

LAPORAN
SAINS DATA INDUSTRI

Project 3 Task 2



KELAS DS-45-GAB

ANGGOTA KELOMPOK 15:

Muhammad Ali Akbar Al - Qahri	(1305210077)
Haura Adzkia Delfina	(1305213006)
Zadosaadi Brahmantio Purwanto	(1305213045)

FAKULTAS INFORMATIKA

**Telkom University, Jl Telekomunikasi No. 1 Terusan Buahbatu - Bojongsoang, Sukapura,
Kec. Dayeuhkolot, Kabupaten Bandung, Jawa Barat 40257**

2023

DAFTAR ISI

DAFTAR ISI	2
BAB I	3
PENDAHULUAN	3
A. Latar Belakang	3
B. Rumusah Masalah	3
BAB II	4
PEMBAHASAN	4
A. Model Pre-Proses	4
1. BG NBD Model	4
2. Gamma Model	5
B. Model Evaluation	5
C. Segmentasi untuk Customer	6
D. Visualisasi	6
1. Top 10 purchase frequency per item	6
2. Average and total CLV per Segment	7
3. Sales Over Time per Segment	8
4. Correlation between frequency, monetary value, and recency	8
BAB III	10
KESIMPULAN	10

BAB I

PENDAHULUAN

A. Latar Belakang

Online retail telah menjadi salah satu industri yang banyak diminati dalam beberapa tahun terakhir. Untuk mempelajari prediksi *customer lifetime value* selama 3 bulan perusahaan-perusahaan *online retail*, kita dapat mengeksplor dan menganalisis dataset yang ada mengenai *online retail*. Dataset yang digunakan merupakan dataset *online retail* yang berisi informasi mengenai penjualan dari perusahaan di beberapa negara. Dataset ini mengandung informasi sekitar 541 ribu data mengenai penjualan dari seperti nomor bukti transaksi, kode stok barang, deskripsi barang, kuantitas barang, tanggal transaksi, harga unit, identitas pelanggan, dan juga negara tujuan produk terjual.

Dengan melakukan eksplorasi dan analisis terhadap dataset *online retail* dapat memberikan pemahaman yang lebih baik mengenai *customer lifetime value*, perilaku pembelian pelanggan, mengidentifikasi tren produk, mengetahui preferensi pelanggan, meningkatkan strategi pemasaran, dan mengembangkan rencana bisnis yang lebih baik. Dengan adanya informasi-informasi tersebut dapat memudahkan perusahaan *online retail* mengetahui bagaimana prediksi *lifetime value* dari pelanggan dan juga dapat mengoptimalkan strategi penjualan, menentukan pelanggan yang berpotensi dalam hal pemasaran, dan mengambil keputusan yang lebih baik mengenai pengelolaan persediaan dan pengembangan produk. Dengan kata lain analisis dataset *online retail* ini dapat membantu perusahaan meningkatkan kinerja bisnis dan memenuhi kebutuhan pelanggan dengan lebih baik.

B. Rumusah Masalah

1. Apa metode yang digunakan untuk mengetahui prediksi *customer lifetime value* selama 3 bulan pada dataset *online retail* ?
2. Bagaimana prediksi *customer lifetime value* selama 3 bulan pada dataset *online retail* ?

C. Tujuan

1. Mempelajari metode yang digunakan untuk mengetahui prediksi *customer lifetime value* selama 3 bulan pada dataset *online retail*
2. Mengetahui prediksi *customer lifetime value* selama 3 bulan pada dataset *online retail*

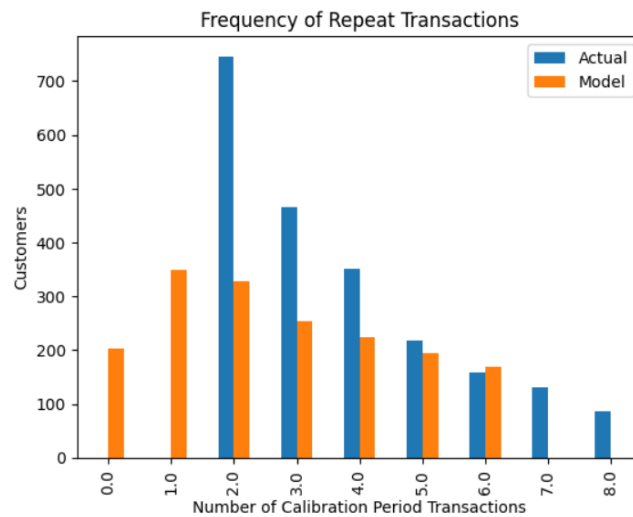
BAB II

PEMBAHASAN

A. Model Pre-Proses

1. BG NBD Model

BG NBD (Beta Geometric Negative Binomial Distribution) adalah model yang digunakan analisis dalam customer lifetime value (CLV) untuk memodelkan perilaku pembelian pelanggan berulang dari perusahaan. Model ini mengasumsikan bahwa pembelian pelanggan terjadi secara stokastik dan bahwa pembelian pelanggan memiliki kecenderungan tertentu untuk melakukan pembelian di masa depan



Dengan menggunakan code tersebut digunakan untuk menghasilkan prediksi jumlah pembelian yang diharapkan dalam satu minggu kedepan untuk setiap pelanggan dalam dataset, serta menampilkan pelanggan teratas berdasarkan prediksi tersebut

CustomerID	recency	T	frequency	monetary_value	expected_purc_1_week
12748.0000	53.1429	53.4286	210	154.1924	3.2530
17841.0000	53.0000	53.4286	124	330.1344	1.9333
13089.0000	52.2857	52.8571	97	606.3625	1.5320
14606.0000	53.1429	53.4286	93	130.1390	1.4588
15311.0000	53.2857	53.4286	91	667.7791	1.4286
12971.0000	52.5714	53.2857	86	127.4859	1.3523
13408.0000	53.0000	53.4286	62	453.5006	0.9829
18102.0000	52.2857	52.5714	60	3859.7391	0.9653
13798.0000	52.8571	53.2857	57	650.9085	0.9082
14527.0000	52.2857	52.8571	55	154.7058	0.8830

2. Gamma Model

Model gamma merupakan model probabilitas yang digunakan dalam analisis statistik dan pemodelan prediksi terutama dalam analisis customer lifetime value (CLV).

	recency	T	frequency	monetary_value	expected_purc_1_week
CustomerID					
12748.0000	53.1429	53.4286	210	154.1924	3.2530
17841.0000	53.0000	53.4286	124	330.1344	1.9333
13089.0000	52.2857	52.8571	97	606.3625	1.5320
14606.0000	53.1429	53.4286	93	130.1390	1.4588
15311.0000	53.2857	53.4286	91	667.7791	1.4286
12971.0000	52.5714	53.2857	86	127.4859	1.3523
13408.0000	53.0000	53.4286	62	453.5006	0.9829
18102.0000	52.2857	52.5714	60	3859.7391	0.9653
13798.0000	52.8571	53.2857	57	650.9085	0.9082
14527.0000	52.2857	52.8571	55	154.7058	0.8830

Dengan menggunakan model gamma yang bertujuan untuk memperkirakan rata-rata keuntungan yang diharapkan per transaksi dan memberikan pemahaman tentang pelanggan dengan potensi keuntungan tertinggi.

B. Model Evaluation

Model ini bertujuan untuk menghitung CLV menggunakan model gamma-gamma, menyatukan hasil CLV dengan data pelanggan yang ada, dan menghasilkan dataset akhir yang berisi informasi CLV untuk setiap pelanggan.

	CustomerID	recency	T	frequency	monetary_value	expected_purc_1_week	exp_avg_profit	clv
2486	18102.0000	52.2857	52.5714	60	3859.7391	0.9653	3870.8387	176133.7843
589	14096.0000	13.8571	14.5714	17	3163.5882	0.7231	3195.9719	105051.1555
2184	17450.0000	51.2857	52.5714	46	2863.2749	0.7451	2874.0455	100934.1543
2213	17511.0000	52.8571	53.4286	31	2933.9431	0.5074	2950.3463	70574.4182
1804	16684.0000	50.4286	51.2857	28	2209.9691	0.4767	2223.6899	49921.7792
406	13694.0000	52.7143	53.4286	50	1275.7005	0.7983	1280.1557	48186.9705
587	14088.0000	44.5714	46.1429	13	3864.5546	0.2597	3916.3832	47750.6104
1173	15311.0000	53.2857	53.4286	91	667.7791	1.4286	669.0782	45080.1888
1485	16000.0000	0.0000	0.4286	3	2335.1200	0.4160	2477.6641	44568.9037
133	13089.0000	52.2857	52.8571	97	606.3625	1.5320	607.4725	43880.6286

C. Segmentasi untuk Customer

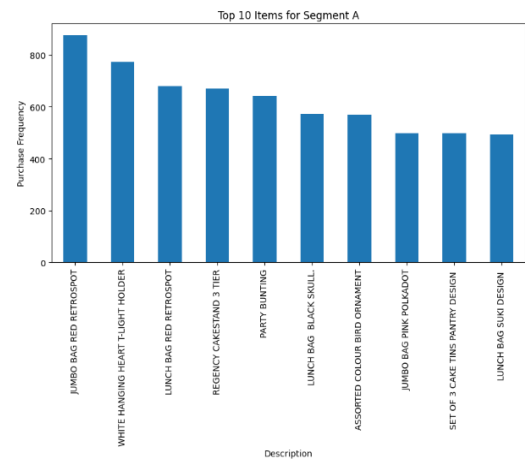
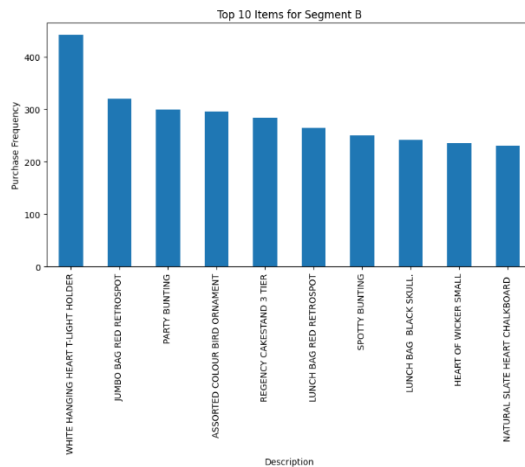
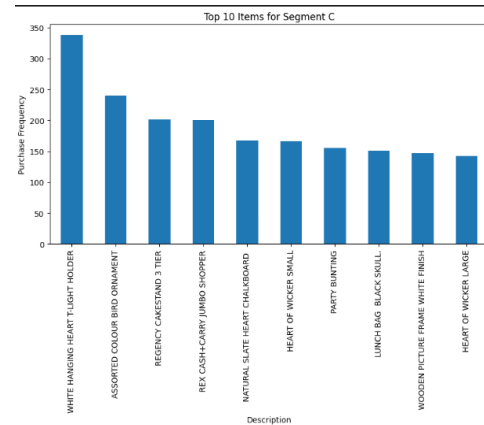
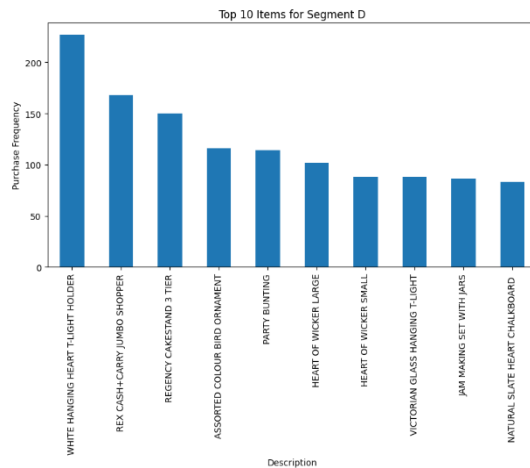
Segmentasi ini digunakan untuk melakukan segmentasi pelanggan berdasarkan nilai CLV yang telah discaling dan memberikan ringkasan statistik untuk setiap segmen pelanggan.

Segment	CustomerID			recency			T			frequency			monetary_value			expected_purc_1_week			exp_avg_profit	
	count	mean	sum	count	mean	sum	count	mean	sum	count	mean	sum	count	mean	sum	count	mean	sum	count	mean
D	643	15700.0653	10095142.0000	643	22.0031	14148.0000	643	40.4897	26034.8571	643	3.0622	1969	643	178.8385	114993.1344	643	0.0709	45.5636	643	193.7562
C	642	15518.5265	9962894.0000	642	30.7655	19751.4286	642	38.0683	24439.8571	642	3.9875	2560	642	260.5130	167249.3501	642	0.1197	76.8738	642	277.9431
B	642	15612.2757	10023081.0000	642	29.9513	19228.7143	642	35.2000	22598.4286	642	5.4626	3507	642	352.7678	226476.9231	642	0.1614	103.6145	642	371.3017
A	643	15383.9114	9891855.0000	643	31.4832	20243.7143	643	34.5528	22217.4286	643	11.2939	7262	643	591.8118	380534.9820	643	0.2724	175.1304	643	614.3773

D. Visualisasi

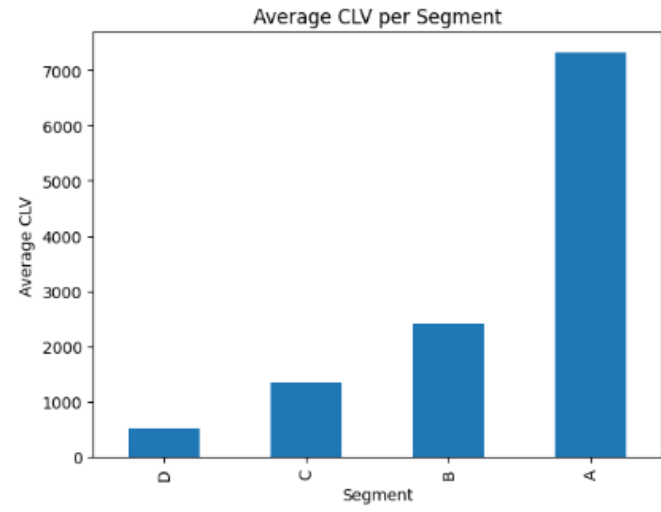
1. Top 10 purchase frequency per item

Untuk melakukan analisis item-item teratas dalam setiap segmen pelanggan berdasarkan frekuensi pembelian, dan memvisualisasikan hasilnya dalam bentuk plot bar.

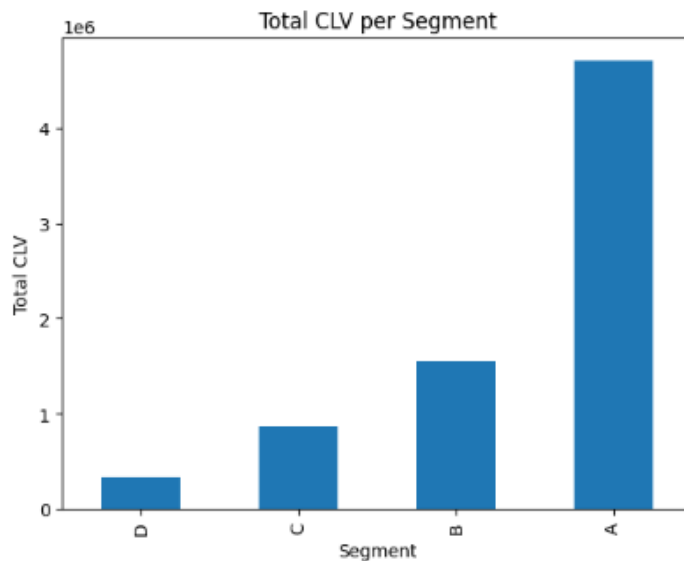


2. Average and total CLV per Segment

Pada bagian ini, dilakukan analisis rata-rata dan total customer lifetime value per segmen pelanggan. CLV merupakan metrik mengukur nilai total yang dapat diharapkan dari setiap pelanggan selama masa hidupnya. Segmen pelanggan dihasilkan dari proses segmentasi sebelumnya.



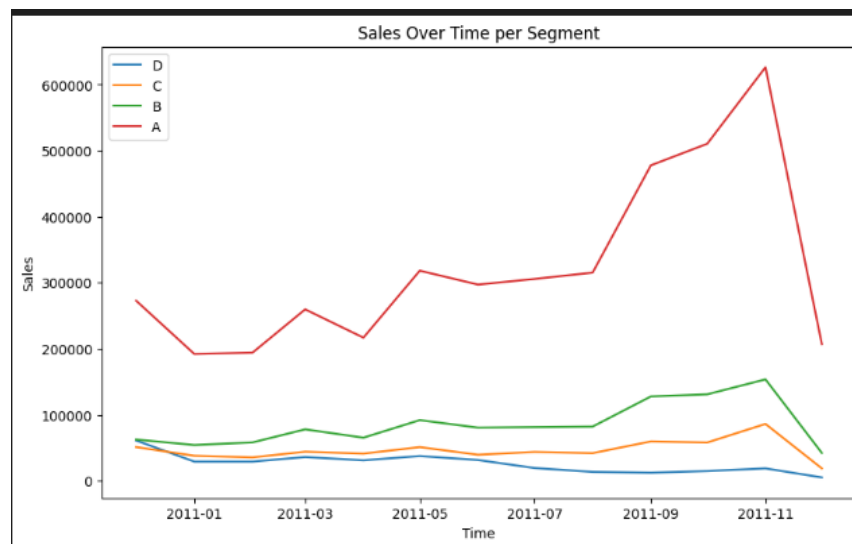
Grafik pertama menampilkan rata-rata CLV per segmen dalam bentuk plot bar. Setiap segmen memiliki batang yang menunjukkan nilai rata-rata CLV. grafik ini memberikan pemahaman tentang nilai rata-rata yang dapat diharapkan dari setiap pelanggan dalam segmen tertentu.



Grafik kedua menampilkan total CLV per segmen dalam bentuk plot bar. Setiap segmen memiliki batang yang menunjukkan nilai total CLV. grafik ini memberikan pemahaman tentang kontribusi total CLV dari setiap segmen pelanggan terhadap keseluruhan nilai CLV.

3. Sales Over Time per Segment

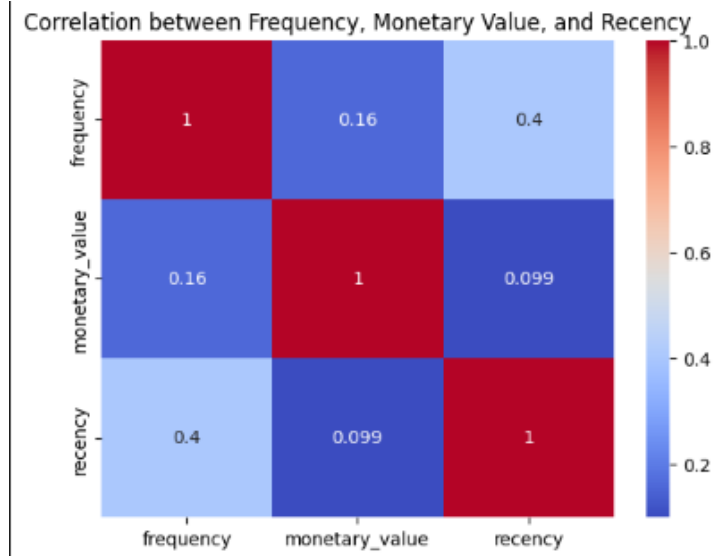
Pada bagian ini dilakukan analisis penjualan seiring waktu untuk setiap segmen pelanggan. Data penjualan dihitung berdasarkan jumlah barang yang terjual dikalikan dengan harga satuan untuk setiap transaksi. Grafik menggunakan seaborn untuk memvisualisasikan tren penjualan per segmen seiring berjalannya waktu.



Analisis penting dalam melihat bagaimana penjualan berkembang dari waktu ke waktu setiap segmen pelanggan. Dengan melihat tren penjualan, kita dapat mengidentifikasi segmen yang mengalami pertumbuhan yang baik dan segmen yang mungkin membutuhkan perhatian lebih dalam upaya meningkatkan penjualan.

4. Correlation between frequency, monetary value, and recency

Pada bagian ini, dilakukan analisis korelasi antara frekuensi pembelian, nilai moneter, dan jangka waktu terakhir pembelian pada segmen pelanggan. Korelasi yang digunakan untuk mengukir hubungan linier antara variabel-variabel tersebut.



Analisis ini penting dalam memahami hubungan antara frekuensi pembelian, nilai moneter, dan jangka waktu terakhir pembelian dalam segmen pelanggan. Korelasi yang kuat antara variabel-variabel tersebut dapat memberikan wawasan tentang perilaku dan kebiasaan pembelian pelanggan dalam segmen tertentu.

BAB III

KESIMPULAN

Berdasarkan pembahasan, terdapat beberapa metode yang dapat digunakan untuk mengetahui prediksi CLV dalam rentang waktu tertentu, dalam kasus ini tiga bulan. Beberapa metode tersebut antara lain BG NBD Model dan Gamma Model. Untuk menghitung CLV, dapat digunakan BG-NBD, yaitu model analisis CLV yang memodelkan perilaku pembelian repeat customer dari perusahaan. Sedangkan Gamma Model adalah model probabilitas yang dipakai untuk analisis statistik. Dengan model di atas, diharapkan dapat memperkirakan rata-rata keuntungan dan pemahaman mengenai pelanggan dengan potensi keuntungan tertinggi.

Setelah melakukan evaluasi nilai expected purchase dalam rentang waktu tertentu, nilai CLV akan memberikan ringkasan statistik dari setiap segmen. Setiap segmen yang berhubungan dapat digambarkan hubungannya dalam bentuk plot bar. Setelah melakukan segmentasi pelanggan, kita kemudian dapat menganalisis korelasi atau hubungan antara frekuensi pembelian, nilai moneter atau spending, dan jangka waktu terakhir pembelian. Hubungan antar variabel tersebut memberikan wawasan mengenai perilaku dan kebiasaan pelanggan.