**DETAIL FINAL PROJECT**

**SIB KAMPUS MERDEKA - DATA ANALYTICS CYCLE 3**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| M. Alwansyah Mardika | - | Informatika | - | Universitas Gunadarma |
| Ayu Nur Baiti | - | Ilmu Ekonomi | - | Universitas Padjadjaran |
| Andesta Wistara S | - | Ekonomi Pembangunan | - | Universitas Sebelas Maret |
| Elsha Nansa Abigail | - | Teknik Kelautan | - | Institut Teknologi Sepuluh Nopember |

**JB Link Telco Customer Churn Problem**

* Dataset: <https://www.kaggle.com/datasets/johnflag/jb-link-telco-customer-churn>
* Link Google Slide:

<https://docs.google.com/presentation/d/1i4t9RMybksW-gqQr-Oe3NnpGIeKpma32/edit#slide=id.p1>

* Link Google Colab:

<https://drive.google.com/file/d/1XYWhdmCdxz7Gdb-ZNFeEWgBsGpAIpgC_/view?usp=sharing>

* Data Understanding:

<https://docs.google.com/spreadsheets/d/1IHGE2Ir7CpO9zv2iGNGtDglaM-ZHfFsqHiTNXFHckmw/edit?usp=sharing>

**BUSINESS UNDERSTANDING**

Sektor telekomunikasi telah menjadi salah satu industri utama di negara maju. Kemajuan teknis dan bertambahnya jumlah operator dapat meningkatkan tingkat persaingan. Banyak perusahaan telekomunikasi seluler menghadapi lingkungan bisnis yang sangat menantang karena pasar sudah jenuh. Akibatnya, retensi *customer* adalah salah satu masalah yang paling penting untuk menghasilkan pendapatan dan margin yang lebih tinggi karena mempertahankan *customer* lebih murah dan lebih menguntungkan daripada menarik *customer* baru (Euler, 2005).

JB Link adalah perusahaan telekomunikasi ukuran kecil yang berlokasi di negara bagian California yang menyediakan layanan Telepon dan Internet kepada *customer* di lebih dari 1.000 kota dan 1.600 kode pos. Perusahaan memiliki jumlah *customer* baru yang diperoleh pada kuartal terakhir mewakili 15% dari total. Namun, pada akhir periode yang sama, hanya 43% dari *customer* ini yang bertahan dengan perusahaan dan kebanyakan dari mereka memutuskan untuk tidak memperbarui kontrak mereka setelah beberapa bulan, yang berarti tingkat pemutusan *customer* sangat tinggi dan perusahaan sekarang menghadapi tantangan besar dalam mempertahankan *customer*nya. Total *customer churn* rate kuartal terakhir adalah sekitar 27%, menghasilkan penurunan hampir 12% dari total jumlah *customer*.

*Customer churn* menjadi perhatian yang cukup besar di sektor jasa dengan layanan kompetitif yang tinggi. Di sisi lain, memprediksi *customer* yang kemungkinan besar meninggalkan perusahaan akan mewakili sumber pendapatan tambahan yang berpotensi besar. Proses *customer churn* merupakan aktivitas pelanggan beralih dari penyedia layanan yang ada ke penyedia layanan lain dalam industri yang sama. Alasan yang cukup besar untuk aktivitas *customer* *churning* adalah ketidakpuasan dengan penyedia layanan jaringan yang ada dan daya tarik penyedia layanan jaringan lainnya. Pentingnya memprediksi detail *customer* yang bersedia melakukan *churn* dari penyedia jaringan tertentu. Karena berlanjutnya aktivitas *customer churning* akan menurunkan tingkat keuntungan dan juga dapat merusak reputasi operator jaringan tersebut dengan mudah.

**Rumusan Masalah**

1. Berapa total seluruh pelanggan JB Link dan berapa pelanggan yang berhenti berlangganan?
2. Mengapa perusahaan JB Link kehilangan pelanggan dan faktor apa saja yang mempengaruhi perusahaan kehilangan pelanggan?
3. Bagaimana upaya dan strategi perusahaan untuk meningkatkan jumlah pelanggan dan menjadi pelanggan tetap?

**Tujuan**

1. Untuk mengetahui jumlah pelanggan yang berhenti berlangganan dan mengapa perusahaan bisa terjadi kehilangan pelanggan.
2. Untuk memprediksi pelanggan yang berlangganan dan menjadi pelanggan tetap.
3. Untuk memberikan solusi strategi kepada perusahaan JB Link dapat meningkatkan jumlah pelanggan dan inovasi baru pada teknologi.

**Target**

* Menurunkan *Customer Churn Rate*, meminimalisir kerugian akibat dari *customer churn* dan membuat perusahaan mampu bersaing dengan kompetitor lain

**Hipotesis**

1. *Customer churn* karena pelayanan yang buruk dan tidak ada perkembangan di bagian inovasi dan teknologi.
2. *Customer churn* dengan tenure in month yang rendah lebih kecenderungan *churn*.
3. Tingkat *monthly charge* yang rendah mempengaruhi kecenderungan *customer churn.*
4. *Contract* pemilihan *contract* mempengaruhi tingkat *churn*.

**DATA UNDERSTANDING**

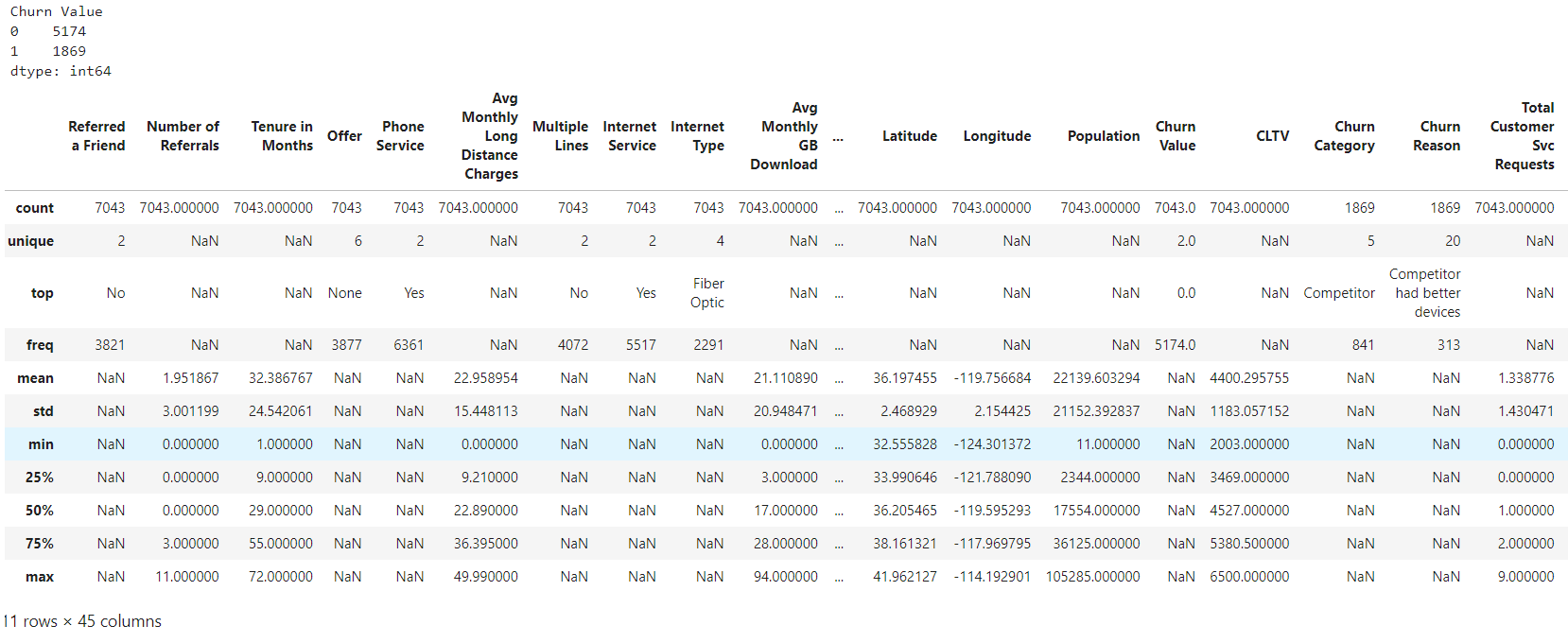
Tim Ilmu Data diberi kumpulan data dengan sampel acak 7.043 *customer* yang dapat membantu mencapai tugas ini. Menggunakan dataset *telco\_churn\_data.* Data juga berisi mengenai demografi *customer*. Para eksekutif menyadari bahwa biaya untuk mendapatkan *customer* baru dapat mencapai lima kali lebih tinggi daripada biaya mempertahankan *customer*, sehingga mereka berharap bahwa hasil dari proyek ini akan menghemat banyak uang bagi perusahaan dan memulainya tumbuh lagi.

**Informasi Data**

* Total pelanggan = 7043
* *Customer churn* = 57% , 7043 x 57% = 4.014,51
* Pelanggan yang bertahan = 43% , 7043 x 43% = 3.028,49

|  |  |  |  |
| --- | --- | --- | --- |
| **No** | **Column Name** | **Type** | **Description** |
| 1 | Customer ID | string | ID setiap pelanggan yang bersifat unik, berbeda dari setiap pelanggan. |
| 2 | Referred a Friend | String | Menampilkan apakah pelanggan memberikan rekomendasi kepada teman atau keluarga nya. |
| 3 | Number of Referrals | Integer | Menampilkan jumlah pelanggan memberikan rekomendasi. |
| 4 | Tenure in Months | Integer | Menampilkan jumlah total bulan pelanggan berlangganan dengan perusahaan. |
| 5 | Offer | String | Menampilkan pilihan paket atau pesanan yang digunakan pelanggan. |
| 6 | Phone Service | String | Menampilkan pelanggan yang menggunakan layanan telpon. |
| 7 | Avg Monthly Long Distance Charges | Integer | Rata-rata biaya jangka panjang bulanan. |
| 8 | Multiple Lines | String | Beberapa baris. |
| 9 | Internet Service | String |  |
| 10 | Internet Type | String | Menampilkan tipe internet yang digunakan pelanggan. |
| 11 | Avg Monthly GB Download | Integer | Menampilkan rata-rata penggunaan internet dalam gb perbulan. |
| 12 | Online Security | String | Menampilkan pelanggan yang menggunakan layanan keamanan online. |
| 13 | Online Backup | String | Menampilkan pelanggan yang menggunakan bantuan secara online. |
| 14 | Device Protection Plan | String | Menunjukan perlindungan perangkat kepada konsumen. |
| 15 | Premium Tech Support | String | Menunjukan dukungan teknis kepada pelanggan. |
| 16 | Streaming TV | String | Menunjukan layanan menonton acara televisi. |
| 17 | Streaming Movies | String | Menunjukan layanan menonton film. |
| 18 | Streaming Music | String | Menunjukan layanan mendengarkan musik. |
| 19 | Unlimited Data | String | Layanan data internet yang tidak terbatas. |
| 20 | Contract | String | Perjanjian antara dua pihak atau lebih. |
| 21 | Paperless Billing | String | Menunjukan sistem tagihan tanpa kertas. |
| 22 | Payment Method | String | Menunjukan metode pembayaran. |
| 23 | Monthly Charge | Integer | Menunjukan biaya bulanan untuk analisis deskriptif. |
| 24 | Total Regular Charges | Integer | Menunjukkan total biaya reguler pelanggan, tidak termasuk biaya tambahan. |
| 25 | Total Refunds | Integer | Menunjukkan total pengembalian uang pelanggan. |
| 26 | Total Extra Data Charges | Integer | Menunjukkan total biaya pelanggan untuk unduhan data tambahan di atas yang ditentukan dalam paket mereka. |
| 27 | Total Long Distance Charges | Integer | Menunjukkan total biaya yang berlaku untuk panggilan telepon yang dilakukan ke lokasi di luar area panggilan lokal yang ditentukan.. |
| 28 | Gender | String | Menunjukkan total biaya pelanggan untuk jarak jauh di atas yang ditentukan dalam paket mereka. |
| 29 | Age | Integer | Usia pelanggan saat ini. |
| 30 | Under 30 | String | Menunjukkan jika pelanggan berusia dibawah 30 tahun. |
| 31 | Senior Citizen | String | Menunjukkan jika pelanggan berusia 65 tahun ke atas. |
| 32 | Married | String | Menunjukkan jika pelanggan sudah menikah. |
| 33 | Dependents | String | Menunjukkan jika pelanggan tinggal dengan tanggungan: Ya, Tidak. Tanggungan bisa anak-anak, orang tua, kakek-nenek, dll. |
| 34 | Number of Dependents | Integer | Menunjukkan jumlah tanggungan yang tinggal bersama pelanggan. |
| 35 | City | String | Menunjukan kota pelanggan. |
| 36 | Zip Code | Integer | Menunjukan kode zip. |
| 37 | Latitude | Integer | Menunjukan garis lintang di permukaan bumi. |
| 38 | Longitude | Integer | Menunjukan garis bujur di permukaan bumi. |
| 39 | Population | Integer | Menunjukan jumlah populasi. |
| 40 | Churn Value | Integer | Kondisi dimana pelanggan meninggalkan atau tetap bersama perusahaan. |
| 41 | CLTV | Integer | Nilai Seumur Hidup Pelanggan. |
| 42 | Churn Category | String | Kategori tingkat tinggi untuk alasan pelanggan melakukan pengadukan. |
| 43 | Churn Reason | String | Menunjukan alasan khusus pelanggan meninggalkan perusahaan. |
| 44 | Total Customer Svc Requests | Integer | Menunjukan jumlah pelanggan menghubungi layanan pelanggan. |
| 45 | Product/Service Issues Reported | Integer | Menunjukan peringkat kepuasan pelanggan perusahaan secara keseluruhan. |
| 46 | Customer Satisfaction | Integer | Menunjukan tingkat kepuasan pelanggan. |

**Mendeskripsikan data *Churn Value* dan data pada kolom lain:**



Analisis gambar :

Menampilkan jumlah *churn value,* yang bernilai 0 itu artinya tidak *churn* berjumlah 5174, dan yang bernilai 1 itu artinya *churn* berjumlah 1869. Menampilkan nilai *count, unique, top, freq, mean, std, min,* 25% adalah nilai kuartil pertama, 50% adalah nilai kuartil kedua, 75% adalah nilai kuartil ketiga, dan *max* pada setiap kolom pada dataset yang digunakan. Jika bernilai nan maka data dari kolom itu bertipe data *object* atau *string.*

**Data Visualisasi Pilihan:**

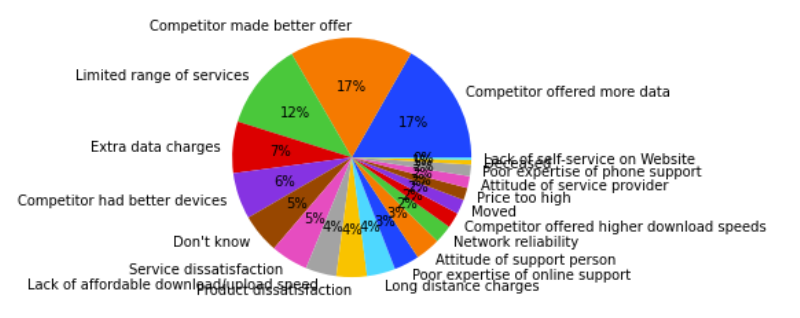
* ***Internet Type***

|  |  |
| --- | --- |
|  | Diagram menunjukan pelanggan yang tidak *churn* paling banyak menggunakan internet tipe *fiber optic* dan yang paling sedikit menggunakan kabel. Untuk pelanggan yang *churn* itu paling banyak menggunakan kabel dan yang paling sedikit *none.* |

* **Diagram *Contract***

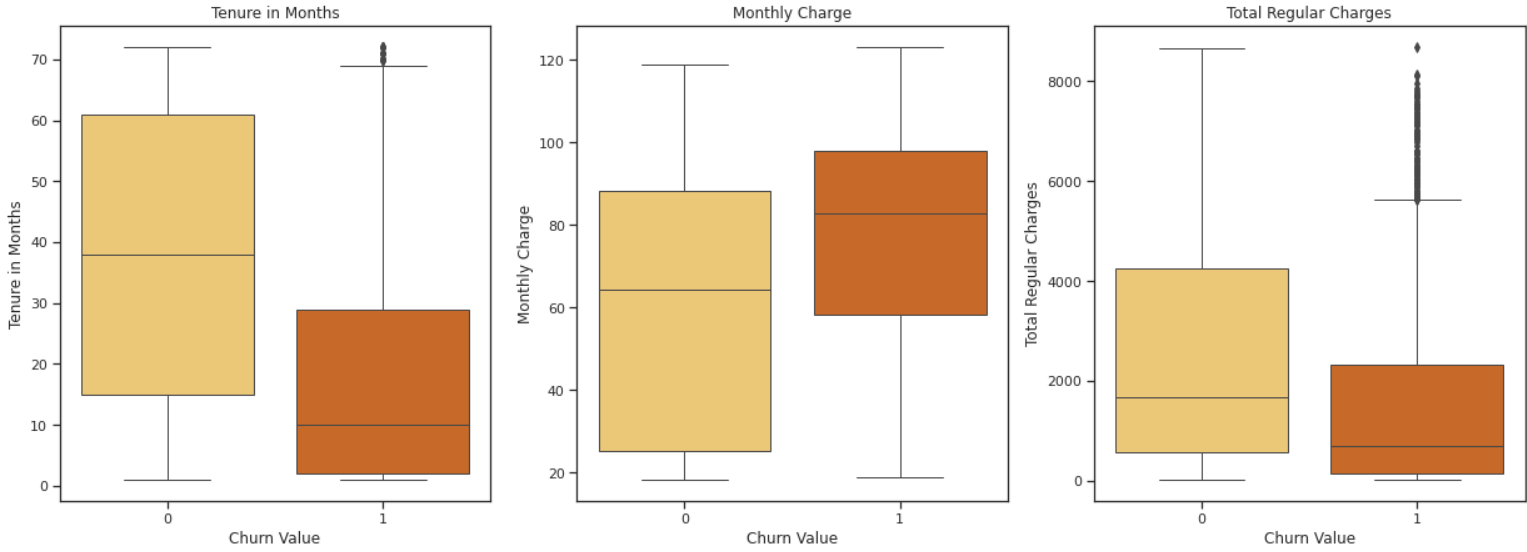
|  |  |
| --- | --- |
|  | Pada diagram tersebut dapat dilihat bahwa pelanggan yang tidak melakukan *churn* didominasi oleh pelanggan yang memiliki kontrak perbulan dan kontrak per 2 tahun. Sedangkan untuk pelanggan yang melakukan *churn* didominasi sebesar lebih dari 1500 pelanggan yang melakukan kontrak bulanan. |

* ***Churn Reason***



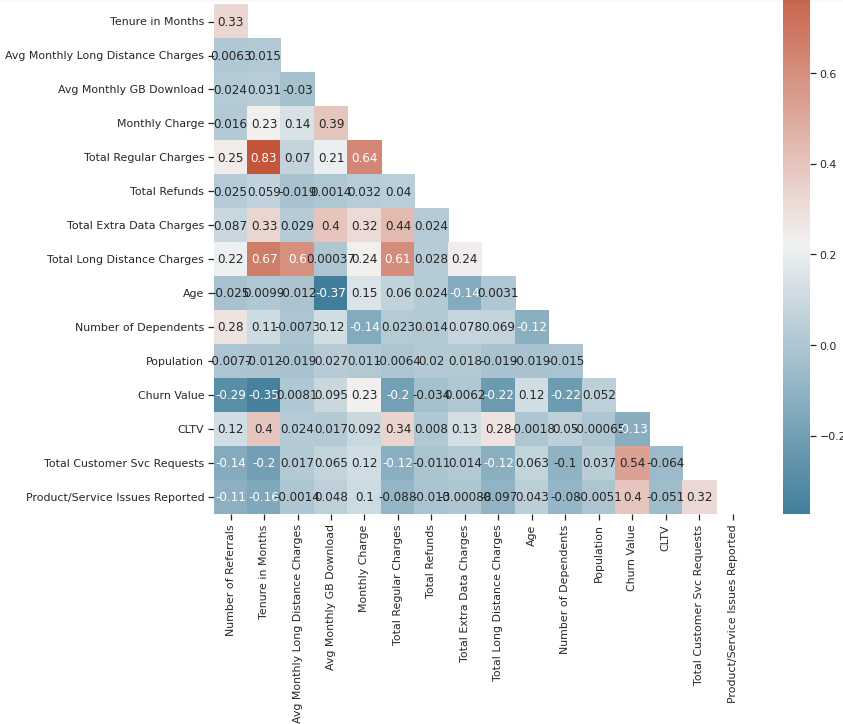
|  |
| --- |
| Dari gambar diatas dapat dilihat persentase alasan *customer* melakukan *churn*, alasan tersebut cenderung beragam dan berbeda beda. |

* ***Tenure in Months, Monthly Charge, Total Regular Charges***

****

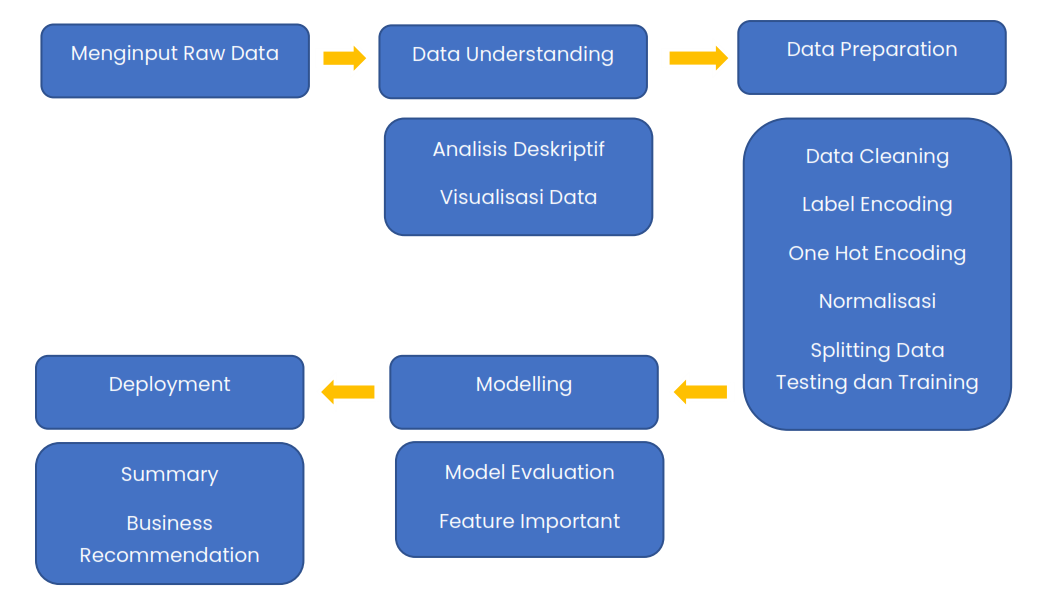
Pada gambar tenure in month, *customer* yang churn jumlah *tenure in months* *customer* lebih rendah daripada *customer* yang tidak *churn* perbedaan yang sangat jauh sehingga terlihat jelas, hipotesis karena pelanggan yang *churn* itu hanya berlangganan beberapa bulan saja. Pada *monthly charge* pelanggan yang *churn* lebih tinggi biaya setiap bulannya daripada pelanggan yang tidak *churn,* hipotesis karena pelanggan menggunakan untuk kegiatan yang banyak. Pada total *reguler charges* pelanggan yang *churn* lebih rendah yang menggunakan layanan biasa daripada yang tidak *churn*, hipotesis karena pelanggan yang *churn* memiliki kebutuhan yang banyak jadi lebih banyak yang menggunakan layanan yang lebih.

* **Korelasi antar kolom**

****

Terdapat korelasi yang tinggi antara *Total Regular Charges* dengan *Tenure in Months* sebesar 0.83 hampir mendekati 1 artinya jika *Total Regular Charges* tinggi, maka *Tenure in Months* juga ikut tinggi dan kedua kolom ini saling berpengaruh. Dan korelasi yang rendah ada pada *Age* dengan *Avg Monthly Long Distance Charges* sebesar -0.37 artinya semakin tinggi umur pelanggan semakin sedikit juga penggunaan layanan bulanannya.

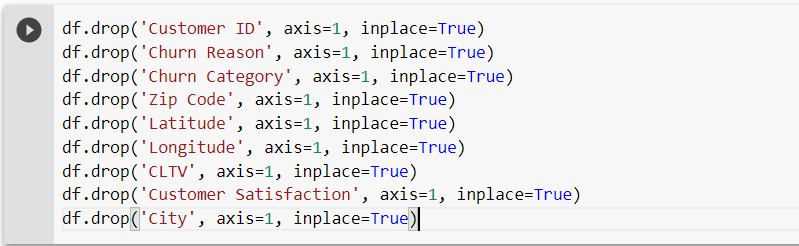
**DATA PREPARATION**

****

Data *preprocessing* merupakan sekumpulan teknik yang diterapkan pada database untuk menghapus *noise, missing value,* dan data yang tidak konsisten. Data *pre processing* dibagi menjadi beberapa langkah, yaitu *cleaning data*, data transformation, dan data reduction. Data *pre processing* yang akan dicakup dalam analisis ini ialah:

* *Data cleaning*

*Data cleaning* berarti menghilangkan informasi yang tidak koheren dari database untuk meningkatkan keseragaman data seperti melakukan *drop duplicates,* melakukan *fill* terhadap null value dan lain lain.

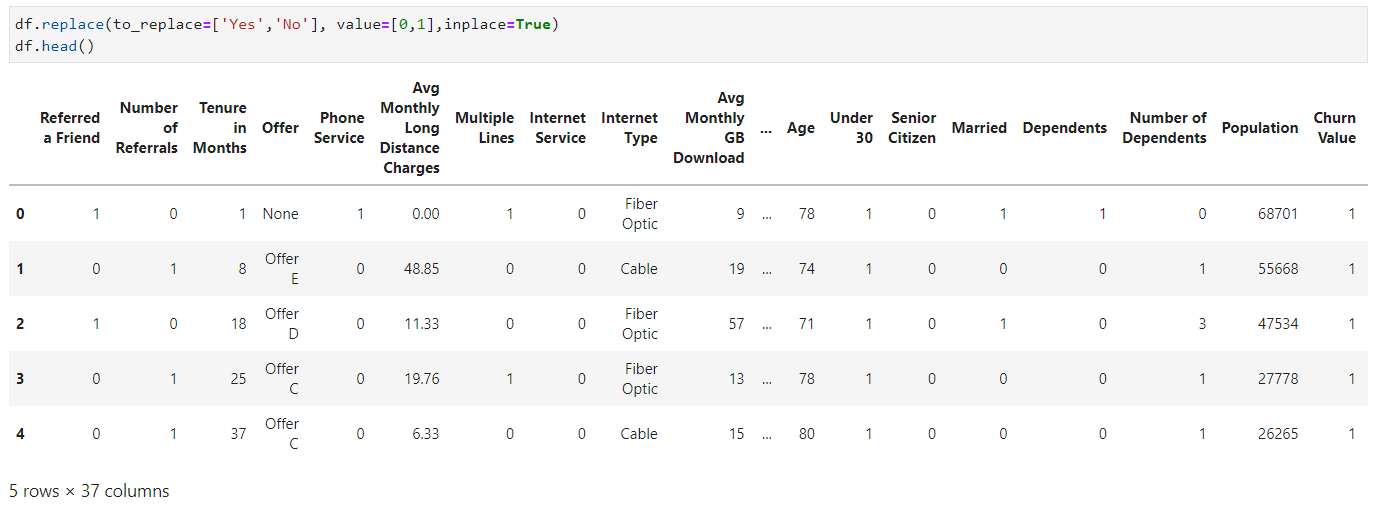


* *Label Encoding*

Mengganti nilai kolom *gender* dari yang berisi *male/female* bertipe data *object* menjadi 0/1 bertipe data *integer*



Mengganti nilai kolom dari yang berisi *yes/no* bertipe data *object* menjadi 0/1 bertipe data *integer*

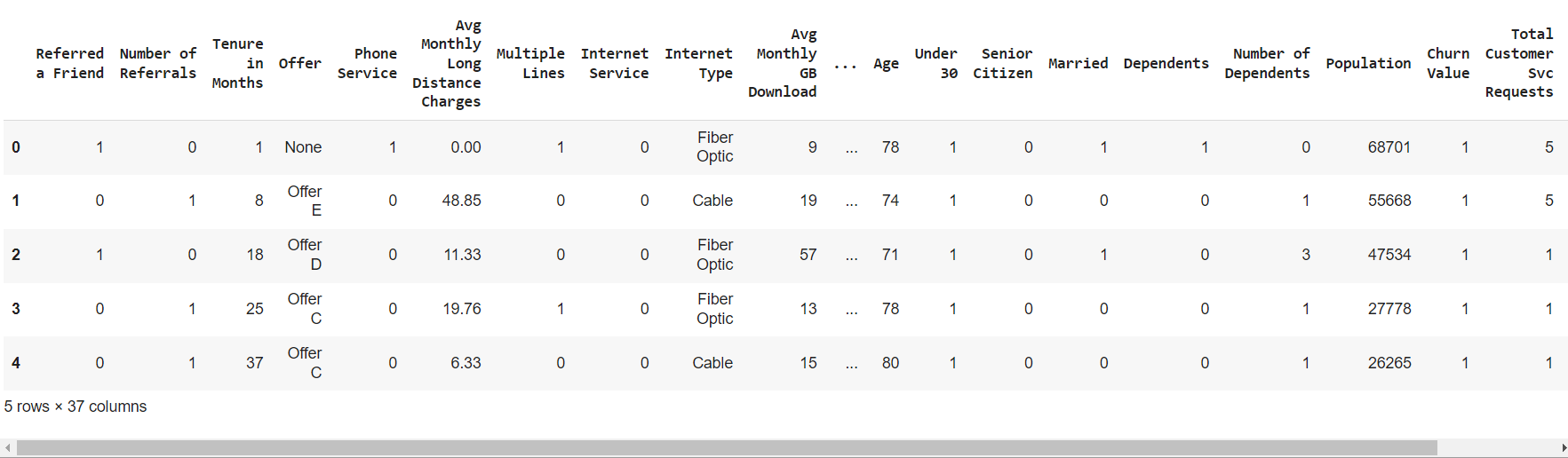


Mengganti nilai kolom *churn value* dari yang isinya 0/1 bertipe data *integer* menjadi *yes/no* bertipe data *object*



* *One Hot Encoding*

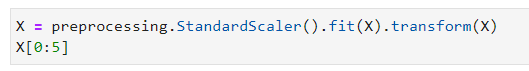
*One Hot Encoding* adalah proses untuk membuat kolom baru dari variabel kategorikal di mana setiap kategori menjadi kolom baru dengan nilai 0 atau 1 (0 mewakili tidak ada dan 1 mewakili ada). Kolom yang kami gunakan adalah '*Referred a Friend', 'Number of Referrals', 'Tenure in Months', 'Phone Service', 'Avg Monthly Long Distance Charges', 'Multiple Lines', 'Internet Service', 'Avg Monthly GB Download', 'Online Security', 'Online Backup', 'Device Protection Plan', 'Premium Tech Support', 'Streaming TV', 'Streaming Movies', 'Streaming Music', 'Unlimited Data', 'Paperless Billing', 'Age', 'Senior Citizen', 'Married', 'Dependents', 'Cable', 'DSL', 'Fiber Optic', 'None', 'Female', 'Male', 'None', 'Offer A', 'Offer B', 'Offer C', 'Offer D', 'Offer E', 'Bank Withdrawal', 'Credit Card','Mailed Check', 'Month-to-Month','One Year'* dan *'Two Year'*.



* Normalisasi

Pada proses normalisasi database yaitu untuk mengurangi atau meminimalisasi redudansi atau pengulangan data. Selain itu juga untuk memastikan data berada pada posisi yang tepat.

Pada *preprocessing*, kami menggunakan *standard scaler* untuk melakukan normalisasi data agar data yang digunakan tidak memiliki penyimpangan yang besar. *fit\_transform method* adalah dua metode yang digabungkan menjadi satu yaitu metode *fit* dan metode *transform*. *Fit* bertujuan untuk melatih model. Maka dari itu, dataset diuji hanya menggunakan *transform.*

**

* *Splitting Data Testing* dan *Training*

Salah satu metode yang dapat digunakan untuk mengevaluasi performa model *machine learning*. Metode evaluasi model ini membagi dataset menjadi dua bagian yakni bagian yang digunakan untuk *training data* dan *testing data* dengan proporsi tertentu.

|  |  |  |
| --- | --- | --- |
| *Data Training* | 70% | 4930 |
| *Data Testing* | 30% | 2113 |

|  |  |
| --- | --- |
| * *K Nearest Neighbor (KNN)* | * *Decision Tree* |
| * *Support Vector Machine* | * *Logistic Regression* |
| * *Gaussian Naïve Bayes* | * *Random Forest Classifier* |
| * *Gradient Boosting Classifier* |  |

**MODELLING**

Tahap *modelling* merupakan tahap pemodelan statistika untuk menguak hubungan matematis pada satu atau lebih variabel. Beberapa penelitian menyebut tahap ini sebagai *shallow learning*. Penerapan pemodelan statistika pada data yang telah diproses, dapat membantu analis untuk melakukan analisa yang sistematis dan terukur. Tujuan utama pemodelan statistika yakni menguak pola-pola tersirat dalam data, antara lain prediksi, *clustering,* dan visualisasi. Dalam studi kasus kali ini beberapa jenis dari *classification* yang digunakan yaitu:

* *K-Nearest Neighbor (KNN)*

*K-Nearest Neighbor* adalah algoritma *supervised learning* dimana hasil dari *instance* yang baru diklasifikasikan berdasarkan mayoritas dari kategori k-tetangga terdekat. Tujuan dari algoritma ini adalah untuk mengklasifikasikan objek baru berdasarkan atribut dan sampel-sample dari *training* data. Algoritma *k-Nearest Neighbor* menggunakan *Neighborhood Classification* sebagai nilai prediksi dari nilai *instance* yang baru.

* *Decision Tree*

*Decision tree* merupakan model yang terinspirasi oleh bagaimana neuron dalam otak manusia bekerja. Tiap *neuron* pada otak manusia saling berhubungan dan informasi mengalir dari setiap *neuron* tersebut. Jika diilustrasikan, *decision tree* memiliki beberapa cabang. Nah, cabang-cabang tersebut mewakili keputusan (*decision*) yang diambil. Sedangkan, simpul-simpul yang mengaitkan cabang merupakan atribut data. Di sisi lain, hasil keputusan diwakili oleh daun. Hasil tersebut bisa berupa nilai kategoris (untuk klasifikasi data) atau kontinu (regresi data).

* *Support Vector Machine*

*Support Vector Machine* atau biasa disingkat dengan SVM adalah teknik *supervised classification* yang cukup rumit tetapi memiliki tingkat keakuratan yang cukup baik. Algoritma *Support Vector Machine* akan menemukan *hyperplane* atau batas antara dua kelas atau lebih dengan memaksimalkan margin di antara kelas-kelas tersebut. Ada banyak batas yang bisa memisahkan kelas-kelas tersebut, tetapi hanya ada satu batas yang dapat memaksimalkan margin atau jarak antar kelas.

* *Logistic Regression*

Regresi logistik adalah teknik analisis data yang menggunakan matematika untuk menemukan hubungan antara dua faktor data. Kemudian menggunakan hubungan ini untuk memprediksi nilai dari salah satu faktor tersebut berdasarkan faktor yang lain. Prediksi biasanya memiliki jumlah hasil yang terbatas, seperti ya atau tidak.

* *Gaussian Naïve Bayes*

*Naive Bayes* adalah algoritma *machine learning* yang digunakan untuk keperluan klasifikasi atau pengelompokan suatu data. Algoritma ini didasarkan pada teorema probabilitas yang dikenalkan oleh ilmuwan Inggris *Thomas Bayes. Naive Bayes* berfungsi memprediksi probabilitas di masa depan berdasarkan pengalaman sebelumnya, sehingga dapat digunakan untuk pengambilan keputusan. Algoritma *Naive Bayes* mempelajari probabilitas suatu objek berdasarkan ciri-ciri tertentu yang termasuk dalam kelompok atau kelas tertentu. Sederhananya, algoritma ini merupakan pengklasifikasian probabilistik. Inilah alasan mengapa *Naive Bayes* disebut dengan “*naive*”, karena membuat asumsi bahwa kemunculan fitur tertentu tidak tergantung pada kemunculan fitur lainnya.

* *Random Forest Classifier*

*Random Forest* adalah algoritma *machine learning* yang menggabungkan keluaran dari beberapa *decision tree* untuk mencapai satu hasil. Sesuai namanya, *Forest* atau 'hutan' dibentuk dari banyak tree (pohon) yang diperoleh melalui proses *bagging* atau *bootstrap aggregating*. Setiap *tree* pada *Random Forest* akan mengeluarkan prediksi kelas. Prediksi kelas dengan vote terbanyak menjadi kandidat prediksi pada model. Semakin banyak jumlah *tree* maka akan menghasilkan akurasi yang lebih tinggi dan mencegah masalah overfitting.

* *Gradient Boosting Classifier*

*Gradient boost* termasuk ke dalam algoritma *ensemble* yang menggunakan peningkatan akurasi prediktor. Beberapa perbedaan algoritma *gradient boost* dengan *adaboost* adalah *gradient boost* membangun tree dengan 8 sampai 32 daun, sedangkan *adaboost* membangun *stumps* dengan dua daun. Perbedaan kedua adalah *gradient boost* menggunakan *boosting* untuk proses pengoptimalan dengan menggunakan *loss function* untuk meminimalisir kesalahan. Algoritma ini disebut algoritma *gradient boost* karena terinspirasi dari penurunan *gradien.* Perbedaan yang terakhir adalah *tree* digunakan untuk memprediksi sisa sampel (hasil prediksi dikurangi aktual). Intinya, cara kerja algoritma *gradient boost* adalah membangun satu tree untuk menyesuaikan data, lalu tree berikutnya dibangun untuk mengurangi residual *(error).*

Berikut hasil modelling yang telah dilakukan pada studi kasus ini:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Model** | **Precision** | | **Recall** | | **F1-score** | | **Support** | |
| **Yes** | **No** | **Yes** | **No** | **Yes** | **No** | **Yes** | **No** |
| K Nearest Neighbour | 0,73 | 0,90 | 0,75 | 0,90 | 0,74 | 0,90 | 589 | 1524 |
| Decision Tree | 0.82 | 0.82 | 0.45 | 0.96 | 0.58 | 0.89 | 589 | 1524 |
| Support Vector Machine | 0,79 | 0,88 | 0,67 | 0,93 | 0,73 | 0,91 | 589 | 1524 |
| Logistic Regression | 0,63 | 0,95 | 0,88 | 0,80 | 0,73 | 0,86 | 589 | 1524 |
| Gaussian Naïve Bayes | 0,63 | 0,95 | 0,88 | 0,80 | 0,73 | 0,86 | 589 | 1524 |
| Random Forest Classifier | 0,80 | 0,87 | 0,65 | 0,94 | 0,72 | 0,91 | 589 | 1524 |
| **Gradient Boosting Classifier** | 0,82 | 0,89 | 0,69 | 0,94 | 0,75 | 0,91 | 589 | 1524 |

*Primary metrics : f1-score*

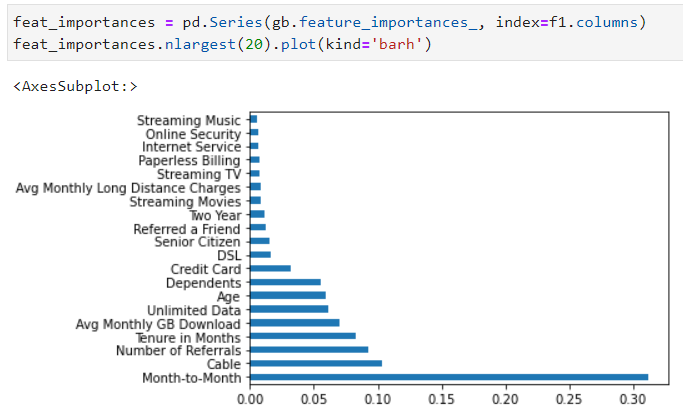
Karena *f1-score balance recall* dengan *precision*

*Secondary metrics : recall*

Karena ingin mendapatkan potensi *churn* sebanyak mungkin

***Feature Important***

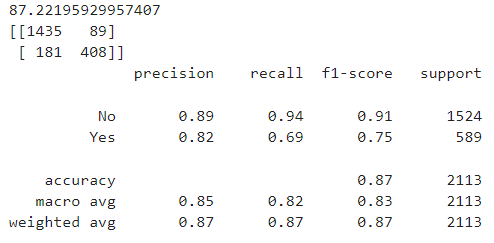
Menunjukan tingkat *customer* yang *churn* dari hasil *modelling* kami yang paling tinggi terdapat *customer* memilih *contract month to month*

****

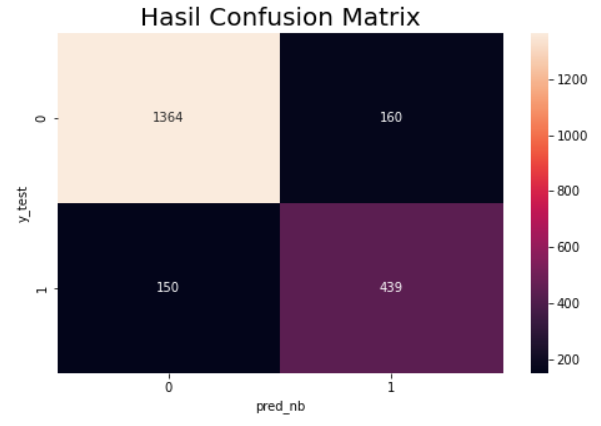
**DEPLOYMENT**

Tahap *deployment* merupakan tahap terakhir dalam proses *data mining.* Tahap ini terdiri dari penggunaan model terbaik yang telah dipilih pada tahap sebelumnya. Hasil model tersebut digunakan dalam pengambilan keputusan, kemudian menerapkannya dalam suatu data baru untuk menghasilkan prediksi yang diharapkan, hasil, dan laporan berdasarkan pengambilan keputusan dari pemodelan tersebut.

* ***Gradient Boosting Classifier***

****

Karena kami memilih *primary metric f1-score* yang memiliki nilai paling tinggi pada *customer* yang *churn*.



1. **Kesimpulan**

Disimpulkan bahwa:

1. Kami menggunakan *modelling Gradient Boosting Classifier* karena memiliki nilai f1-score yang tinggi pada *customer* yang *churn* sehingga dapat mengurangi tingkat *churn* pada *customer*.
2. *Customer* yang *churn* itu paling banyak pada *customer* yang memilih berlangganan bulan ke bulan dari hasil dari *feature important*.
3. Perusahaan ini kalah bersaing dengan kompetitor lain sehingga menjadi salah satu alasan bisa terjadi customer churn dari hasil data visualisasi.
4. **Business Recommendation**
5. Perusahaan dapat memberikan diskon harga terhadap *contract customer* yang berlangganan *month to month*.
6. Perusahaan perlu memperbaiki dan meningkatkan kualitas jaringanyang digunakan untuk jenis tipe internet pelanggan.
7. Perusahaan dapat memberikan *voucher* diskon atau *cashback* apabila *customer* berhasil mengajak teman atau keluarganya untuk berlangganan kepada perusahaan JB Link Telco.
8. Perusahaan perlu meningkatkan kecepatan internet dan menjaga kualitas internet agar tidak melambat dan memberikan batas FUP internet yang tinggi.