 E-ISSN : 2722-0346

PREDIKSI HARGA *SMARTPHONE* BERDASARKAN SPESIFIKASI MENGGUNAKAN *K-NEAREST NEIGHBORS*

Dea Fitra Ningrum[-1], Shabrina Putri Ramadhani[-2], Iman Paryudi[-3], Ionia Veritawati[-4], Sri Rezeki Candra Nursari[-5]
Program Studi Teknik Informatika
Universitas Pancasila, Jakarta, Indonesia
Jl. Lenteng Agung Raya No.56 Jakarta, Indonesia 12630
4520210023@univpancasila.ac.id[-1]
4520210024@univpancasila.ac.id[-2]
iman.paryudi@univpancasila.ac.id[-3]
ionia.doc@gmail.com[-4]
dosen.it2011@gmail.com[-5]

***Abstract* — Di era teknologi informasi yang terus berkembang, pasar ponsel pintar menjadi salah satu pasar konsumen yang paling dinamis dan beragam. Pembeli seringkali dihadapkan pada banyak pilihan dalam memilih smartphone baru yang sesuai dengan kebutuhan dan budgetnya. Penelitian ini bertujuan untuk memprediksi harga smartphone berdasarkan spesifikasi. Metodologi yang digunakan adalah algoritma K-Nearest Neighbor dengan menggunakan Euclidean distance, membagi dataset menjadi 70% data latih dan 30% data uji. Model ini telah diuji sebanyak 2 kali, pengujian pertama menggunakan k sebesar 1 dan menghasilkan akurasi sebesar 57%, sedangkan pengujian kedua menggunakan nilai k sebesar 3 dan memperoleh akurasi sebesar 65%.**

Keywords—Algoritma K-Nearest Neighbor, Smartphone, Prediksi

I. PENDAHULUAN

Smartphone menjadi salah satu kebutuhan masyarakat modern saat ini untuk menunjang aktivitasnya. Perusahaan elektronik memperhatikan kebutuhan ini, itulah sebabnya banyak merek ponsel pintar bermunculan. Semakin majunya teknologi informasi dan meningkatnya taraf hidup masyarakat menyebabkan meningkatnya tuntutan masyarakat terhadap kualitas pelayanan dan produk yang digunakan. Kebutuhan untuk menggunakan smartphone sudah menjadi kebutuhan hidup yang dianggap penting bagi sebagian masyarakat modern saat ini. Fenomena tersebut mendukung munculnya banyak smartphone yang menawarkan produknya untuk memenuhi kebutuhan teknologi komunikasi masyarakat.

Banyak merek smartphone yang saling berlomba-lomba untuk meningkatkan spesifikasi smartphone tersebut untuk menarik minat para pembeli seperti merek Infinix, Motorola, Samsung, Redmi, Vivo, Realme dan masih banyak lagi. Biasanya pembeli tertarik untuk membeli smartphone dengan kamera depan dan belakang, namun ada juga yang tertarik dengan RAM dan prosesor. Oleh karena itu, pada penelitian ini kami melakukan analisis untuk memprediksi harga smartphone berdasarkan spesifikasi.

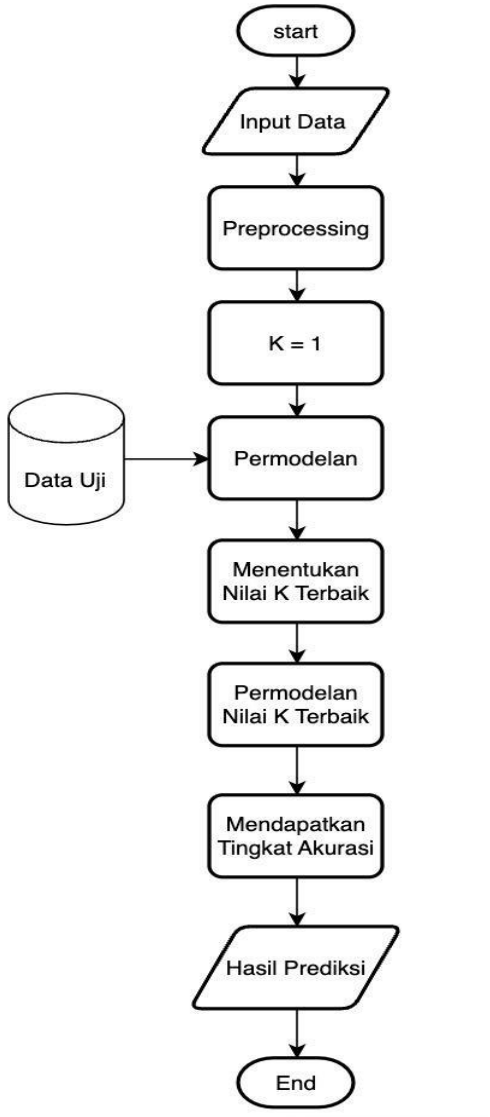
II. STUDI LITERATUR

Salah satu teknik yang bisa diterapkan dalam data mining adalah teknik klasifikasi.Teknik klasifikasi adalah proses menemukan definisi karakteristik serupa dalam suatu kelompok atau kelas. Klasifikasi merupakan salah satu metode yang paling umum digunakan dalam data mining, yang bertujuan untuk memperkirakan kelas suatu objek yang labelnya tidak diketahui. Ada beberapa metode klasifikasi yang umum digunakan, salah satunya adalah *K-Nearest Neighbor*.

Algoritma *K-Nearest Neighbor* (KNN) merupakan algoritma pembelajaran mesin non-parametrik dan l*azy learning*. Metode non-parametrik berarti tidak membuat asumsi tentang distribusi data yang mendasarinya. Dengan kata lain, tidak ada jumlah parameter atau estimasi parameter yang tetap dalam model, tidak peduli seberapa kecil atau besar datanya. Algoritma KNN mengasumsikan bahwa ada sesuatu yang serupa di dekatnya atau di dekatnya. Artinya, data dengan tren serupa akan berdekatan satu sama lain. KNN menggunakan seluruh data yang tersedia dan mengklasifikasikan data atau kasus baru berdasarkan ukuran kemiripan atau fungsi jarak. Data baru kemudian ditugaskan ke kelas yang memiliki sebagian besar data tetangganya

 E-ISSN : 2722-0346

III. METODOLOGI PENELITIAN



Gambar. 1 Metodologi Penelitian

Pada penelitian ini, data yang digunakan merupakan data sekunder, dimana data sekunder adalah jenis data yang diperoleh dari sumber lain, bukan dari sumber asli atau pertama. Data yang digunakan bersumber dari kaggle.com dan memiliki beberapa spesifikasi seperti Model, Colour, Memory, RAM, Battery, Rear Camera, Front Camera, AI Lens, Mobile Height, Processor dan Prize. Penelitian ini menggunakan 642 data dan digunakan keseluruhan datanya untuk data testing dan data training.

Dalam penelitian ini, digunakan salah satu algoritma dalam data mining untuk mengolah data dengan tujuan menghasilkan prediksi harga smartphone berdasarkan spesifikasi yang dapat bermanfaat bagi perusahaan atau pembeli, serta memperoleh pengetahuan dari data yang tersedia. Algoritma yang digunakan adalah Algoritma K-Nearest Neighbor, dimana algoritma ini dapat memprediksi harga produk dengan cara bekerja berdasarkan prinsip bahwa objek yang mirip cenderung berada dalam jarak yang dekat satu sama lain yang kemudian data dilatih dan diuji untuk menghasilkan nilai prediksi.

Langkah pertama dalam metodologi ini adalah memasukkan data yang akan digunakan. Proses preprocessing data melibatkan serangkaian tahapan seperti pembersihan data, transformasi data, normalisasi, dan langkah-langkah lain untuk mempersiapkan data agar sesuai dengan algoritma K-Nearest Neighbors (KNN). Tujuannya adalah memastikan bahwa data dalam kondisi yang optimal untuk digunakan dalam pemodelan.

Langkah selanjutnya yaitu memilih nilai k yang akan digunakan untuk model KNN. Nilai k menunjukkan jumlah tetangga terdekat yang akan dipertimbangkan dalam proses klasifikasi. Pemodelan dilakukan dengan menggunakan data yang telah diproses dan dengan nilai k yang telah ditentukan. Kemudian, melakukan eksperimen dengan berbagai nilai k untuk menemukan nilai k terbaik yang menghasilkan hasil yang optimal. Proses pemilihan nilai k terbaik dengan menggunakan elbow method untuk menghitung Mean Squared Error (MSE) dari model K-Nearest Neighbors Regressor pada berbagai nilai k yang berbeda.. Setelah menemukan nilai k terbaik, model KNN dilatih kembali dengan nilai k tersebut.

Setelah melalui tahapan tersebut, model KNN yang telah dilatih akan digunakan untuk melakukan prediksi berdasarkan tetangga terdekat yang dipilih sesuai dengan nilai k yang telah ditentukan. Hasil prediksi ini diukur dengan menghitung tingkat akurasi model, yang mencerminkan sejauh mana model dapat memprediksi dengan benar. Setelah itu, model KNN yang telah terlatih akan digunakan untuk membuat prediksi pada data baru atau data pengujian.

IV. HASIL DAN PEMBHASAN

Data yang digunakan adalah data yang bersumber dari *kaggle*, yaitu menggunakan dataset *Mobile Price Prediction*, dimana data memiliki 541 baris dan 11 kolom. Atribut yang tersedia pada dataset tersebut diantaranya *Model, Colour, Memory, RAM, Battery, Rear Camera, Front Camera, AI Lens, Mobile Height, Processor, Price.*

TABLE I
DATASET MOBILE PREDICTION

E-ISSN : 2722-0346

index	Brand	Colour	Memory	RAM	Battery	primary camera	secondary camera	AI Lens	Mobile Height	Processor_
0	Infinix	Night Black	64	4	6000	13	5	1	0,719444444	Unisoc Spreadtrum SC9863A1
1	Infinix	Azure Blue	64	4	6000	13	5	1	0,719444444	Unisoc Spreadtrum SC9863A1
2	MOTOROLA	Mineral Gray	128	8	5000	50	16	0	0,711111111	Qualcomm Snapdragon 680
3	POCO	Royal Blue	32	2	5000	8	5	0	16.56	Mediatek Helio A22
4	Infinix	Marigold	128	8	5000	50	5	1	0,719444444	G37
5	Infinix	Glacier Blue	128	8	5000	50	5	1	0,719444444	G37

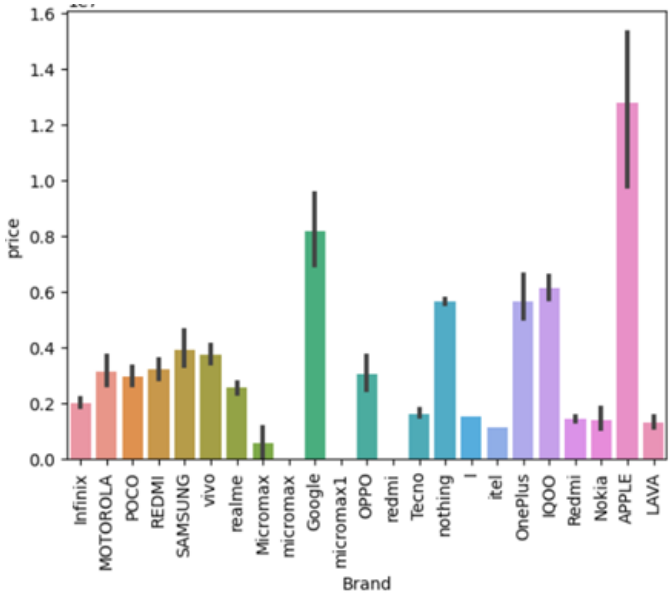
Dari dataset yang diperoleh, akan dilakukan proses pengecekan data apakah terdapat data yang duplikat maupun data kosong (null). Setelah dilakukan pengecekan, dataset diatas tidak memiliki data ganda maupun data kosong. Maka akan dilanjutkan ke tahap berikutnya yaitu melihat nilai yang kosong (null).

```
df.isnull().sum()
```

```
Brand              0
Colour             0
Memory             0
RAM                0
Battery_           0
primary camera     0
secondary camera   0
AI Lens            0
Mobile Height      0
Processor_         0
price              0
dtype: int64
```

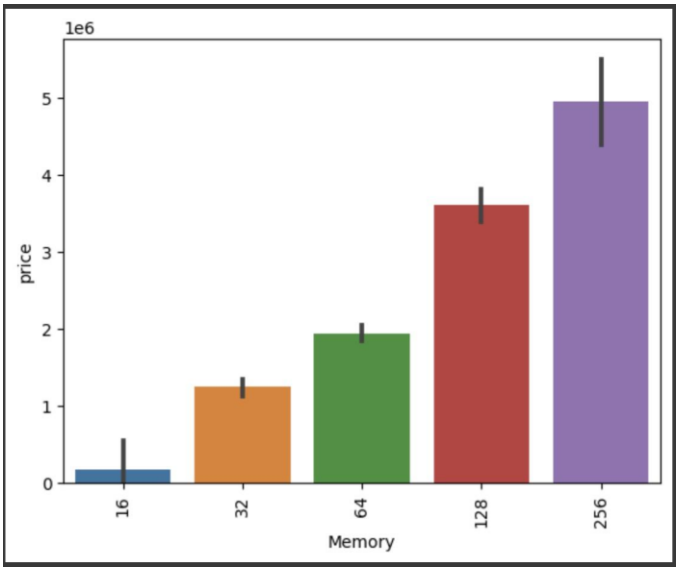
Gambar. 2 Melihat nilai null

Dataset diatas tidak memiliki data ganda maupun data kosong. Sebelum memasuki tahap prediksi harga laptop, akan dilakukan tahap analisis spesifikasi terhadap harga *smartphone*.



Gambar 3. Perbandingan harga antar *brand*

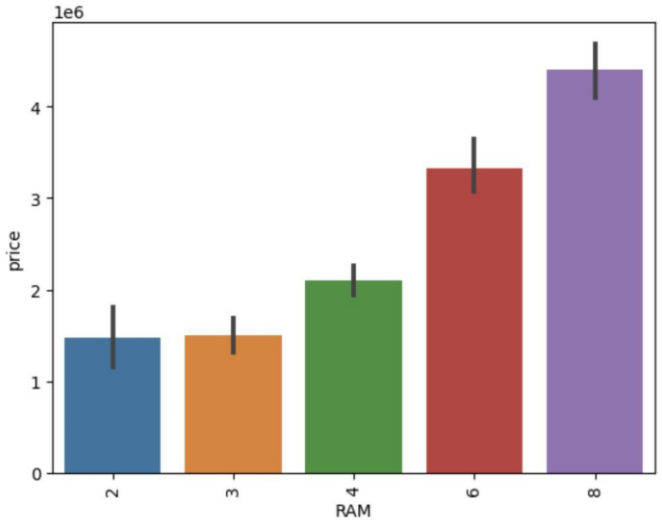
Terdapat berbagai macam jenis brand yang ada pada dataset. Untuk kisaran harga, Brand dari APPLE memiliki harga yang signifikan tinggi dibanding harga dari *smartphone* lainnya.



Gambar 4. harga *Memory*

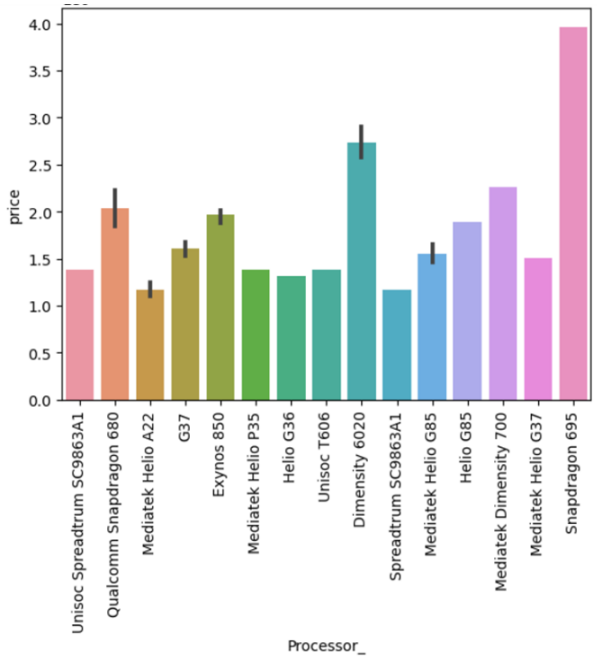
Berdasarkan dataset tersebut, smartphone dengan kapasitas penyimpanan 256GB memiliki harga yang cukup tinggi yaitu berkisar 4-5 juta rupiah.

E-ISSN : 2722-0346



Gambar 5. harga RAM

Rata-rata RAM yang terdapat pada dataset tersebut sebesar 3GB dengan kisaran harga 1-3 Juta Rupiah. RAM sebesar 8GB memiliki harga yang cukup tinggi, yaitu berada di rentang 4-5 Juta Rupiah.



Gambar 6. Perbandingan *Processor*

Terdapat beberapa *processor* yang digunakan pada *smartphone* di dataset tersebut. Smartphone dengan *processor* Snapdragon 695 memiliki harga yang paling tinggi dibandingkan *smartphone* dengan *processor* lain. Rata-rata *processor* yang digunakan yaitu processor Dimensity 6020, Qualcomm snapdragon 680, dan exynos dengan kisaran harga 1,5-3 juta rupiah.

Setelah melakukan analisis spesifikasi *smartphone* dengan harga, dilakukan pemrosesan berupa pengkodean label (label encoding) untuk mengubah data atribut dengan tipe data object menjadi data numerik. Tabel 2 merupakan hasil dari proses pengkodean label menggunakan python.

TABLE 2
HASIL SETELAH DIKONVERSI

index	Brand	Colour	Memory	RAM	Battery	primary camera	secondary camera	AI Lens	Mobile Height	Processor_	price
0	4	159	64	4	6000	13	5	1	0,719444444	113	1377394
1	4	20	64	4	6000	13	5	1	0,719444444	113	1377394
2	6	149	128	8	5000	50	16	0	0,711111111	75	2264331
3	11	201	32	2	5000	8	5	0	16.56	56	1066023
4	4	130	128	8	5000	50	5	1	0,719444444	14	1698201

Setelah melakukan proses konversi atribut dengan tipe data object menjadi data numerik, dilakukan proses normalisasi untuk mengubah rentang atau skala data sehingga semua atribut memiliki pengaruh yang seimbang pada klasifikasi dan memastikan bahwa tidak ada fitur yang mendominasi perhitungan jarak hanya karena besarnya skala nilai yang dimilikinya. Tabel 3 merupakan hasil dari proses normalisasi menggunakan python.

TABLE 3
HASIL SETELAH PROSES NORMALISASI

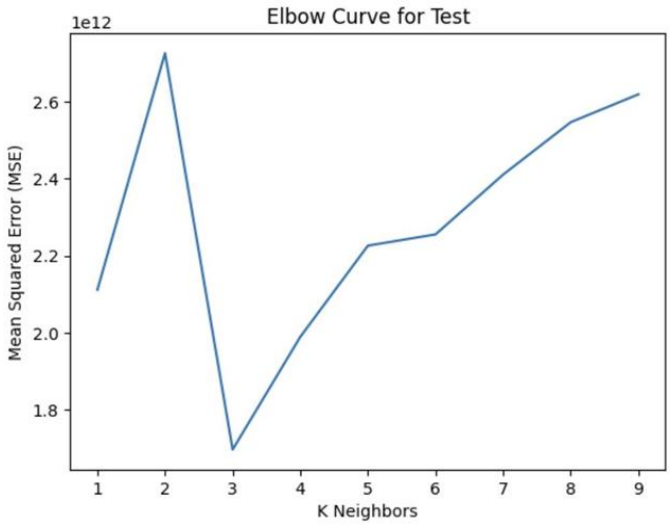
Index	Brand	Colour	Memory	RAM	Battery	primary camera	secondary camera	AI Lens	Mobile Height	Processor_
0	0,18	0,58	0,20	0,33	0,84	0,07	0,08	1,00	0,33	0,93
1	0,18	0,07	0,20	0,33	0,84	0,07	0,08	1,00	0,33	0,93
2	0,27	0,54	0,47	1,00	0,68	0,17	0,27	0,00	0,32	0,61
3	0,50	0,73	0,07	0,00	0,68	0,04	0,08	0,00	0,32	0,46
4	0,18	0,47	0,47	1,00	0,68	0,17	0,08	1,00	0,33	0,11

Selanjutnya akan menguji pengaruh spesifikasi terhadap harga *smartphone* dengan melakukan split data testing dan data training. train_x akan berisi subset pelatihan dari fitur, test_x akan berisi subset pengujian dari fitur, train_y akan berisi

 E-ISSN : 2722-0346

subset pelatihan dari variabel target, dan test_y akan berisi subset pengujian dari variabel target. Dataset akan dibagi sehingga 30% dari data akan menjadi data uji, sementara 70% akan menjadi data latih.

Setelah melakukan split dataset, selanjutnya akan melakukan perhitungan prediksi harga *smartphone* berdasarkan spesifikasi menggunakan metode *K-Nearest Neighbor*. Nilai k yang akan digunakan yaitu k=1, dengan menghasilkan tingkat akurasi sebesar 57%.Untuk meningkatkan tingkat akurasi, pemilihan nilai k akan menggunakan *elbow method* untuk membantu menentukan jumlah tetangga terdekat k yang optimal dalam algoritma *K-Nearest Neighbors* (KNN).



Gambar 7. Melihat Nilai K Terbaik

Grafik diatas menandakan bahwa dalam pemilihan nilai k dengan metode elbow, nilai k yang paling optimal adalah 3. Ini berdasarkan analisis grafik elbow method yang menunjukkan bahwa pada nilai k=3, MSE mencapai titik terendah atau paling optimal. MSE yang lebih rendah menunjukkan bahwa model KNN dengan k=3 memberikan tingkat kesalahan yang lebih kecil ketika memprediksi data pengujian, yang berarti model KNN memiliki kinerja yang lebih baik dalam memprediksi data. Tingkat akurasi yang diperoleh dari perhitungan dengan menggunakan metode K-Nearest Neighbor dengan k=3 cukup baik yaitu sebesar 65%. Tabel 4 merupakan hasil prediksi yang diperoleh dari perhitungan menggunakan metode K-Nearest Neighbor.

TABLE 4
PERBANDINGAN NILAI AKTUAL DAN NILAI PREDIKSI

INDEX	ACTUAL VALUE	PREDICTED VALUE
326	6227241	5661111
287	3207881	3583603
209	5661111	5661111
365	5260669	5340116
196	3207881	3207881
252	2264331	2264331
430	2169976	2830461
459	3113526	3113526

V. KESIMPULAN

Dataset spesifikasi ponsel dari Kaggle dengan 541 baris dan 11 kolom telah dibersihkan dari data duplikat dan kosong. Dilakukan pengkodean label untuk mengubah data kategori menjadi numerik dan normalisasi data untuk keseimbangan atribut. Dataset dibagi menjadi data pelatihan (70%) dan pengujian (30%). Model K-Nearest Neighbor digunakan untuk prediksi harga awalnya dengan k=1, menghasilkan akurasi 57%. Penggunaan metode elbow menemukan nilai k optimal sebesar 3, berdasarkan penurunan MSE. Dengan k=3, akurasi meningkat menjadi 65%, meningkatkan kinerja prediksi harga ponsel berdasarkan spesifikasi.

VI. DAFTAR PUSTAKA

[1] Y. Wiyli, "Algoritma K-Nearest Neighbour untuk Memprediksi Harga Jual Tanah," Jurnal Matematika,Statistika, & Komputasi, Vol 9 No.1, 2012.

[2] K. Dartia, and I.N. Suarmanayasa, "PENGARUH KUALITAS PRODUK, HARGA, DAN DESAIN PRODUK TERHADAP KEPUTUSAN PEMBELIAN SMARTPHONE MEREK OPPO," Jurnal Manajemen, Vol.8 No.1 April 2022.

[3] R. I Made, and V. Ayu, "Impelementasi Algoritma KNN untuk Memprediksi Permorma Siswa Sekolah," Jurnal Nasional Teknologi Informasi dan Aplikasinya, Vol. 1 No. 3 Mei 2023.

[4] Y. Mukhlisin, M. Imrona, dan D. Triantoro. 2020. "Prediksi Harga Beras Premium dengan Metode Algoritma K-Nearest Neighbor". e-Proceeding of Engineering. Vol.7, No.1 April 2020.

[5] Y. Mukhlisin, M. Imrona, dan D. Triantoro. 2020. "Prediksi Harga Beras Premium dengan Metode Algoritma K-Nearest Neighbor". e-Proceeding of Engineering. Vol.7, No.1 April 2020.

[6] F. Ristianto, Nurmalasari, dan A. Yoraeni. 2021. Implementasi Metode Naive Bayes untuk Prediksi Harga Emas. Computer Science (CO-SCIENCE). Vol. 1 No, 1 Januari 2021.

[7] D. Fitriati, "Implementasi Data Mining untuk Menentukan Kombinasi Media Promosi Barang Berdasarkan Perilaku Pembelian Pelanggan Menggunakan Algoritma Apriori," Annual Research Seminar, Vol 2 No.1, 2016.