

Nama : I Gede Teguh Permana

Nim : 2008561015

1. Data Collecting

Pada progress minggu ini, saya ada proses scraping kembali yaitu menambahkan data hasil scraping yang sebelumnya berjumlah 150 data menjadi 1151 data dengan 8 fitur. Adapun hasil pada proses scraping sebagai berikut :

	Description	Processor	RAM	Operating System	Storage	Display	Warranty	Price
0	Primebook 4G Android Based MediaTek MT8788 - (...)	Intel Core i3 Processor (11th Gen)	4 GB DDR4 RAM	Prime OS Operating System	512 GB SSD	29.46 cm (11.6 Inch) Display	1 Year Onsite Warranty	16,990
1	Primebook 4G Android Based MediaTek MT8788 - (...)	Intel Core i5 Processor (11th Gen)	4 GB DDR4 RAM	Prime OS Operating System	512 GB SSD	29.46 cm (11.6 Inch) Display	1 Year Onsite Warranty	18,990
2	ASUS Vivobook 15 Core i3 11th Gen - (8 GB/512 ...)	Intel Core i3 Processor (11th Gen)	8 GB DDR4 RAM	64 bit Windows 11 Operating System	512 GB SSD	39.62 cm (15.6 Inch) Display	1 Year Onsite Warranty	38,990
3	ASUS Vivobook 15 Core i5 11th Gen - (8 GB/512 ...)	AMD Ryzen 5 Hexa Core Processor	8 GB DDR4 RAM	64 bit Windows 11 Operating System	1 TB SSD	39.62 cm (15.6 Inch) Display	1 Year Onsite Warranty	47,990
4	HP 14s Intel Core i3 11th Gen - (8 GB/512 GB S...	Intel Core i7 Processor (12th Gen)	8 GB DDR4 RAM	64 bit Windows 11 Operating System	256 GB SSD	35.56 cm (14 inch) Display	1 Year Onsite Warranty	38,990

Berikut adalah gambaran data yang didapatkan dengan diambil datasetnya (head(5)), dapat dilihat bahwa terdiri dari deskripsi, merepresentasikan deskripsi dari laptop, karakteristik processor dari laptop, type RAM dan versi DDR nya sebagai size slot ram, jenis sistem operasi yang digunakan, penyimpanannya, tampilan, warranty, harga.

2. Data preparation

Data preparation saya baru di tahap feature engineer dengan mentransformasi beberapa fitur, adapun fitur yang ditransformasi sebagai berikut :

- RAM : Membuat columns untuk kapasitas ram dalam GB dan versi DDR
- Processor : Membuat columns untuk nama processor, tipe processor, generasi
- OS : menguraikan os dari column dan membuat new column
- Storage : Membuat new column untuk type dari disk drive dan kapasitas disk drive
- Display : Membuat new column untuk size laptop (in inches) and touchscreen
- Description : Membuat new columns untuk company and graphic card

Jumlah fitur yang sebelumnya adalah 8 menjadi 12 fitur, adapun hasil dari proses feature engineer adalah sebagai berikut :

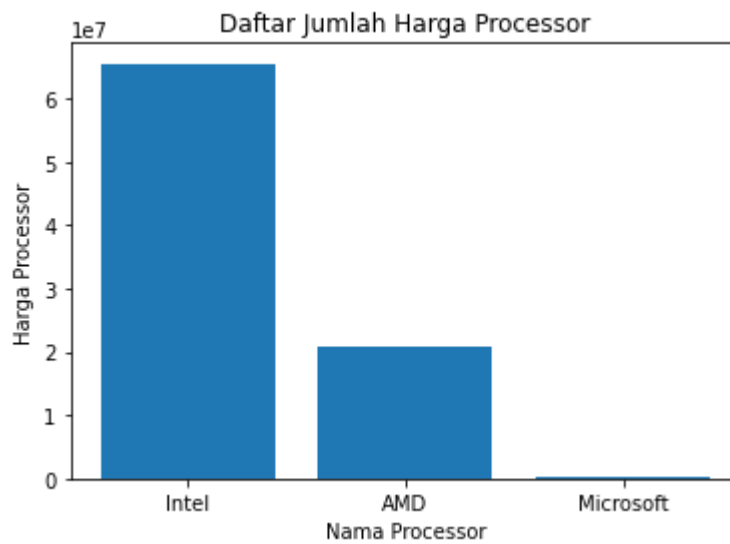
	Price	ram_in_gb	ddr_version	processor_name	gen_type	OS	disk_drive	storage_in_gb	size_in_inches	touchscreen	company	graphic_card
0	16,990	4	4	Intel	10	Windows	SSD	512	11.6	No	Primebook	No
1	18,990	4	4	Intel	10	Windows	SSD	512	11.6	No	Primebook	No
2	38,990	8	4	Intel	10	Windows	SSD	512	15.6	No	Asus	No
3	47,990	8	4	AMD	Not Mentioned	Windows	SSD	1000	15.6	No	Asus	No
4	38,990	8	4	Intel	10	Windows	SSD	256	14	No	HP	No
...
1146	20,990	4	4	AMD	Not Mentioned	Windows	SSD	128	14	No	Lenovo	No
1147	97,990	8	4	Intel	9	Windows	Both	1000+256	15.6	No	Acer	Yes
1148	94,990	8	4	Intel	8	Windows	SSD	512	13.3	Yes	Lenovo	No
1149	99,990	16	4	Intel	10	Windows	SSD	1000	13.3	No	Asus	No
1150	2,29,990	16	3	Intel	7	Mac	SSD	512	12	No	Apple	Yes

Setiap fitur di ekstrak informasinya sehingga setiap informasi pada fitur didapatkan informasinya dalam setiap column.

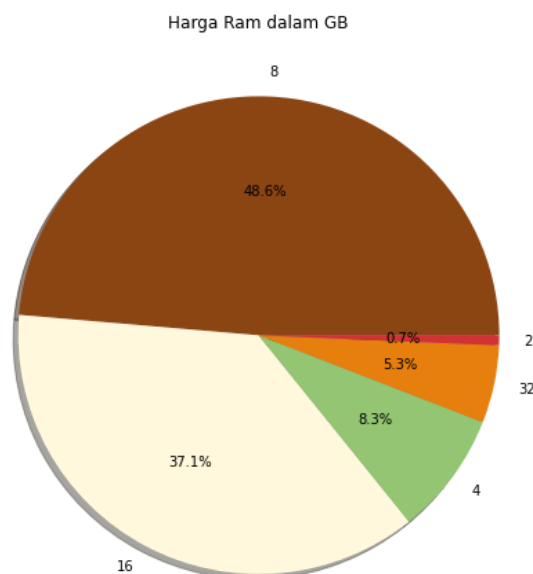
3. Data Visualisasi

Adapun data yang di visualisasi menggabungkan teknik group by untuk mendapatkan intisari dari setiap informasi antara column pada dataset;

a. Harga Processor

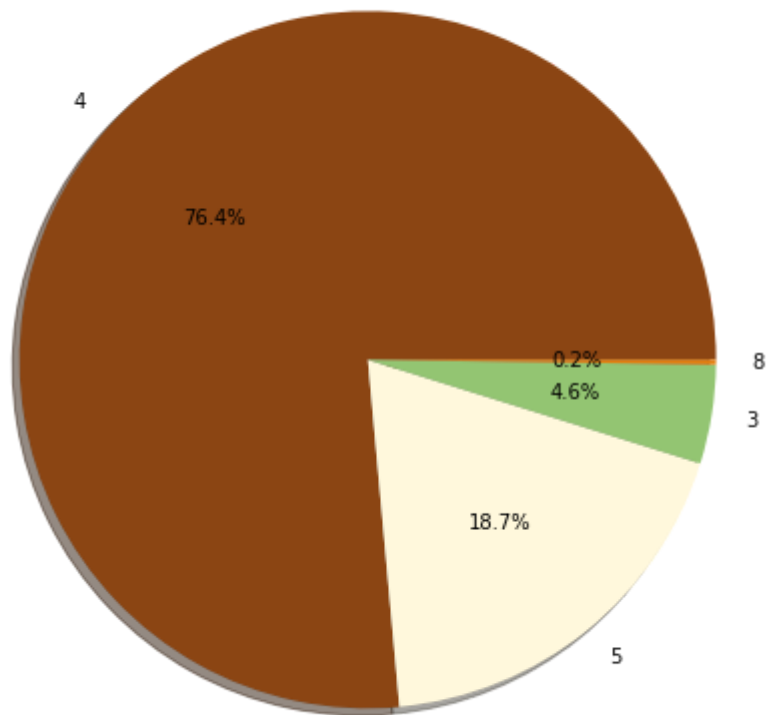


b. List Harga ram dalam GB

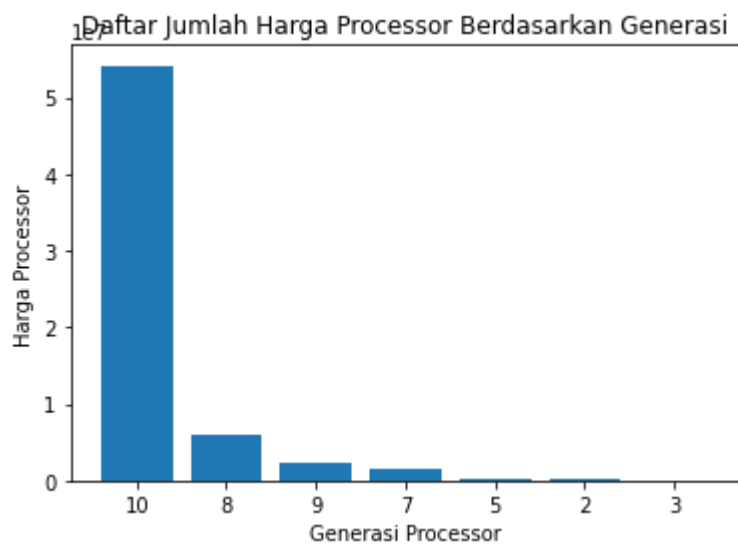


c. Harga Ram berdasarkan DDR version

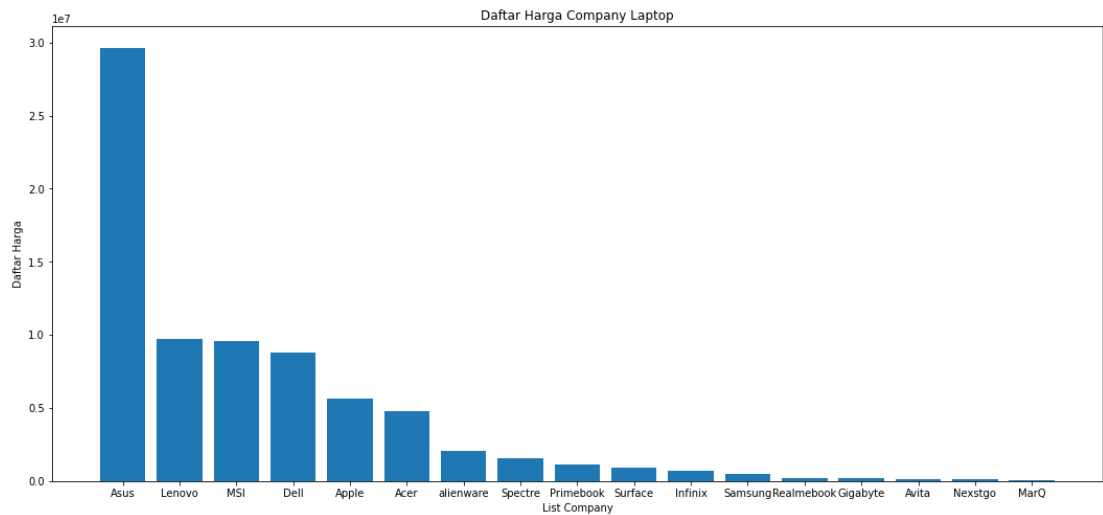
Harga Ram dalam GB Berdasarkan DDR-version



d. Type generasi processor



e. Harga laptop disetiap perusahaan



4. Modeling

Pada tahap modeling baru proses tahapan riset mengenai model apa yang baik digunakan jika datanya lebih condong terhadap representasi value, categorical ataupun data yang bersifat pilihan ke dalam list tuning model agar model yang di tuning tidak terlalu banyak.