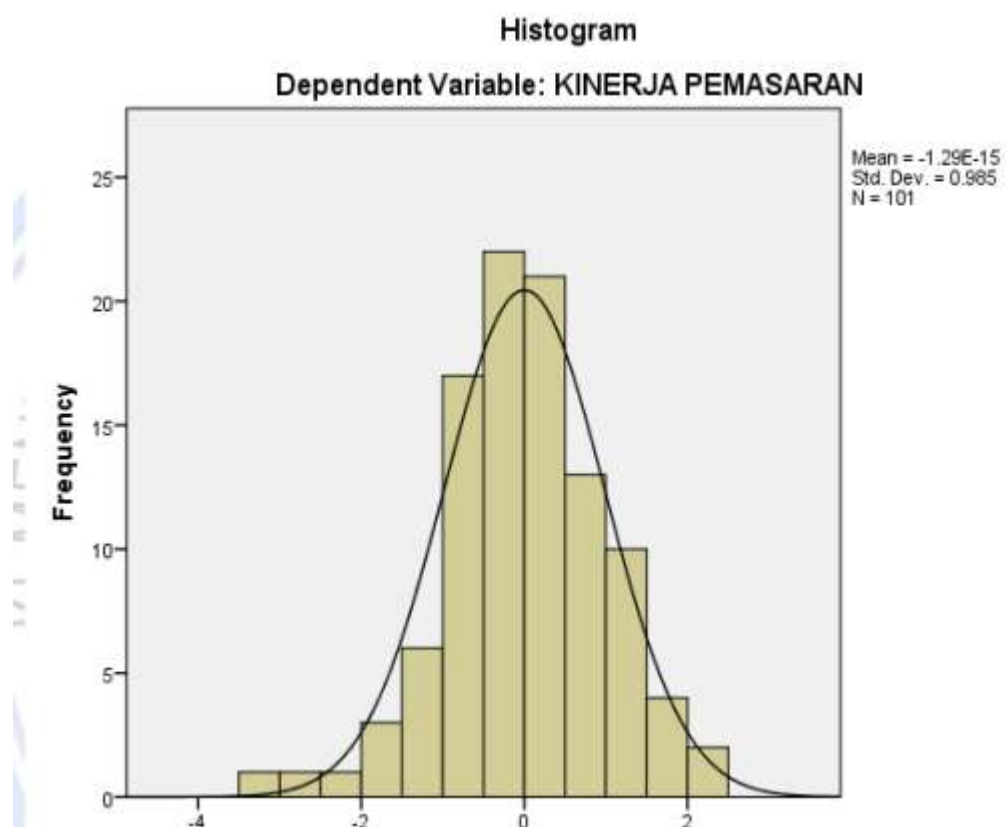
	Column Name	Missing Values	Percentage Missing (%)
nama_toko	nama_toko	0	0.00000
harga_pound	harga_pound	0	0.00000
harga_per_unit	harga_per_unit	129	0.00545
unit	unit	129	0.00545
nama	nama	0	0.00000
kategori	kategori	0	0.00000
brand_sendiri	brand_sendiri	0	0.00000
tanggal	tanggal	0	0.00000

Gambar 3. 7 Dataset yang Memiliki Missing Value

	Column Name	Missing Values	Percentage Missing (%)
nama_toko	nama_toko	0	0.0
harga_pound	harga_pound	0	0.0
harga_per_unit	harga_per_unit	0	0.0
unit	unit	0	0.0
nama	nama	0	0.0
kategori	kategori	0	0.0
brand_sendiri	brand_sendiri	0	0.0
tanggal	tanggal	0	0.0

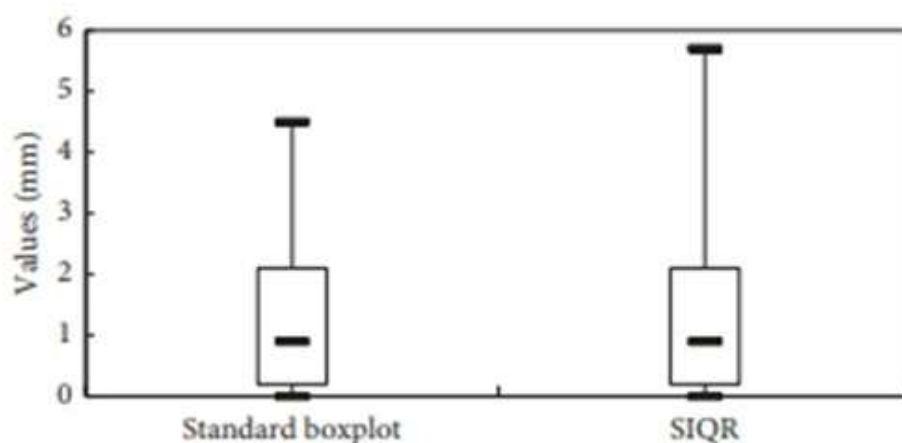
Gambar 3. 8 Dataset yang Tidak Memiliki *Missing Value*

- c) *Data Distribution*: Menganalisis nilai-nilai pada dataset, dengan menggunakan *plot histogram*, untuk memudahkan dalam memahami data (Furqoni Yudhistira et al., 2023). Dapat dilihat pada Gambar 3.9 menunjukkan distribusi salah satu variabel yang ada pada dataset menggunakan *plot histogram*.



Gambar 3. 9 Distribusi Data Menggunakan *Histogram*

- d) *Outlier*: Data yang memiliki nilai yang sangat berbeda dari sebagian besar data lainnya dalam suatu dataset, untuk menangani nilai tersebut dapat dilakukan dengan cara menghapusnya (Ribeiro et al., 2020). Dapat dilihat pada Gambar 3.10, menunjukkan aliran data untuk melihat nilai *outlier*.



Gambar 3. 10 Dataset yang Mempunyai Nilai *Outlier*

- e) *Feature Engineering*: Sebelum mengimplementasikan pemodelan, untuk meningkatkan akurasi dari prediksi dilakukannya tahapan *feature engineering* dengan mengubah atau menambahkan variabel data (Faiz et al., 2022). Pada Gambar 3.11, merupakan teknik *feature engineering*, untuk menambahkan variabel baru.

	nama_toko	total_products
0	0	965071
1	1	240530

Gambar 3. 11 *Feature Engineering*

3. *Splitting Data*

Setelah melalui tahapan *data preparation*, langkah berikutnya adalah melakukan implementasi model. Sebagai tahapan awal *modelling*, membagi dataset

menjadi data *train* dan data *test*, dengan perbandingan 80:20% (Hidayah et al., 2024).

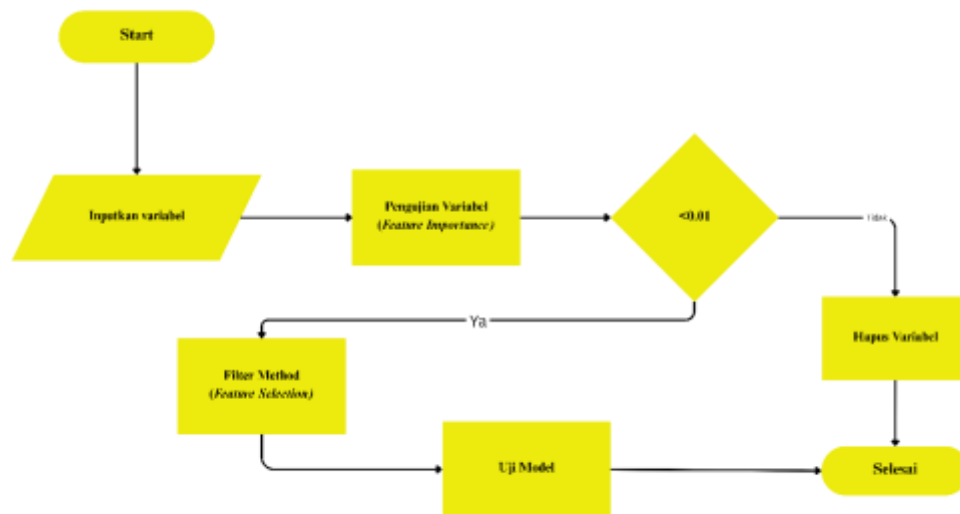
4. *Modelling*

Mengimplementasikan model pada data *train* dan *test* yang telah dilakukan pada tahapan *splitting data* sebelumnya. Pada penelitian ini, model yang digunakan yaitu *Random Forest*, *XGBoost*, *Gradient Boosting*, *Decision Tree*, dan *Regresi Linear* untuk mencari model yang mempunyai akurasi awal yang cukup tinggi. Kemudian dilakukan proses *model selection* untuk mengetahui model terbaik dengan membandingkan akurasi model yang telah di implementasikan (Adıgüzel, 2023; Priyatno et al., 2024).

Hasil dari pemodelan dan rekomendasi membantu toko acuan untuk memprediksi harga produk berdasarkan tren harga toko pembanding, agar setiap harga produk mampu bersaing kompetitif dengan toko pembanding dan tidak melebihi harga dari toko pembanding agar menarik minat pembeli terhadap produk di toko acuan sehingga keuntungan mampu di optimalkan.

5. *Optimasi*

Setelah mendapatkan model yang mempunyai akurasi awal cukup tinggi, dilakukan tahapan optimasi untuk meningkatkan performa model (Setia Budi et al., 2024).



Gambar 3. 12 Proses Optimasi Model

Gambar 3.12, merupakan proses optimasi dilakukan pengujian variabel dengan menggunakan *feature importance*, bilamana nilai batas setiap variabel > 0.01 disarankan untuk dihapus karena tidak ada pengaruhnya ke nilai akurasi (Budiman et al., 2021). Kemudian setelah melakukan pengujian variabel selanjutnya dilakukannya *feature selection* dengan menggunakan *filter method* untuk menggunakan variabel yang mempengaruhi terhadap hasil akurasi awal dari proses *feature importance* (Rahman et al., 2024).

6. Evaluation

Pengukuran seberapa efektif metode atau algoritma yang digunakan pada tahapan *modelling*, yaitu diukur dengan evaluasi hasil model tersebut dengan menggunakan *matrix evaluation*, R^2 untuk menguji seberapa efektif model yang digunakan, MAPE digunakan untuk mengukur kesalahan dalam peramalan, dan RMSE untuk mengevaluasi seberapa baik model (Rosita & Moonlight, 2024; Triya et al., 2024).

7. Visualisasi

Setelah mengukur seberapa efektif model yang digunakan, langkah terakhir yaitu mempublikasi hasil untuk memudahkan publik dalam memahami hasil dari pemodelan tersebut. Pada penelitian ini, untuk publikasi terhadap hasil dari *modelling* tersebut, dengan menggunakan *platform Tableau* kepada supermarket agar memudahkan dalam menyusun strategi pemasarannya.