

CUSTOMER BONUS MODELLING **PADA PERUSAHAAN TELEKOMUNIKASI**

LAPORAN

Laporan ini sebagai salah satu syarat untuk lulus dari mata kuliah
MA3271 Pemodelan Matematika Tahun Akademik 2020/2021



Disusun oleh:

Yayan Febriansyah	(10118017)
Muhamad Daffa Wicaksana	(10118062)
Mia Agustina Nurfadilah	(10118083)
Arya Pratama Putra	(10118104)

PROGRAM STUDI MATEMATIKA
FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM
INSTITUT TEKNOLOGI BANDUNG
2021

DAFTAR ISI

DAFTAR ISI.....	2
BAB I PENDAHULUAN.....	2
1.1 Latar Belakang.....	3
1.2 Rumusan Masalah.....	3
1.3 Tujuan Penulisan.....	3
BAB II TINJAUAN PUSTAKA.....	4
2.1 Kerangka Teori.....	4
2.1.1 Min-max <i>Scaling</i>	4
2.1.2 Teorema Bayes.....	4
2.1.3 K-means <i>Clustering</i>	5
2.1.4 <i>Elbow Method</i>	5
2.1.5 Klaster <i>Risk-Value Customer</i>	6
2.2 Siklus Pemodelan Utuh.....	7
BAB III HASIL DAN PEMBAHASAN.....	8
3.1 Analisis Data.....	8
3.2 Asumsi.....	12
3.3 Analisis Model.....	12
3.3.1 Model <i>Value</i>	13
3.3.2 Model Risiko.....	14
3.3.3 Model <i>Clustering</i>	15
3.3.4 Identifikasi Karakteristik Pelanggan.....	22
3.3.5 Skema Pemberian Bonus.....	29
BAB IV KESIMPULAN DAN SARAN.....	32
4.1 Kesimpulan.....	32
4.2 Saran.....	33
DAFTAR PUSTAKA.....	34

BAB I

PENDAHULUAN

1.1 Latar Belakang

Pelanggan merupakan salah satu pemangku kepentingan (*stakeholder*) yang memiliki pengaruh sangat besar terhadap suatu perusahaan. Bahkan, dapat dikatakan bahwa memenuhi kebutuhan dan keinginan pelanggan merupakan kegiatan utama dari sebuah perusahaan. Salah satu indikator terpenuhinya kebutuhan dan keinginan pelanggan adalah tingginya kepuasan pelanggan. Semakin tinggi tingkat kepuasan pelanggan terhadap layanan suatu perusahaan, semakin unggul perusahaan tersebut dibandingkan dengan pesaing-pesaingnya. Hill, dkk. (2007) menyebutkan bahwa tingkat kepuasan pelanggan merupakan barometer yang dapat memprediksi perilaku pelanggan di masa depan. Oleh karena itu, cara menjaga dan meningkatkan kepuasan pelanggan menjadi salah satu konsentrasi utama setiap perusahaan.

Di sektor perusahaan jasa, tren kepuasan pelanggan yang tinggi mengarah pada daya saing yang semakin kuat. Salah satu sektor tersebut adalah sektor industri telekomunikasi. Dalam keberlangsungan bisnis industri telekomunikasi, pelanggan mampu memilih di antara beberapa penyedia layanan dan secara aktif menggunakan hak mereka untuk beralih ke satu layanan penyedia lain, sehingga kepuasan pelanggan berperan penting untuk menjaga kelangsungan bisnis agar dapat menghadapi persaingan. Oleh karena itu, untuk memastikan keberlangsungan bisnis, perusahaan telekomunikasi perlu berupaya untuk menjaga loyalitas pelanggan.

Berbagai cara dilakukan oleh perusahaan untuk menjaga loyalitas pelanggan, salah satunya adalah pemberian bonus kepada pelanggan. Karena perusahaan memiliki *budget* yang terbatas, perusahaan tidak bisa mengalokasikan *budget* pemberian bonus untuk seluruh pelanggan. Oleh karena itu, perusahaan perlu menentukan pelanggan mana yang paling layak untuk dipertahankan. Karena perusahaan mencari keuntungan, tentu saja perusahaan akan berfokus pada pelanggan yang bernilai tinggi untuk perusahaan. Selain itu, perusahaan juga perlu melihat risiko pelanggan akan beralih, sehingga perusahaan dapat lebih memfokuskan sumber daya yang ada untuk mempertahankan pelanggan yang memiliki kemungkinan pergi yang lebih tinggi.

Berdasarkan hal-hal yang telah dipaparkan di atas, penulis memutuskan untuk memodelkan segmentasi pelanggan dan skema pemberian bonusnya pada pelanggan perusahaan telekomunikasi.

1.2 Rumusan Masalah

1. Pelanggan mana yang layak untuk diberikan bonus?
2. Bagaimana skema pemberian bonus yang terbaik?

1.3 Tujuan Penulisan

1. Menentukan pelanggan yang layak untuk diberikan bonus.
2. Menentukan skema pemberian bonus yang terbaik.

BAB II

TINJAUAN PUSTAKA

2.1 Kerangka Teori

2.1.1 Min-max *Scaling*

Algoritma klasterisasi, terutama yang menggunakan jarak sebagai patokan, cenderung akan memiliki performa yang baik apabila variabel masukan berada pada *range* standar. Hal ini disebut sebagai *scaling*. Salah satu bentuk dari *scaling* adalah normalisasi yang mengubah nilai pada variabel masukan menjadi berada pada *range* [0,1]. Salah satu dari metode normalisasi yang umum dan relatif mudah untuk diterapkan adalah Min-max *scaling*. Min-max *scaling* akan mengubah nilai pada variabel masukan menjadi berada pada *range* [0,1] menggunakan formula berikut:

$$x' = \frac{x - \min(x)}{\max(x) - \min(x)}$$

Keterangan:

x : variabel masukan

2.1.2 Teorema Bayes

Teorema Bayes merupakan sebuah teorema yang dikemukakan oleh seorang ilmuwan Inggris yaitu Thomas Bayes, untuk memprediksi probabilitas di masa depan berdasarkan pengalaman di masa sebelumnya. Asumsi dasar yang digunakan dalam Teorema Bayes adalah variabel-variabelnya saling bebas. Teorema Bayes memiliki formula:

$$P(y|X) = \frac{P(X|y) P(y)}{P(X)}$$

dengan variabel y adalah variabel yang diprediksi, sedangkan variabel X adalah variabel yang sudah diketahui kejadiannya, di mana variabel X dapat terdiri dari n buah kejadian ($X = (x_1, x_2, \dots, x_n)$). Dengan mensubstitusi X dan memperluas persamaan menggunakan aturan rantai, diperoleh:

$$P(y|x_1, x_2, \dots, x_n) = \frac{P(x_1|y)P(x_2|y)\dots P(x_n|y)P(y)}{P(x_1)P(x_2)\dots P(x_n)}$$

dengan y adalah variabel yang diprediksi, dan x_i $i = 1, 2, \dots, n$ adalah variabel yang memprediksinya, dalam hal ini terdapat n buah variabel prediktor.

2.1.3 K-means *Clustering*

K-means *clustering* adalah salah satu metode klasterisasi yang berprinsip pada kedekatan jarak. Karenanya, titik-titik yang berjarak sama akan berada pada satu klaster. Langkah-langkah dalam K-means *clustering*:

1. Inisialisasi klaster sebanyak k klaster (dipilih secara bebas).
2. Tentukan sebanyak k sentroid awal, yaitu titik pusat klaster awal secara acak.
3. Hitung jarak suatu titik pada plot dengan masing-masing sentroid dengan menggunakan jarak Euclid, yaitu:

$$d_{(p,q)} = \sqrt{\sum_{i=1}^n (p_i - q_i)^2}$$

Keterangan:

$d_{(p,q)}$:Jarak Euclid

n :Banyak dimensi

p_i :Titik p pada koordinat i

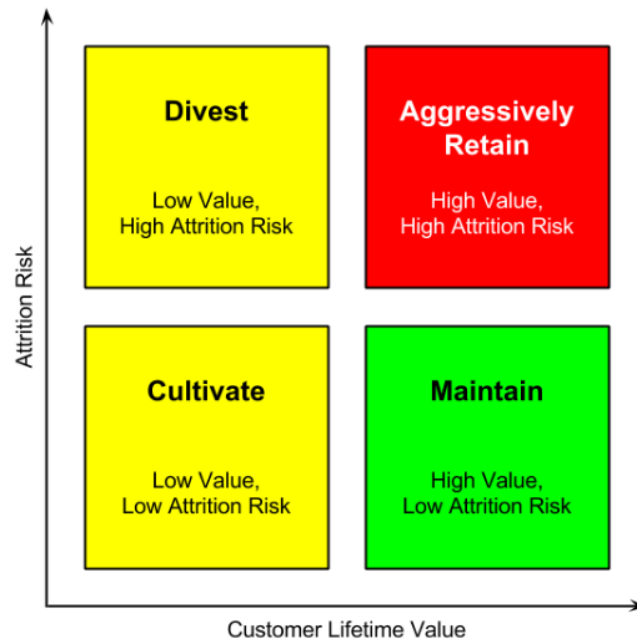
q_i :Titik q pada koordinat i

4. Kelompokkan setiap titik dengan sentroid yang jaraknya paling dekat, sehingga menjadi satu klaster dengan sentroid tersebut.
5. Setelah mendapatkan sejumlah k klaster tersebut kemudian cari rata-rata tiap klaster untuk mendapatkan sentroid yang baru, lalu hitung jarak suatu titik dengan masing masing sentroid yang baru (kembali ke langkah 2) dan seterusnya sampai sentroid tidak mengalami perubahan nilai lagi sehingga didapatkan sejumlah k klaster akhir.

2.1.4 *Elbow Method*

Elbow method merupakan suatu metode yang digunakan untuk menghasilkan informasi dalam menentukan jumlah klaster terbaik dengan cara melihat persentase hasil perbandingan antara jumlah klaster yang akan membentuk siku pada suatu titik. Nilai perbandingan tersebut didapatkan dengan menghitung SSE (*Sum of Square Error*) dari masing-masing nilai klaster. Semakin besar jumlah klaster k , maka nilai SSE akan semakin kecil.

2.1.5 Klaster *Risk-Value Customer*



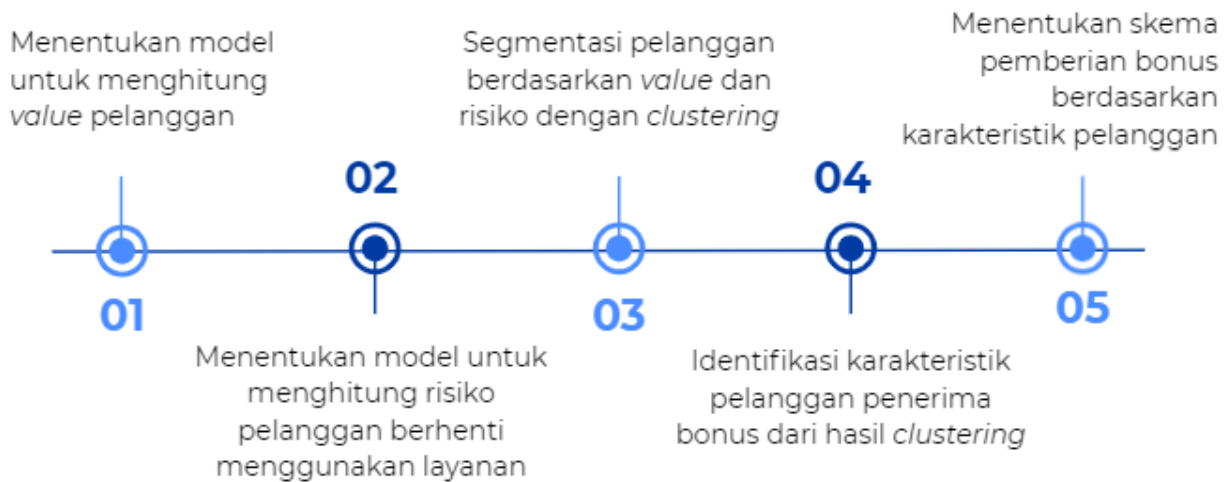
Jika perusahaan dapat mengidentifikasi pelanggan mana yang kemungkinan besar akan pergi, perusahaan dapat secara proaktif memfokuskan upaya pemasaran untuk mempertahankan mereka. Dengan pemikiran ini, gambar di atas menunjukkan empat kuadran pemasaran yang menggambarkan segmentasi pelanggan. Sumbu horizontal menunjukkan *customer lifetime value*. Sumbu vertikal menunjukkan *attrition risk*. Berdasarkan hal tersebut, daerah ini dapat dibagi menjadi empat kuadran:

1. *Maintain*: berisi pelanggan dengan nilai tinggi dan risiko peralihan rendah. Biasanya ini adalah pelanggan terbaik perusahaan, dan perusahaan harus membuat mereka puas tanpa terlalu memusatkan perhatian pada mereka jika memungkinkan.
2. *Cultivate*: berisi pelanggan dengan nilai rendah dan risiko peralihan rendah. Perusahaan kemungkinan besar tidak akan kehilangan pelanggan ini dalam waktu dekat, tetapi pelanggan seperti ini kurang bernilai untuk perusahaan. Oleh karena itu perusahaan harus mengembangkan pelanggan ini agar memiliki nilai yang lebih tinggi.
3. *Divest*: berisi pelanggan dengan nilai rendah dan risiko peralihan tinggi. Perusahaan tidak rugi jika kehilangan pelanggan seperti ini, karena nilai mereka rendah dan perlu usaha lebih untuk mempertahankannya karena risiko peralihan yang tinggi.
4. *Aggressively maintain*: berisi pelanggan dengan nilai tinggi dan risiko gesekan tinggi, dan merekalah yang harus paling diperhatikan. Jika perusahaan kehilangan pelanggan seperti ini, hal tersebut akan membuat perbedaan besar pada ukuran aset perusahaan.

Menggunakan ide ini, penulis akan mengadaptasikannya pada model *clustering* yang akan dibuat.

2.2 Siklus Pemodelan Utuh

Setelah ditemukan data yang akan digunakan dalam pemodelan, berikut siklus pemodelan yang penulis rancang. Dikatakan siklus karena pada setiap langkahnya, dilakukan secara iteratif sampai penulis memutuskan bahwa model yang dibuat sudah cukup memuaskan.



BAB III

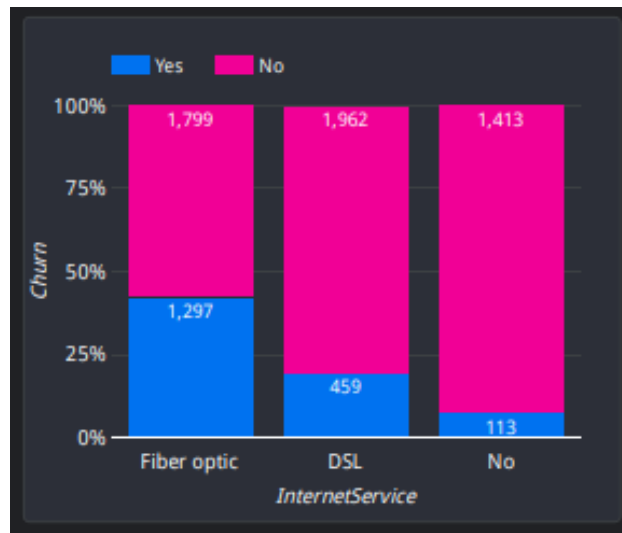
HASIL DAN PEMBAHASAN

3.1 Analisis Data

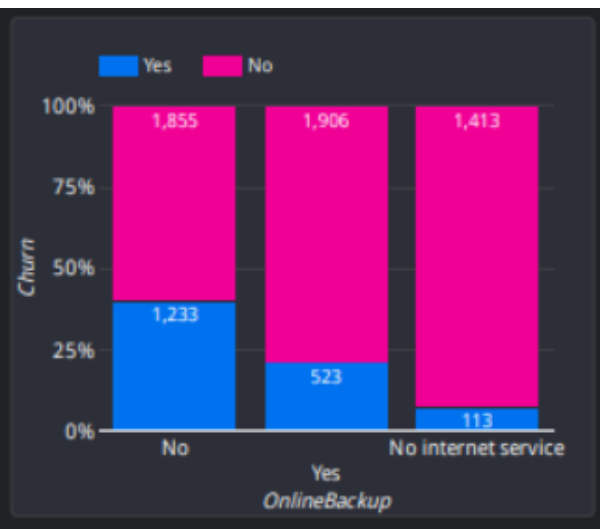
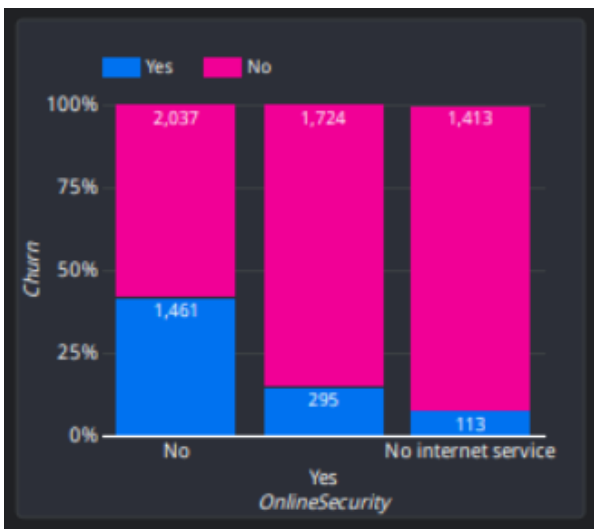
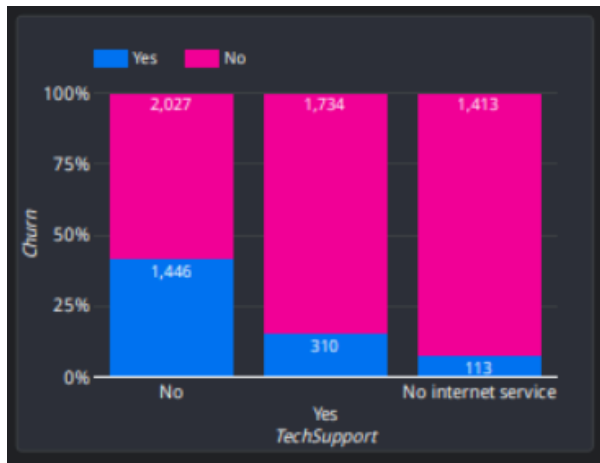
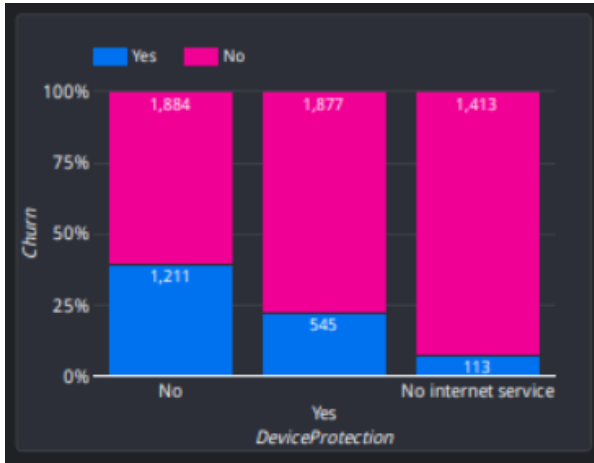
Penulis menggunakan data pelanggan suatu perusahaan telekomunikasi dari situs IBM Accelerator Catalog. Data yang penulis gunakan terdiri dari lima bagian utama, yaitu informasi akun pelanggan, demografi pelanggan, layanan yang digunakan, informasi penggunaan, dan informasi apakah pelanggan telah berhenti menggunakan layanan (*churn*). Berikut variabel-variabel pada masing-masing bagian tersebut.

1. Informasi akun pelanggan – customerID.
2. Demografi pelanggan – gender, age range, partners, dan dependents.
3. Layanan yang digunakan – phone, multiple lines, internet, online security, online backup, device protection, tech support, streaming TV, dan streaming movies.
4. Informasi pembayaran – tenure, contract, payment method, paperless billing, monthly charges, dan total charges.
5. Informasi apakah pelanggan telah berhenti menggunakan layanan – Churn.

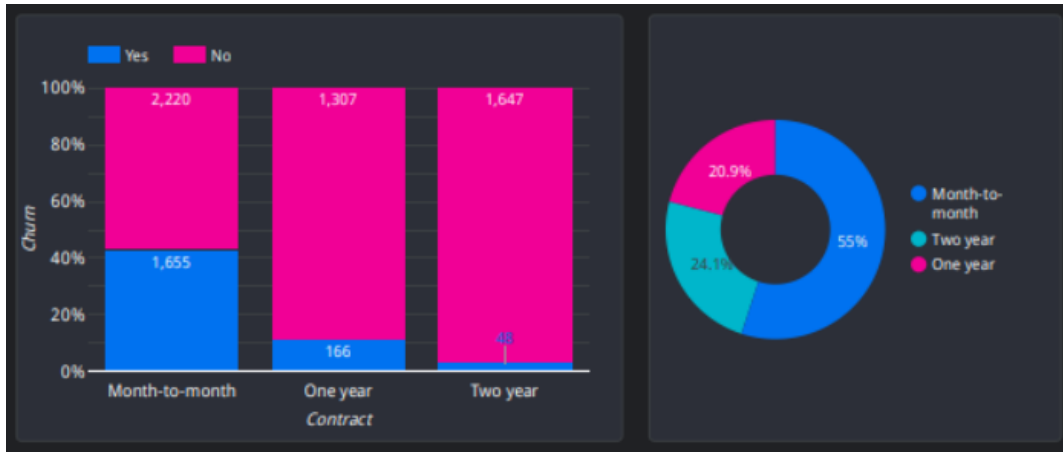
Data tersebut terdiri atas 7043 baris dan 21 kolom. Setiap baris pada data merepresentasikan secara unik informasi seorang pelanggan, sedangkan setiap kolomnya merupakan nilai dari variabel-variabel yang telah disebutkan sebelumnya pada setiap pelanggan. Berikut merupakan beberapa visualisasi dan analisis dari data yang penulis gunakan.



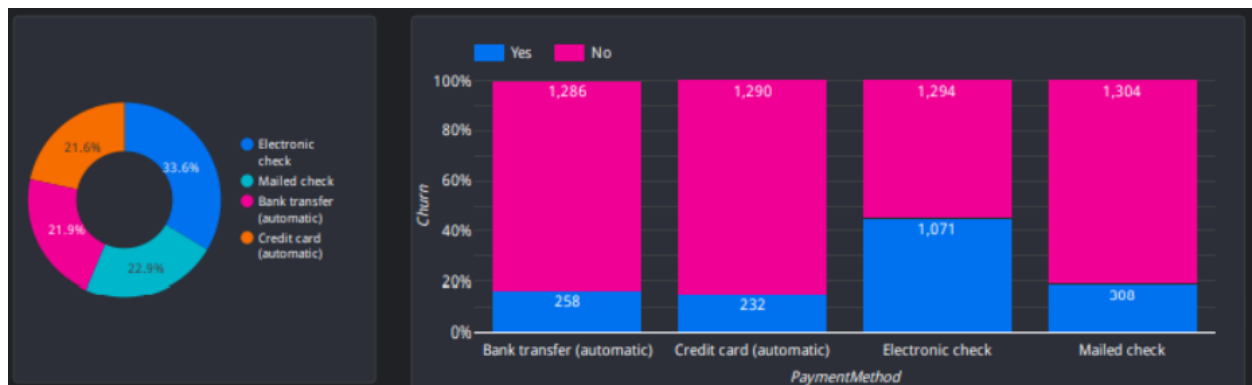
Berdasarkan layanan yang digunakan pelanggan, pada visualisasi terlihat bahwa pelanggan yang menggunakan layanan internet *Fiber optic* lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan dibandingkan dengan pelanggan dengan layanan internet DSL dan pelanggan yang tidak menggunakan layanan internet.



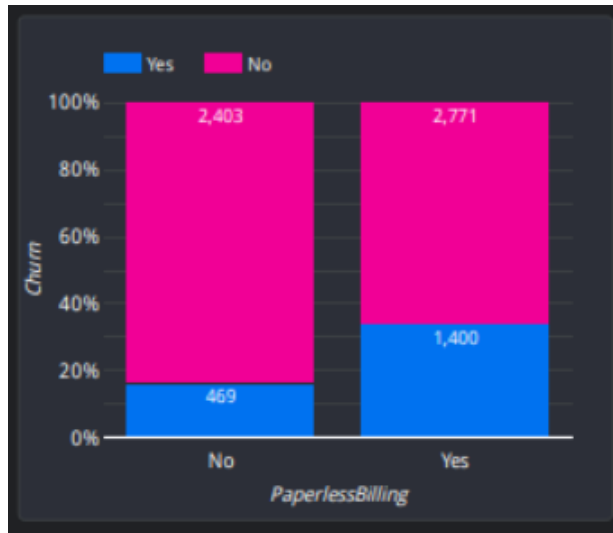
Variabel-variabel pada empat visualisasi tersebut merupakan layanan-layanan tambahan dari layanan internet. Terlihat pada setiap visualisasi bahwa pelanggan yang tidak menggunakan layanan tambahan perusahaan cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan.



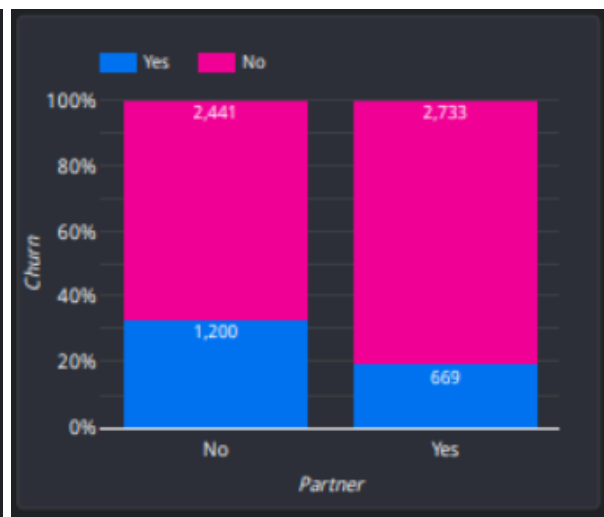
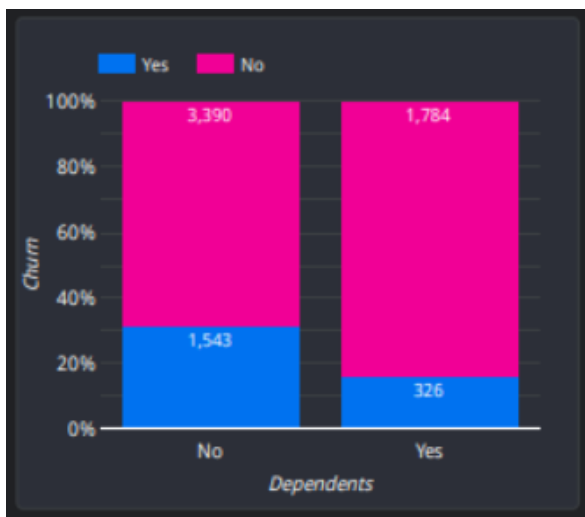
Visualisasi di atas memperlihatkan perbandingan pelanggan dengan kontrak bulanan (*Month-to-month*), satu tahunan (*One year*), dan dua tahunan (*Two year*). Berdasarkan diagram lingkaran pada visualisasi, terlihat bahwa mayoritas pelanggan memiliki kontrak bulanan (sekitar 55%), sedangkan pada diagram batang, pelanggan dengan kontrak bulanan cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan. Di sisi lain, walaupun jumlahnya lebih sedikit dibandingkan pelanggan dengan kontrak bulanan, pelanggan dengan kontrak tahunan cenderung lebih kecil kemungkinannya untuk berhenti menggunakan layanan perusahaan.

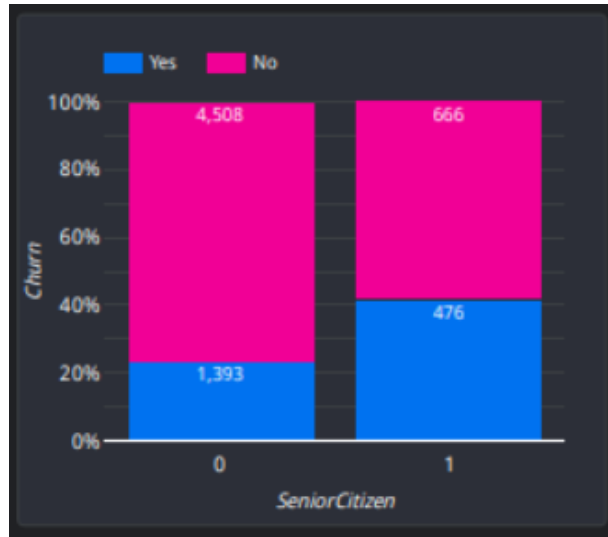


Pada visualisasi tersebut, diperlihatkan perbandingan pelanggan dengan metode pembayaran *Bank transfer* (otomatis), *Credit card* (otomatis), *Electronic check*, dan *Mailed check*. Pada diagram batang terlihat secara jelas bahwa pelanggan yang metode pembayarannya menggunakan *Electronic check* cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan. Dapat diamati juga bahwa jumlah pelanggan dengan metode pembayaran *Electronic check* lebih banyak dibandingkan tiga metode lainnya (sekitar 33%).



Selanjutnya, dapat diamati juga bahwa pelanggan dengan tagihan tanpa kertas (*Paperless Billing*) cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan.





Terakhir, berdasarkan demografinya, terlihat bahwa pelanggan yang tidak memiliki tanggungan (*Dependents*) cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan. Serupa, pelanggan yang belum memiliki pasangan pun cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan. Selain itu, walaupun minoritas, pelanggan berusia lanjut (*Senior Citizen*) juga cenderung lebih besar kemungkinannya untuk berhenti menggunakan layanan perusahaan.

3.2 Asumsi

Asumsi yang digunakan adalah sebagai berikut.

1. *Value* pelanggan dilihat berdasarkan loyalitas dan *revenue* yang dihasilkan dari tiap pelanggan
2. Peluang pelanggan berhenti berlangganan (*churn*) hanya dipengaruhi oleh variabel-variabel kategorikal yang ada dalam data
3. Variabel-variabel kategorikal dalam data bersifat saling bebas
4. Tidak ada faktor eksternal yang memengaruhi *behavior* pelanggan

3.3 Analisis Model

Untuk melakukan segmentasi pelanggan, penulis membuat tiga jenis model, yaitu

1. Model *value*
dibuat untuk menentukan seberapa bernilai seorang pelanggan terhadap perusahaan telekomunikasi yang diteliti
2. Model risiko
dibuat untuk menentukan peluang seorang pelanggan berhenti menggunakan layanan perusahaan telekomunikasi yang diteliti, dan

3. Model *clustering*

dibuat untuk melakukan segmentasi pelanggan berdasarkan model *value* dan model risiko sebelumnya

Selanjutnya, setelah dilakukan segmentasi pelanggan dan penulis mendapatkan data pelanggan yang layak untuk diberikan bonus, penulis melakukan analisis kembali terhadap data tersebut untuk mengidentifikasi karakteristik pelanggan pada setiap segmen. Hasil identifikasi karakteristik pelanggan tersebut selanjutnya penulis gunakan untuk menentukan skema pemberian bonus terbaik pada setiap segmennya.

3.3.1 Model *Value*

Tujuan dari model *value* adalah mengetahui seberapa bernilai seorang pelanggan berdasarkan loyalitas dan potensi *revenue* yang dapat dihasilkan pelanggan tersebut. Penulis merumuskan beberapa model *value* untuk kemudian ditentukan model terbaik di antara model-model yang telah dibuat.

1. Model *Value* 1

Pada model *value* 1, penulis menggunakan hubungan loyalitas yang sebanding dengan lama berlangganan (*tenure*) dan *revenue* yang sebanding dengan tagihan bulanan (*Monthly Charges*) pelanggan. Kemudian, penulis merumuskan *value* pelanggan sebagai berikut.

$$Value = tenure \cdot MonthlyCharges$$

Keterangan:

Value: nilai pelanggan bagi perusahaan

tenure: lama berlangganan seorang pelanggan

MonthlyCharges: besar tagihan setiap bulannya seorang pelanggan

2. Model *Value* 2

Pada model *value* 2, penulis kembali menggunakan hubungan loyalitas yang sebanding dengan lama berlangganan (*tenure*), tetapi untuk *revenue* kali ini sebanding dengan total tagihan (*Total Charges*) pelanggan.

Kali ini, penulis menggunakan Min-max *scaling* dan mengalikan kedua variabel dengan 0.5 untuk menyeimbangkan bobot keduanya, sehingga penulis merumuskan *value* pelanggan sebagai berikut.

$$Value = 0.5 \frac{tenure - \min(tenure)}{\max(tenure) - \min(tenure)} + 0.5 \frac{tenure - \min(TotalCharges)}{\max(TotalCharges) - \min(TotalCharges)}$$

Keterangan:

Value: nilai pelanggan bagi perusahaan

tenure: lama berlangganan seorang pelanggan

Total Charges: besar tagihan seorang pelanggan secara kumulatif

3. Model *Value* 3

Pada model *value* 3, penulis kembali menggunakan hubungan loyalitas yang sebanding dengan lama berlangganan (*tenure*), sedangkan untuk *revenue* penulis kembali menggunakan hubungannya yang sebanding dengan tagihan bulanan (*Monthly Charges*) pelanggan.

Kemudian, seperti pada model *value* 2, penulis menggunakan Min-max *scaling* dan mengalikan kedua variabel dengan 0.5 untuk menyeimbangkan bobot keduanya sehingga penulis merumuskan *value* pelanggan sebagai berikut.

$$Value = 0.5 \frac{tenure - \min(tenure)}{\max(tenure) - \min(tenure)} + 0.5 \frac{tenure - \min(MonthlyCharges)}{\max(MonthlyCharges) - \min(MonthlyCharges)}$$

Keterangan:

Value: nilai pelanggan bagi perusahaan

tenure: lama berlangganan seorang pelanggan

MonthlyCharges: besar tagihan setiap bulannya seorang pelanggan

3.3.2 Model Risiko

Tujuan dari model risiko adalah menentukan peluang seorang pelanggan berhenti menggunakan layanan perusahaan. Dalam memodelkan risiko pelanggan, penulis menggunakan pembobotan pada tiap variabel kategorikal yang ada dalam data, di mana bobot ditentukan berdasarkan peluang pelanggan berhenti menggunakan layanan. Seperti pada model *value*, penulis merumuskan beberapa model risiko untuk kemudian ditentukan model terbaik di antara model-model yang telah dibuat.

1. Model risiko 1

Pada model risiko 1, penulis melakukan pembobotan pada tiap variabel kategorikal berdasarkan peluang empiris kejadian pada variabel yang bersangkutan. Setelah dilakukan pembobotan, nilai peluang dari setiap variabel dijumlahkan.

Contoh: Pelanggan X memiliki tanggungan (*dependents*), menggunakan layanan internet DSL, menggunakan metode pembayaran *credit card*. Risiko pelanggan tersebut berhenti menggunakan layanan adalah sebagai berikut:

$$\begin{aligned} P(\text{Churn} = \text{Yes} \mid \text{Dependents} = \text{Yes}, \text{InternetService} = \text{DSL}, \text{PaymentMethod} = \text{CreditCard}) \\ = P(\text{Churn} = \text{Yes} \mid \text{Dependents} = \text{Yes}) \\ + P(\text{Churn} = \text{Yes} \mid \text{InternetService} = \text{DSL}) \\ + P(\text{Churn} = \text{Yes} \mid \text{PaymentMethod} = \text{CreditCard}) \end{aligned}$$

2. Model risiko 2

Pada model risiko 2, penulis melakukan pembobotan pada tiap variabel kategorikal berdasarkan peluang empiris kejadian pada variabel yang bersangkutan seperti pada

model risiko 1. Perbedaannya yaitu setelah dilakukan pembobotan, nilai peluang dari setiap variabel dikalikan.

Contoh : Pelanggan X memiliki tanggungan (*dependents*), menggunakan layanan internet DSL, menggunakan metode pembayaran *credit card*. Risiko pelanggan tersebut berhenti menggunakan layanan adalah sebagai berikut:

$$\begin{aligned}
 P(\text{Churn} = \text{Yes} \mid \text{Dependents} = \text{Yes}, \text{InternetService} = \text{DSL}, \text{PaymentMethod} = \text{CreditCard}) \\
 &= P(\text{Churn} = \text{Yes} \mid \text{Dependents} = \text{Yes}) \\
 &\times P(\text{Churn} = \text{Yes} \mid \text{InternetService} = \text{DSL}) \\
 &\times P(\text{Churn} = \text{Yes} \mid \text{PaymentMethod} = \text{CreditCard})
 \end{aligned}$$

3. Model risiko 3

Pada model risiko 3, penulis menerapkan Teorema Bayes terhadap variabel-variabel kategorikal yang ada sebagai perhitungan peluang pelanggan berhenti menggunakan layanan perusahaan. Dengan asumsi variabel-variabel tersebut saling bebas, diperoleh peluang pelanggan berhenti menggunakan layanan perusahaan adalah sebagai berikut:

$$P(y \mid x_1, x_2, \dots, x_{16}) = \frac{P(x_1 \mid y) \cdot P(x_2 \mid y) \cdot \dots \cdot P(x_{16} \mid y) \cdot P(y)}{P(x_1) \cdot P(x_2) \cdot \dots \cdot P(x_{16})}$$

Keterangan:

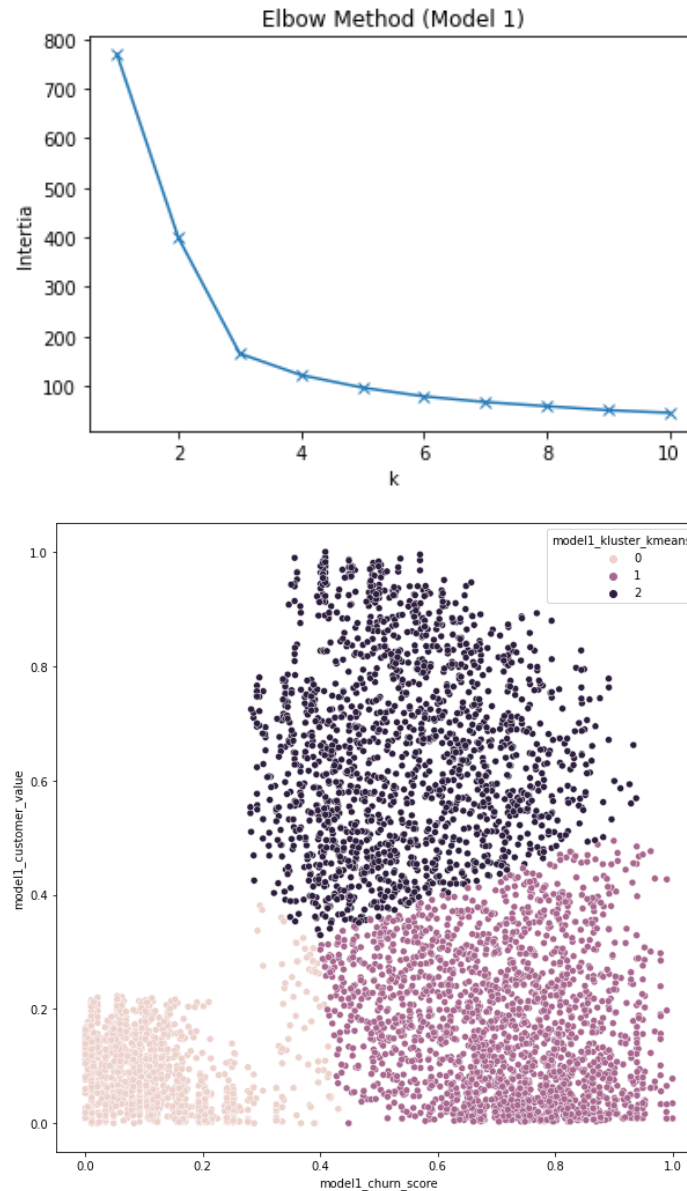
x_i = variabel ke - i

y = churn pelanggan

3.3.3 Model Clustering

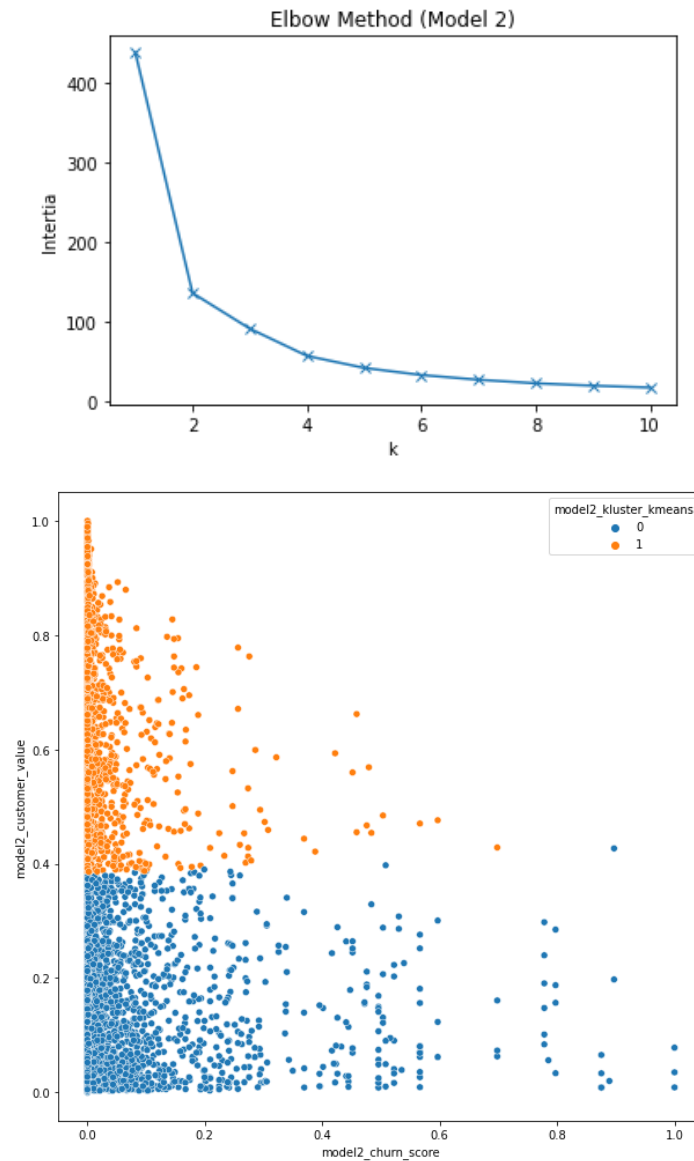
Tujuan dilakukannya klasterisasi pelanggan adalah untuk melakukan segmentasi pelanggan berdasarkan model *value* dan model risiko yang telah dibuat sebelumnya. Kemudian akan diputuskan pelanggan mana yang layak diberikan bonus berdasarkan hasil segmentasi yang diperoleh. Pada model klasterisasi, penulis mengombinasikan model *value* dan model risiko yang telah dibuat sebelumnya, lalu penulis melakukan plot dua dimensi dengan risiko sebagai sumbu horizontal dan *value* sebagai sumbu vertikal. Kemudian dengan menggunakan algoritma K-Means *clustering*, penulis mendapatkan beberapa hasil klasterisasi dan perhitungan *Elbow method* sebagai berikut.

1. Model 1: model *value* 1 dan model risiko 1



Pada model pertama, penulis mengombinasikan model *value* 1 dan model risiko 1. Berdasarkan perhitungan pada *Elbow method*, terlihat bahwa banyak kluster terbaik yang direkomendasikan adalah tiga kluster. Selanjutnya penulis melakukan klasterisasi menggunakan algoritma K-means *clustering*, sehingga didapatkan klasterisasi seperti pada gambar. Pada model 1 ini, terlihat bahwa masih banyak pelanggan yang terkonsentrasi di sudut kiri bawah (*value* dan risiko rendah). Selain pelanggan yang masih banyak terkonsentrasi di satu bagian, hasil klasterisasi ini juga belum sesuai dengan ide awal penulis yang ingin membagi pelanggan menjadi empat kluster. Oleh karena itu, penulis memutuskan untuk memperbaiki model.

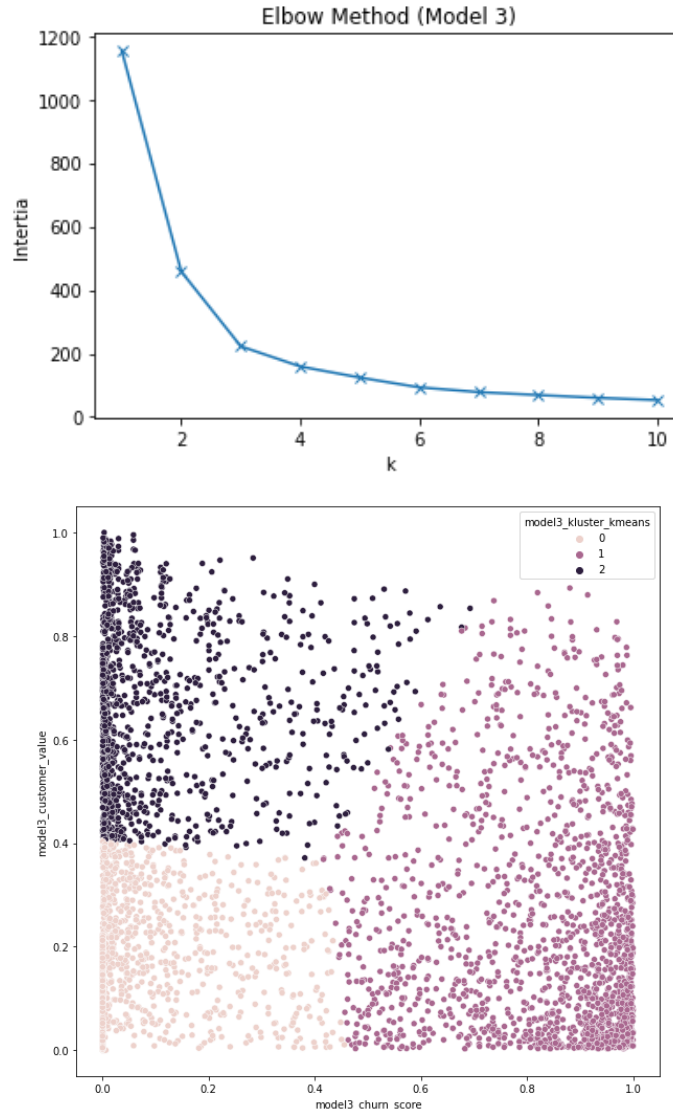
2. Model 2: model *value* 1 dan model risiko 2



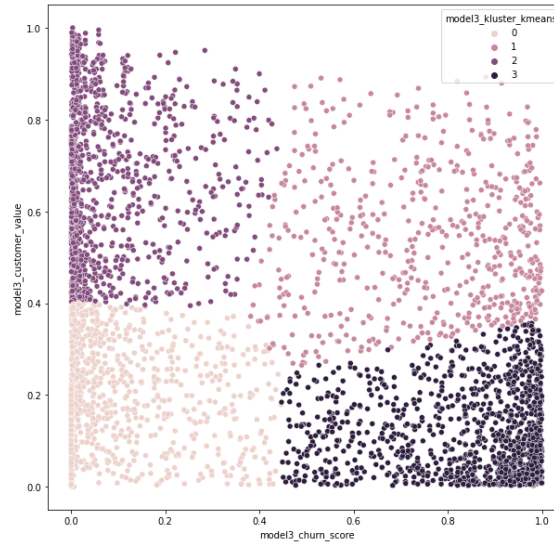
Pada model kedua, penulis memperbaiki model risiko yang telah dibuat, kemudian mengombinasikan model *value* 1 dan model risiko 2. Berdasarkan perhitungan pada *Elbow method*, terlihat bahwa banyak kluster terbaik yang direkomendasikan adalah dua kluster. Selanjutnya, penulis melakukan klusterisasi menggunakan algoritma K-means *clustering* sehingga didapatkan klusterisasi seperti pada gambar. Pada model 2 ini, model risiko yang baru membuat pelanggan terkonsentrasi di bagian kiri (risiko rendah). Hal ini masuk akal karena risiko pelanggan yang awalnya pada model risiko 1 merupakan penjumlahan peluang empiris, sekarang pada model risiko 2 diubah menjadi perkalian sehingga risiko tiap pelanggan akan mengecil. Hal ini tentunya tidak diinginkan penulis karena pelanggan dengan risiko rendah dan tinggi tidak terpisahkan secara jelas. Selain itu, hasil klusterisasi ini juga masih belum sesuai dengan ide awal penulis yang ingin

membagi pelanggan menjadi empat klaster. Oleh karena itu, penulis kembali memutuskan untuk memperbaiki model.

3. Model 3: model *value* 1 dan model risiko 3

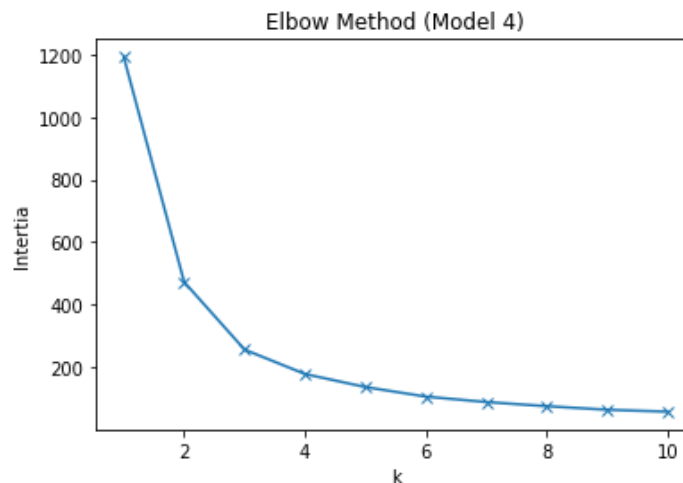


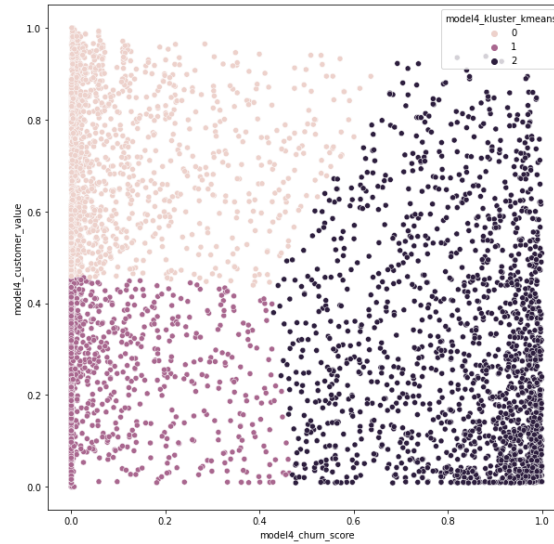
Penulis menyadari bahwa model-model risiko sebelumnya masih belum sesuai dengan teori peluang. Karenanya, pada model risiko 3, penulis menerapkan Teorema Bayes untuk menghitung risiko pelanggan berhenti menggunakan layanan perusahaan. Kemudian pada model ketiga, penulis mengombinasikan model *value* 1 dan model risiko 3. Berdasarkan perhitungan pada *Elbow method*, terlihat bahwa banyak klaster terbaik yang direkomendasikan adalah tiga klaster. Walaupun begitu, terlihat bahwa setelah dilakukan plot menggunakan algoritma K-means *clustering*, pelanggan pada sisi kanan plot masih tidak terpisahkan secara jelas antara pelanggan dengan *value* tinggi dan rendah. Karenanya, penulis melakukan klasterisasi kembali dengan empat klaster seperti pada plot di bawah.



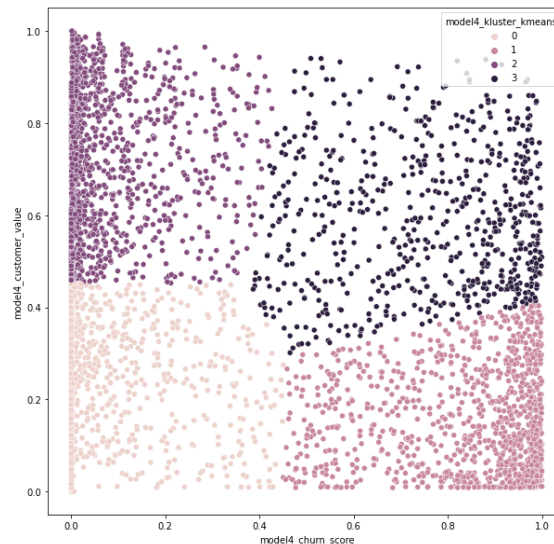
Pada hasil klaster kali ini, banyak klaster sudah sesuai dengan ide awal penulis untuk membagi pelanggan menjadi empat klaster. Namun terlihat bahwa banyak pelanggan yang terkonsentrasi pada bagian sudut kanan bawah (risiko tinggi, *value* rendah). Selain itu, setelah melakukan ekstraksi data, penulis menemukan bahwa pelanggan yang telah sedari lama berlangganan pada perusahaan namun tagihan bulanannya rendah, terkategori dengan *value* rendah dan risiko rendah. Untuk menentukan *value* pelanggan, penulis ingin menyeimbangkan bobot antara loyalitas pelanggan yang direpresentasikan oleh lama berlangganan dan potensi *revenue* yang dihasilkan oleh pelanggan yang pada model ini direpresentasikan oleh tagihan bulanan. Oleh karena itu, penulis kembali memutuskan untuk memperbaiki model.

4. Model 4: model *value* 2 dan model risiko 3



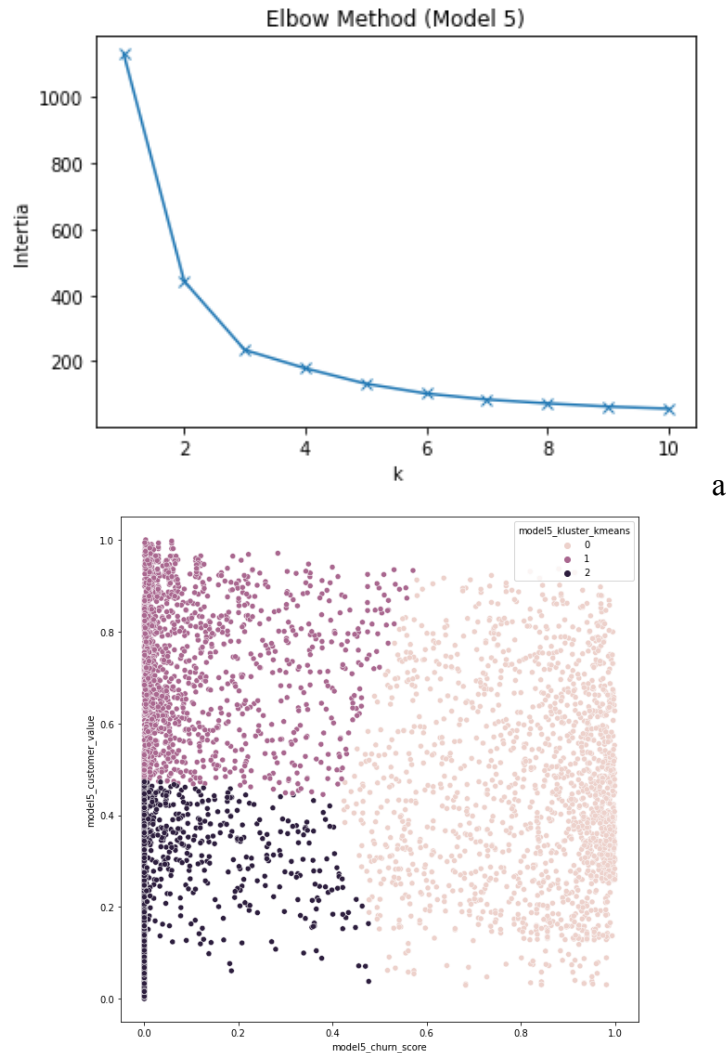


Pada model *value* 1, penulis menentukan *value* pelanggan dengan mengalikan antara lama berlangganan dan tagihan bulanan pelanggan. Dari hal tersebut, didapatkan pada model 3 bahwa pelanggan yang cenderung loyal namun tagihan bulannya rendah dikategorikan pada pelanggan dengan *value* rendah. Karenanya, penulis memperbaiki model *value* 1 dan berusaha untuk menyeimbangkan antara loyalitas dan potensi *revenue* yang dihasilkan pelanggan, kemudian didapatkan model *value* 2. Pada model 4, penulis mengombinasikan model *value* 2 dan model risiko 3. Berdasarkan perhitungan pada *Elbow method*, terlihat bahwa banyak kluster terbaik yang direkomendasikan adalah tiga kluster. Walaupun begitu, terlihat bahwa setelah dilakukan plot di atas menggunakan algoritma K-means *clustering*, serupa dengan kasus pada model 3, pelanggan pada sisi kanan plot masih tidak terpisahkan secara jelas antara pelanggan dengan *value* tinggi dan rendah. Karenanya, penulis melakukan klusterisasi kembali dengan empat kluster seperti pada plot di bawah sehingga sesuai dengan ide awal penulis untuk membagi pelanggan menjadi empat kluster.



Sekarang, pelanggan loyal dengan total tagihan yang rendah memiliki *value* yang lebih tinggi dibandingkan pada model sebelumnya, namun masih banyak pelanggan yang terkonsentrasi pada bagian sudut kanan bawah (risiko tinggi, *value* rendah). Oleh karena itu, penulis kembali memutuskan untuk memperbaiki model.

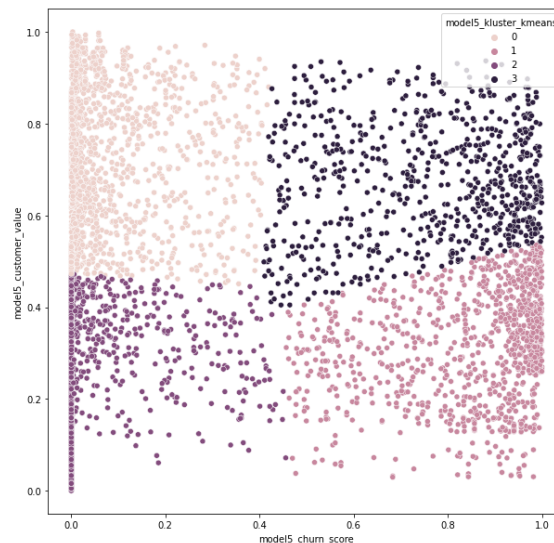
5. Model 5: model *value* 3 dan model risiko 3



Pada model *value* 2, penulis menggunakan total tagihan setiap pelanggan untuk mengukur potensi *revenue* yang dihasilkan oleh seorang pelanggan. Karena total tagihan merupakan kombinasi antara lama berlangganan dengan tagihan bulanan, penulis berpikir bahwa penggunaan total tagihan berpotensi menyebabkan bias pada hasil plot yang dilakukan. Karenanya, penulis kembali menggunakan tagihan bulanan sebagai ukuran potensi *revenue* yang dihasilkan oleh seorang pelanggan, kemudian didapatkan model *value* 3. Pada model 5, penulis mengombinasikan model *value* 3 dan model risiko 3. Berdasarkan perhitungan pada *Elbow method*, terlihat bahwa banyak kluster terbaik yang direkomendasikan adalah tiga kluster. Serupa dengan kasus pada dua model sebelumnya,

terlihat bahwa setelah dilakukan plot di atas menggunakan algoritma K-means *clustering*, pelanggan pada sisi kanan plot masih tidak terpisahkan secara jelas antara pelanggan dengan *value* tinggi dan rendah.

Karenanya, penulis melakukan klasterisasi kembali dengan empat klaster seperti pada plot di bawah sehingga sesuai dengan ide awal penulis untuk membagi pelanggan menjadi empat klaster.



Sekarang, sudah tidak ada konsentrasi pelanggan di sudut kanan maupun sudut kiri plot. Karena hasil plot model 5 sudah sesuai dengan ide awal klasterisasi penulis dan pelanggan sudah relatif lebih tersebar dibandingkan dengan model-model sebelumnya, penulis memutuskan bahwa model 5 merupakan model terbaik dibandingkan model-model sebelumnya. Oleh karena itu, penulis melakukan ekstraksi data pelanggan menggunakan model 5 untuk kemudian diidentifikasi karakteristik pelanggan pada setiap klasternya.

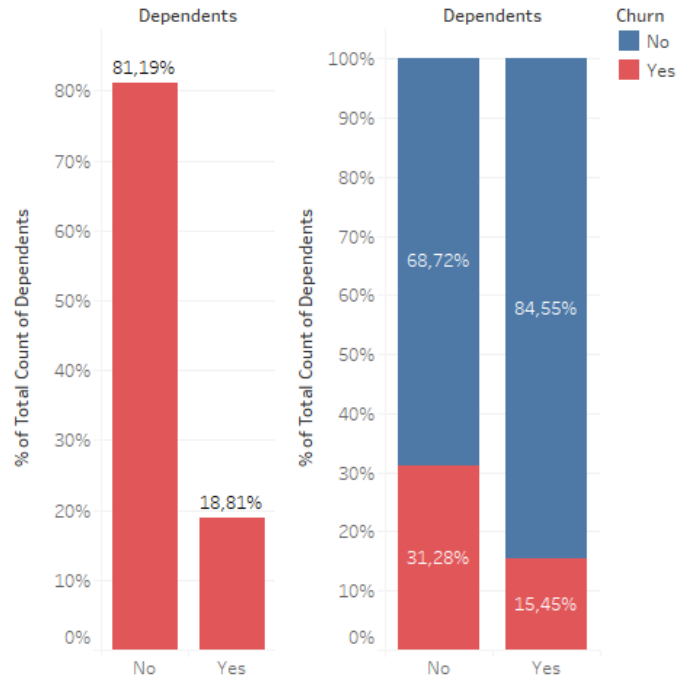
3.3.4 Identifikasi Karakteristik Pelanggan

1. Karakteristik pelanggan klaster 1 (*value* tinggi, risiko tinggi)

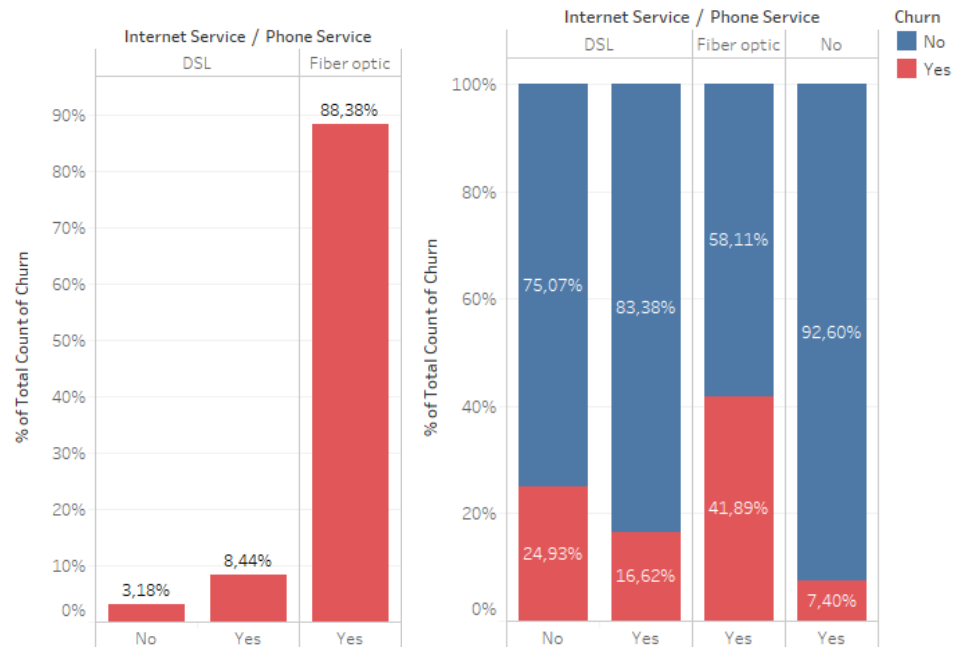
Pada klaster ini, akan dicari beberapa hal yang menyebabkan pelanggan mempunyai risiko yang tinggi

a. Demografi

Penulis menemukan hal menarik dari demografi pelanggan klaster 1. Sebagian besar pelanggan tidak memiliki tanggungan (*dependents*). Jika kita lihat dari *churn rate*-nya, pelanggan yang tidak memiliki tanggungan lebih berisiko untuk berhenti berlangganan dibandingkan dengan pelanggan yang memiliki tanggungan.



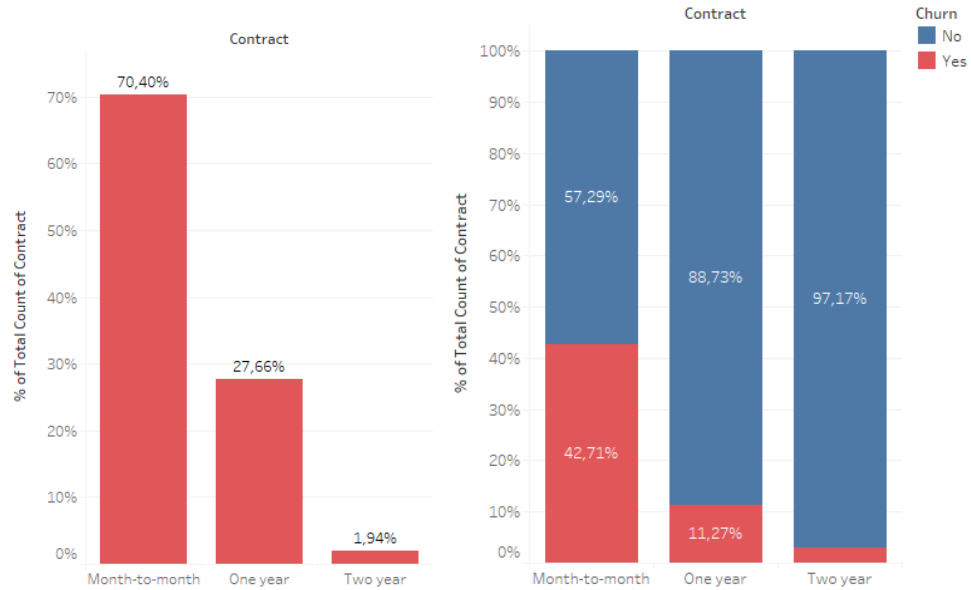
b. Layanan yang digunakan



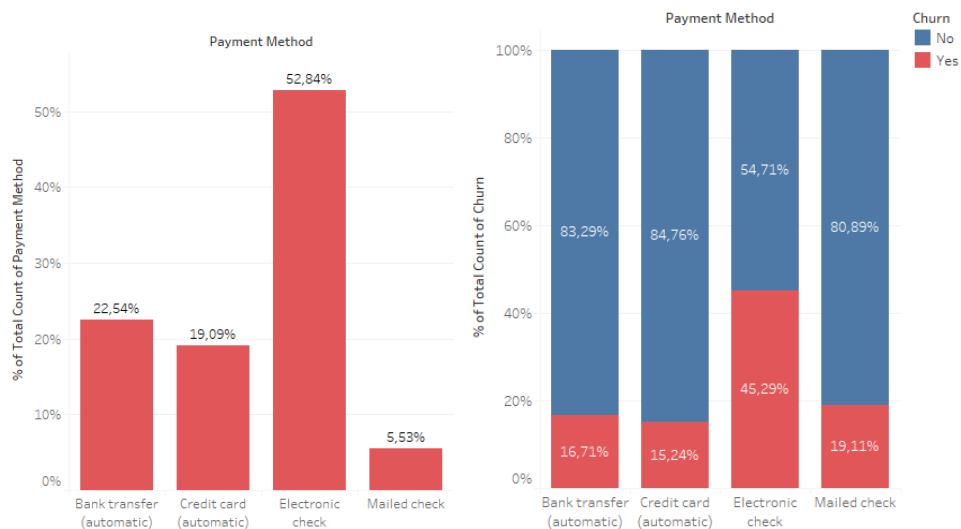
Selain demografi, faktor yang cukup berpengaruh adalah layanan yang digunakan. Pelanggan di klaster 1 sebagian besar menggunakan layanan *fiber optic*, sisanya adalah layanan DSL dan tidak ada yang tidak menggunakan layanan internet. Jika dilihat dari *churn rate* masing-masing layanan, kita dapat melihat bahwa layanan fiber optic adalah layanan dengan tingkat risiko berhenti berlangganan yang paling tinggi.

c. Kontrak

Kontrak juga merupakan variabel yang perlu diperhatikan pada klaster ini. Sebagian besar pelanggan klaster 1 menggunakan kontrak bulanan dan sangat sedikit yang kontraknya dua tahunan, sisanya menggunakan kontrak satu tahunan. Ini menunjukkan bahwa pelanggan di klaster ini memang tidak banyak yang mau berkomitmen jangka panjang. Jika dilihat dari *churn rate*-nya, pelanggan dengan kontrak bulanan sangat berisiko membuat pelanggan berhenti berlangganan.



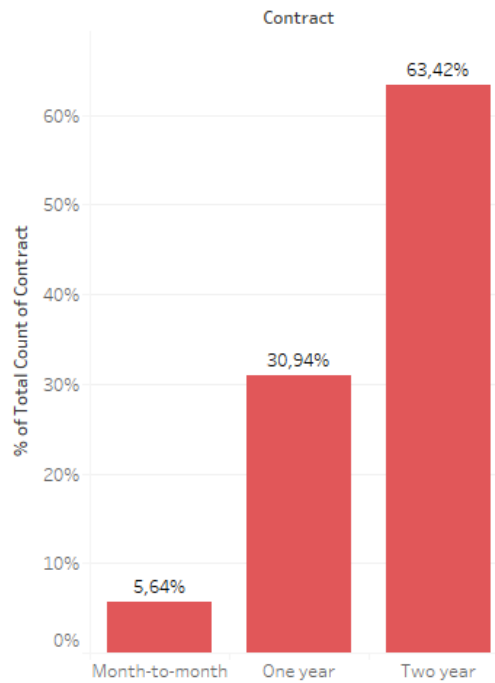
d. Metode pembayaran



Metode pembayaran juga menjadi variabel yang perlu diamati. Dilihat dari *churn rate*-nya, metode pembayaran electronic check menjadi metode yang sangat berisiko membuat pelanggan berhenti berlangganan.

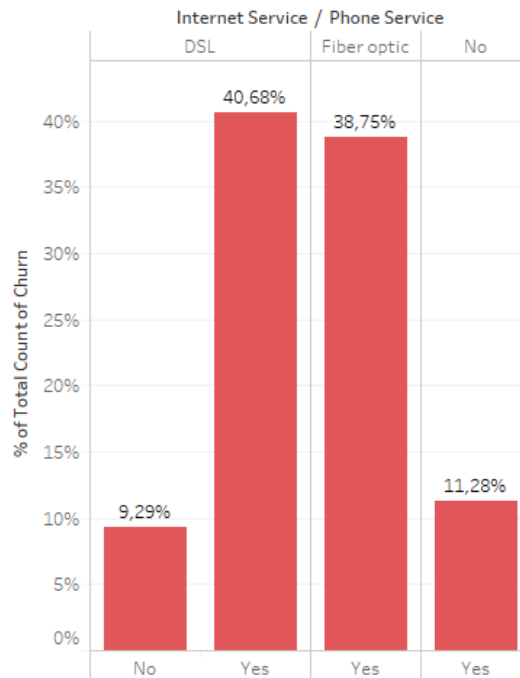
2. Karakteristik pelanggan klaster 2 (*value* tinggi, risiko rendah)

a. Kontrak



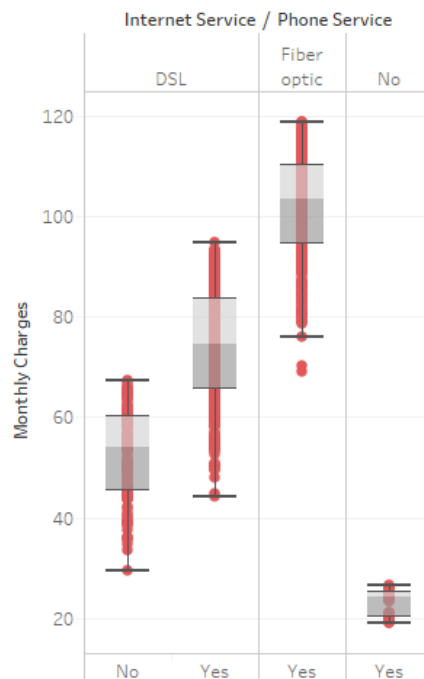
Berdasarkan diagram batang tersebut, terlihat bahwa mayoritas pelanggan pada klaster 2 memiliki kontrak tahunan, bahkan dua tahunan. Hal ini sangat kontras apabila dibandingkan dengan kontrak pelanggan pada klaster 1 sebelumnya. Berdasarkan fakta *churn rate* kontrak sebelumnya, hal ini mengindikasikan bahwa pelanggan pada klaster 2 merupakan pelanggan yang cenderung loyal sehingga perusahaan perlu mempertahankan mereka.

b. Layanan yang digunakan



Berdasarkan layanan yang digunakan, mayoritas pelanggan di klaster 2 telah menggunakan layanan internet, dengan penggunaan tertinggi adalah DSL.

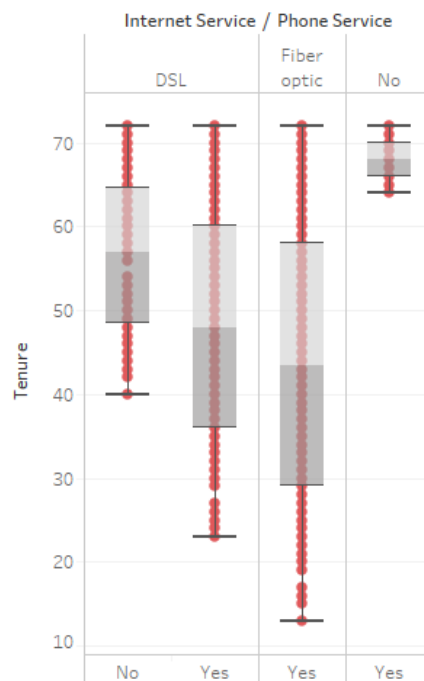
c. Tagihan bulanan



Untuk tagihan bulanan, jelas bahwa pelanggan yang tidak menggunakan layanan internet memiliki rata-rata yang lebih rendah dibandingkan pelanggan pengguna layanan internet. Seperti pada diagram penggunaan layanan sebelumnya, mayoritas pelanggan pada klaster 2 menggunakan layanan internet DSL dan *Fiber optic* sehingga rata-rata tagihan bulanan mereka secara keseluruhan relatif tinggi.

d. Lama berlangganan

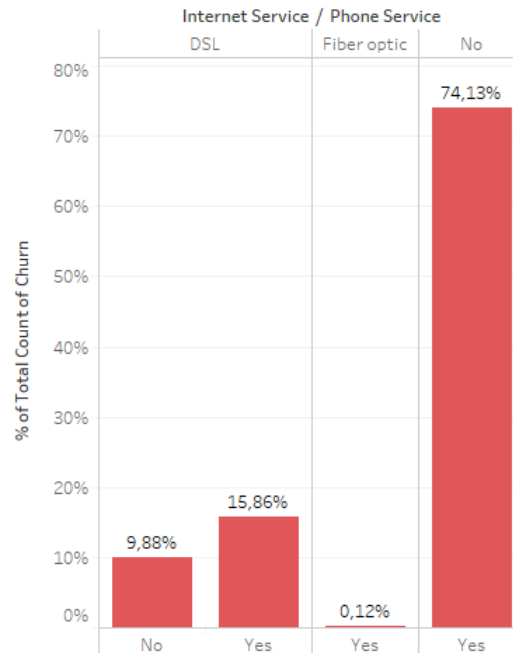
Terlihat bahwa pelanggan pada kluster 2 rata-rata telah relatif lama berlangganan (*tenure*) di perusahaan telekomunikasi yang diteliti. Lalu, teramati pada *box plot* *Fiber optic* terhadap lama berlangganan bahwa *range*-nya cukup tinggi, sebagian sangat loyal dan sebagian lagi belum cukup loyal. Walaupun begitu, penggunaan layanan *Fiber optic* membuat pelanggan yang belum lama berlangganan di perusahaan tetap memiliki *value* tinggi karena potensi *revenue* yang dihasilkan dari mereka cenderung tinggi. Di sisi lain, pelanggan yang tidak menggunakan layanan internet relatif tidak memberikan *revenue* yang signifikan apabila dibandingkan dengan pelanggan pengguna layanan internet. Akan tetapi, terlihat pada *box plot* bahwa pelanggan dengan karakteristik ini telah relatif lama berlangganan di perusahaan sehingga mereka dikategorikan loyal dan memiliki *value* tinggi bagi perusahaan.



3. Karakteristik pelanggan kluster 3 (*value* rendah, risiko rendah)

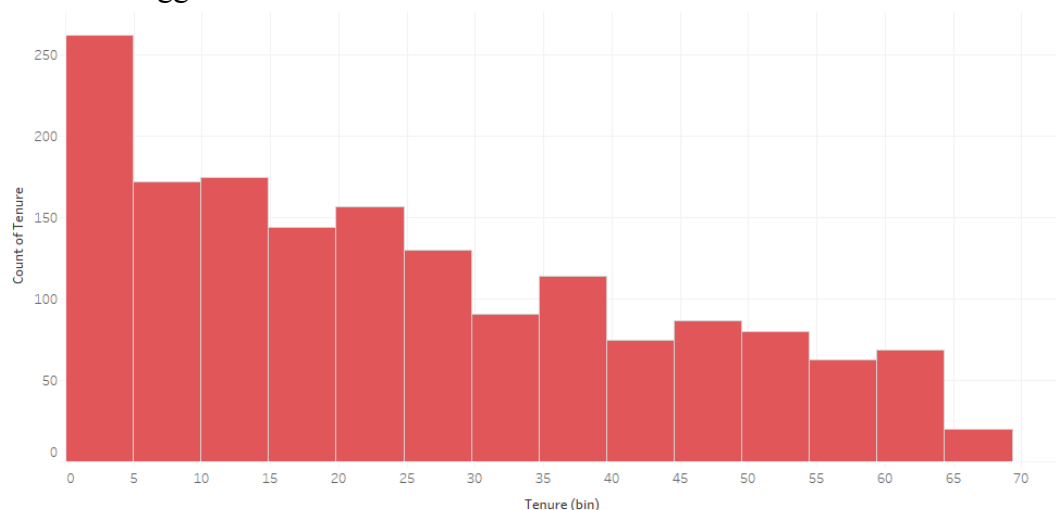
Pada kluster ini, akan dicari mengapa pelanggan memiliki *value* yang rendah.

a. Layanan yang digunakan



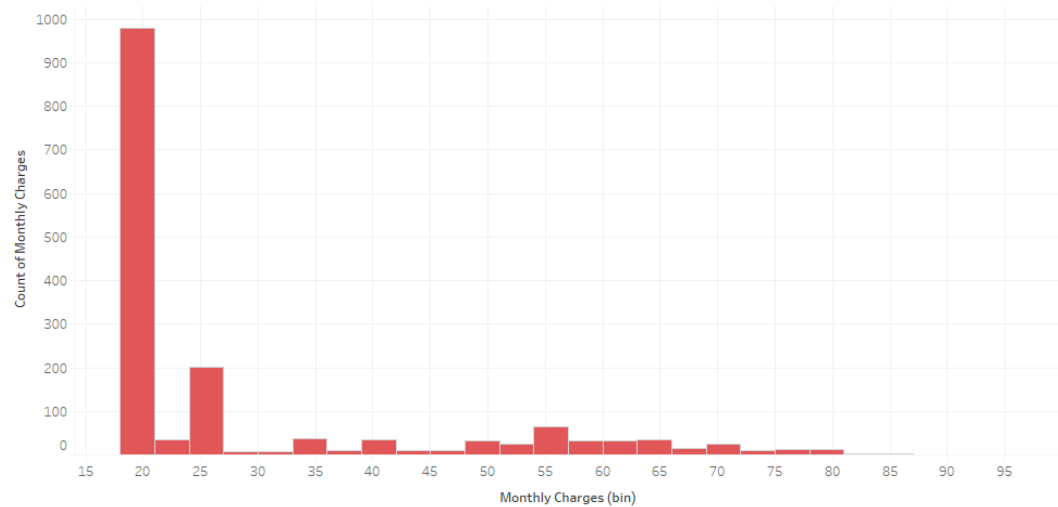
Pelanggan di klaster 3 sebagian besar tidak menggunakan layanan internet, kemudian diikuti dengan pengguna internet DSL dan sangat sedikit yang menggunakan fiber optic.

b. Lama berlangganan

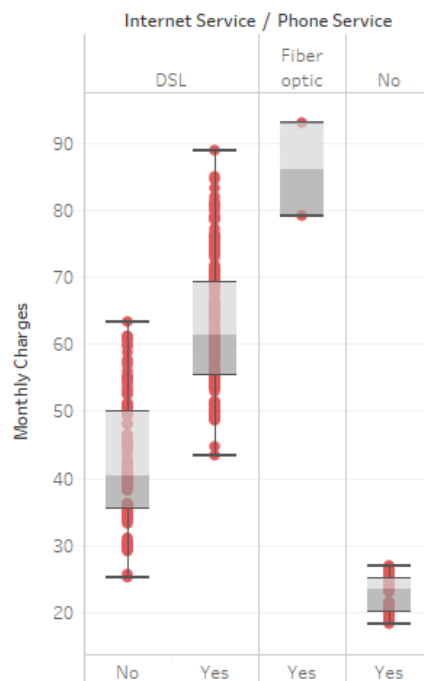


Jika melihat persebaran data lama berlangganan, terlihat bahwa persebarannya cukup menyebar dari yang baru berlangganan beberapa bulan hingga yang sangat lama. Namun, persebaran cukup menumpuk di lama berlangganan yang masih sebentar. Oleh karena itu, lama berlangganan dirasa bukan hal utama yang menyebabkan *valuenya* rendah.

c. Tagihan bulanan



Jika melihat persebaran data tagihan bulanan, jelas terlihat bahwa data sangat terpusat di tagihan bulanan yang rendah. Ini yang menyebabkan value pelanggan menjadi rendah. Kemudian kami ingin melihat perbandingan persebaran tagihan bulanan untuk setiap layanan. Hasilnya seperti pada gambar berikut.



Terdapat perbedaan tagihan bulanan yang cukup signifikan untuk masing-masing layanan. Kita tahu bahwa pelanggan di klaster ini didominasi oleh pelanggan tanpa layanan internet dan sangat sedikit yang menggunakan layanan fiber optic sehingga persebaran datanya sangat terpusat pada tagihan yang rendah.

3.3.5 Skema Pemberian Bonus

Berdasarkan karakteristik pelanggan yang diperoleh, penulis merumuskan skema pemberian bonus yang berbeda untuk setiap klaster dengan rincian sebagai berikut.

1. Skema bonus klaster 1

Tujuan:

- a. Membuat pelanggan lebih loyal dengan memberikan diskon pada layanan yang dipakai
- b. Membuat pelanggan dengan kontrak bulanan berpindah menjadi kontrak tahunan
- c. Membuat pelanggan dengan metode pembayaran *electronic check* menjadi *credit card*/bank transfer

Skema:

- a. Pemberian potongan tagihan bulanan untuk setiap pelanggan
- b. Penawaran diskon tambahan untuk pelanggan dengan kontrak bulanan supaya berpindah menjadi tahunan
- c. Penawaran diskon tambahan untuk pelanggan dengan metode pembayaran *electronic check* supaya berganti menjadi *credit card*/bank transfer

Contoh penerapan:

- a. Potongan tagihan bulanan sebesar 20% untuk setiap pelanggan
- b. Diskon tambahan sebesar 20% untuk pelanggan dengan kontrak bulanan yang mau berganti menjadi tahunan
- c. Diskon tambahan sebesar 40% untuk pelanggan dengan metode pembayaran *electronic check* yang mau berganti menjadi *credit card*/bank transfer

Contoh tersebut akan memberikan hasil sebagai berikut.

Kasus 1: Tidak ada pelanggan dengan kontrak bulanan yang menerima penawaran bonus menjadi tahunan sehingga perusahaan berpotensi mengalami penurunan *revenue*, maksimum sebesar 1.7% atau sekitar \$53,688.8.

Kasus 2: Semua pelanggan menerima penawaran bonus sehingga perusahaan berpotensi mengalami kenaikan *revenue*, maksimum sebesar 4.75% atau sekitar \$150,113.1.

2. Skema bonus klaster 2

Tujuan:

- a. Mempertahankan loyalitas pelanggan dengan memberikan potongan tagihan bulanan pada layanan yang dipakai

Skema:

- a. Pemberian potongan tagihan bulanan untuk setiap pelanggan pada perpanjangan kontrak berikutnya

Contoh penerapan:

- a. Potongan tagihan bulanan sebesar 20% untuk setiap pelanggan pada kontrak berikutnya

Contoh tersebut akan memberikan hasil sebagai berikut.

Perusahaan memberikan bonus kepada seluruh pelanggan di klaster 2 sehingga tidak ada penambahan *revenue*, sebaliknya, perusahaan diproyeksikan mengalami penurunan *revenue* sebesar 2.83% atau sekitar \$270,575.

3. Skema bonus klaster 3

Tujuan:

- Membuat pelanggan menambah layanan internet jika belum berlangganan internet
- Membuat pelanggan menambah layanan telepon jika belum berlangganan telepon

Skema:

- Penawaran diskon berlangganan internet untuk pelanggan yang belum berlangganan internet
- Penawaran diskon berlangganan telepon untuk pelanggan yang belum berlangganan telepon

Contoh penerapan:

- Diskon berlangganan internet sebesar 20% untuk pelanggan yang belum berlangganan internet
- Diskon berlangganan telepon sebesar 20% untuk pelanggan yang belum berlangganan telepon

Contoh tersebut akan memberikan hasil sebagai berikut.

Jika seluruh pelanggan yang mendapat penawaran bonus menerima penawaran, perusahaan berpotensi mengalami kenaikan *revenue*, maksimum sebesar 9.55% atau sekitar \$138,812.4.

Berdasarkan skema-skema tersebut, tabel berikut menyajikan hasil klasterisasi beserta pemberian bonus dari sampel sebanyak dua puluh pelanggan perusahaan telekomunikasi yang diteliti.

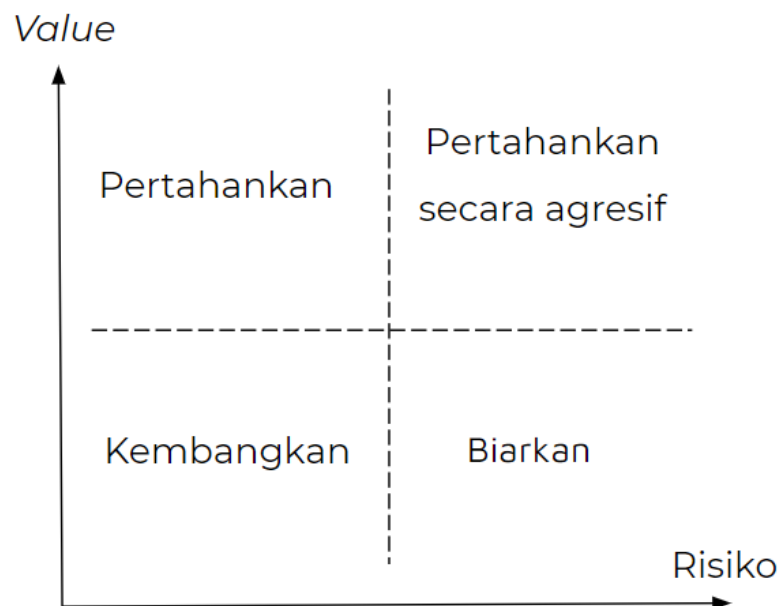
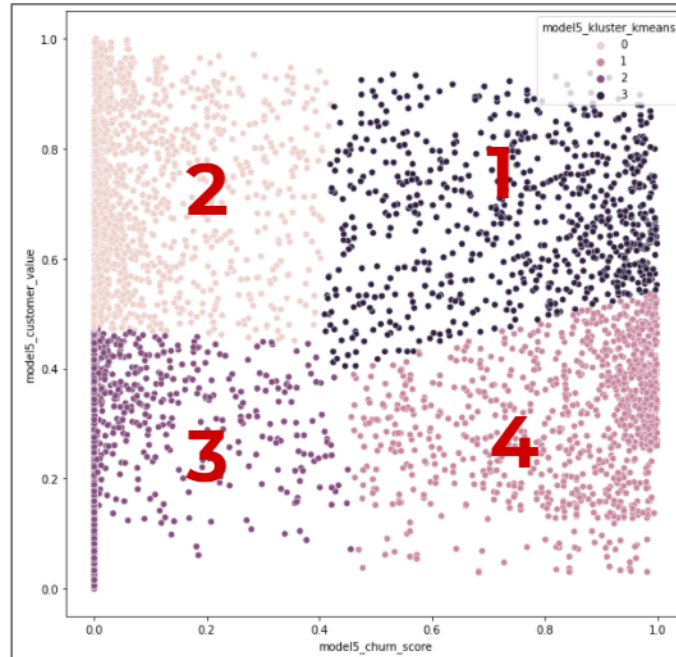
Cluster	CustomerID	Dependents	Tenure	Phone Service	Internet Service	Contract	Payment Method	Monthly Charges	Bonus
1	5129-JLPIS	No	25	Yes	Fiber optic	Month-to-month	Electronic check	105,5	Potongan tagihan bulanan, penawaran diskon pergantian metode pembayaran, penawaran diskon pergantian kontrak
	3714-NTNFO	No	49	Yes	Fiber optic	Month-to-month	Electronic check	84,5	
	6572-ADKRS	No	46	Yes	Fiber optic	Month-to-month	Credit card (auton	74,8	Potongan tagihan bulanan, penawaran diskon pergantian kontrak
	6575-SUVOI	No	25	Yes	DSL	Month-to-month	Credit card (auton	69,5	
	3070-BDOQC	No	60	Yes	Fiber optic	Two year	Electronic check	99,65	Potongan tagihan bulanan, penawaran diskon pergantian metode pembayaran
2	6388-TABGU	Yes	62	Yes	DSL	One year	Bank transfer (aut	56,15	Potongan tagihan bulanan
	8091-TTVAX	No	58	Yes	Fiber optic	One year	Credit card (auton	100,35	
	3655-SNQYZ	Yes	69	Yes	Fiber optic	Two year	Credit card (auton	113,25	
	9959-WOFKT	Yes	71	Yes	Fiber optic	Two year	Bank transfer (aut	106,7	
	3638-WEABW	No	58	Yes	DSL	Two year	Credit card (auton	59,9	
3	7469-LKBCI	No	16	Yes	No	Two year	Credit card (auton	18,95	Penawaran diskon layanan internet
	8191-XWSZG	No	52	Yes	No	One year	Mailed check	20,65	Penawaran diskon layanan telepon
	7795-CFOCW	No	45	No	DSL	One year	Bank transfer (aut	42,3	
	5575-GNVDE	No	34	Yes	DSL	One year	Mailed check	56,95	Tidak mendapat bonus
	9763-GRSKD	Yes	13	Yes	DSL	Month-to-month	Mailed check	49,95	
4	7590-VHVEG	No	1	No	DSL	Month-to-month	Electronic check	29,85	Tidak mendapat bonus
	1452-KIOVK	Yes	22	Yes	Fiber optic	Month-to-month	Credit card (auton	89,1	
	6713-OKOMC	No	10	No	DSL	Month-to-month	Mailed check	29,75	
	4183-MYFRB	No	21	Yes	Fiber optic	Month-to-month	Electronic check	90,05	
	6865-JZDKO	No	30	Yes	DSL	Month-to-month	Bank transfer (aut	55,3	

BAB IV KESIMPULAN DAN SARAN

4.1 Kesimpulan

Berdasarkan hasil yang diperoleh, penulis mendapatkan kesimpulan sebagai berikut.

1. Pelanggan yang layak untuk mendapatkan bonus adalah pelanggan yang berada di kluster 1, 2, dan 3.



2. Skema pemberian bonus untuk masing-masing klaster adalah
 - a. Klaster 1.
 - 1) Pemberian potongan tagihan bulanan untuk setiap pelanggan.
 - 2) Penawaran diskon tambahan untuk pelanggan dengan kontrak berlangganan bulanan supaya berpindah menjadi tahunan.
 - 3) Penawaran diskon tambahan untuk pelanggan dengan metode pembayaran *electronic check* supaya berganti menjadi *credit card*/bank transfer.
 - b. Klaster 2.
 - 1) Pemberian potongan tagihan bulanan untuk setiap pelanggan pada kontrak berikutnya.
 - c. Klaster 3.
 - 1) Penawaran diskon berlangganan layanan internet untuk pelanggan yang belum berlangganan internet.
 - 2) Penawaran diskon berlangganan layanan telepon untuk pelanggan yang belum berlangganan telepon.

4.2 Saran

Dengan hasil yang diperoleh, penelitian ini masih memiliki keterbatasan dan dapat dikembangkan lagi. Saran yang dapat penulis berikan adalah sebagai berikut.

1. Melengkapi data yang ada agar model lebih akurat.
2. Mengembangkan model agar tidak bergantung pada asumsi yang kurang realistis.
3. Mencari referensi strategi bisnis yang dapat diterapkan secara real.
4. Menggunakan analisis model kesintasan (*survival analysis*) yang dapat memprediksi peluang loyalitas pelanggan hingga beberapa waktu kedepan, tidak hanya pada keadaan sekarang.

DAFTAR PUSTAKA

- Brownlee, J. 2020. *How to Use StandardScaler and MinMaxScaler Transforms in Python*. Diakses tanggal 25 Mei 2021 pukul 14.30 WIB, dari <https://machinelearningmastery.com/standardscaler-and-minmaxscaler-transforms-in-python/>
- Brownlee, J. 2019. *A Gentle Introduction to Bayes Theorem for Machine Learning*. Diakses tanggal 23 Mei 2021 pukul 18.15 WIB, dari <https://machinelearningmastery.com/bayes-theorem-for-machine-learning/>
- Chauhan, N. 2020. *Naïve Bayes Algorithm: Everything you need to know*. Diakses tanggal 28 Maret 2021 pukul 14.15 WIB, dari <https://www.kdnuggets.com/2020/06/naive-bayes-algorithm-everything.html>
- Hill, N., Roche, G. & Allen, R. 2007. *Customer Satisfaction: The customer experience through the customer's eyes*. London: Cogent Publishing Ltd.
- Khadka & Maharjan. 2017. *Customer Satisfaction and Customer Loyalty*. Jakobstad: Centria University of Applied Sciences Pietarsaari.
- L. Al Shalabi & Z. Shaaban. 2006. *Normalization as a Preprocessing Engine for Data Mining and The Approach of Preference Matrix*. Proc International Conference in Dependability of Computer Systems.