Data yang digunakan merupakan data total pendapatan suatu perusahaan telekomunikasi (data fiktif). Pendapatan perusahaan ini berdasarkan 120 variabel prediktor yang dikumpulkan dari berbagai subdistrict atau area jangkauan layanan telekomunikasi yang ditawarkan perusahaan. Terdapat total 156.591 data yang akan dianalisis.

1. **Pada Subdistrict 2**

Sebelum pemodelan, terlebih dahulu dilakukan preprocessing data dengan:

1. Melakukan penghapusan variabel yang memiliki nilai konstan, yaitu dt\_id, revenue\_category, eNodebid\_x, video\_streaming\_start\_delay, bad\_MR\_UMTS, good\_MR\_UMTS, normal\_MR\_UMTS, dan subdistrict.
2. Melakukan feature selection menggunakan uji ANOVA pada model regresi linier untuk variabel kategori dan korelasi untuk variabel numerik. Diperoleh variabel yang tidak memiliki hubungan dengan variabel respon (ditunjukkan dengan uji korelasi untuk p-value > α), yaitu TTI.CELL.30, serverside\_dl\_TCP\_Packet\_loss\_rate subsdistrict\_population, bad\_Voice) sehingga variabel tersebut tidak dimasukkan ke dalam analisis.
3. Melakukan uji multikolinearitas untuk mengetahui hubungan linear yang tinggi antar variabel.
4. Melakukan pengecekan missing value. Hasil menunjukkan tidak ada missing value pada data subdistrict 2.
5. Melakukan split data dengan proporsi 80% data training dan 20% data testing.
6. Melakukan scaling data pada data numerik karena untuk menyamakan rentang data antara variabel numerik.

Selanjutnya, pada data khusus subdistrict 2 dengan variabel respon total pendapatan, dilakukan pemodelan menggunakan tiga macam metode yaitu Generalized Linear Model Elastic Net, Classification and Regression Trees (CART), dan Gradient Boosting. Berikut merupakan hasil pemodelan yang diperoleh.

1. Generalized Linear Model Elastic Net

Pemodelan menggunakan metode GLM Elastic Net menghasilkan performa model sebagai berikut.

|  |  |  |
| --- | --- | --- |
| **RMSE data training** | **RMSE data testing** | **Koefisien Determinasi (R2)** |
| 313179.3 | 323033 | 0.9776592 |

Nilai RMSE pada data testing tidak cukup jauh dibandingkan dengan RMSE data training. Sedangkan untuk nilai koefisien determinasi menunjukkan angka 97,77% variabel yang digunakan dapat menjelaskan model yang dihasilkan dan sisanya 2.23% dijelaskan variabel lain yang tidak dimasukkan pada model.

Dari model yang dihasilkan, dapat dilihat variabel yang berpengaruh terhadap model yaitu sebagai berikut.



Variabel yang paling berpengaruh adalah data\_traffic\_gb, good\_Video\_Mes, dan normal\_Data.

1. Classification and Regression Trees (CART)

Pemodelan menggunakan metode CART menghasilkan performa model sebagai berikut.

|  |  |  |
| --- | --- | --- |
| **RMSE data training** | **RMSE data testing** | **Koefisien Determinasi (R2)** |
| 605348.3 | 625420.4 | 0.9209279 |

Nilai RMSE pada data testing tidak cukup jauh dibandingkan dengan RMSE data training. Sedangkan untuk nilai koefisien determinasi menunjukkan angka 92,09% variabel yang digunakan dapat menjelaskan model yang dihasilkan dan sisanya 7.91% dijelaskan variabel lain yang tidak dimasukkan pada model.

Dari model yang dihasilkan, dapat dilihat variabel yang berpengaruh terhadap model yaitu sebagai berikut.



Dari gambar dapat dilihat bahwa variabel yang paling berpengaruh adalah data\_traffic\_gb dan data\_traffic.

1. Gradient Boosting

Pemodelan menggunakan metode Gradient Boosting menghasilkan performa model sebagai berikut.

|  |  |  |
| --- | --- | --- |
| **RMSE data training** | **RMSE data testing** | **Koefisien Determinasi (R2)** |
| 341539.2 | 358620.6 | 0.9419206 |

Nilai RMSE pada data testing tidak cukup jauh dibandingkan dengan RMSE data training. Sedangkan untuk nilai koefisien determinasi menunjukkan angka 94,19% variabel yang digunakan dapat menjelaskan model yang dihasilkan dan sisanya 5.84% dijelaskan variabel lain yang tidak dimasukkan pada model.

Dari model yang dihasilkan, dapat dilihat variabel yang berpengaruh terhadap model yaitu sebagai berikut.



Dari gambar dapat dilihat bahwa variabel yang paling berpengaruh adalah data\_traffic\_gb.

Dari ketiga model yang dihasilkan dapat dilihat perbandingan dari masing-masing model sebagai berikut.



Nilai performa model RMSE yang paling kecil adalah pada model Generalized Linear Model Elastic Net sehingga model tersebut menjadi model terbaik yang dapat digunakan untuk meregresikan dan memodelkan total pendapatan pada Subdistrict 2. Variabel yang berpengaruh dalam model ini adalah sebagai berikut.



Dari gambar dapat dilihat bahwa variabel yang paling berpengaruh adalah data\_traffic\_gb, good\_Video\_Mes, dan normal\_Data.

1. **Pada Seluruh Subdistrict**

Sebelum pemodelan, terlebih dahulu dilakukan preprocessing data dengan:

1. Melakukan penghapusan variabel yang memiliki nilai konstan, yaitu dt\_id, revenue\_category, eNodebid\_x, video\_streaming\_start\_delay, bad\_MR\_UMTS, good\_MR\_UMTS, normal\_MR\_UMTS, dan subdistrict.
2. Melakukan feature selection menggunakan uji ANOVA pada model regresi linier untuk variabel kategori dan korelasi untuk variabel numerik. Diperoleh variabel yang tidak memiliki hubungan dengan variabel respon (ditunjukkan dengan uji korelasi untuk p-value > α), yaitu AVAIL<98, im\_interactive\_delay\_ms, dan bad\_Voice. sehingga variabel tersebut tidak dimasukkan ke dalam analisis.
3. Melakukan uji multikolinearitas untuk mengetahui hubungan linear yang tinggi antar variabel.
4. Melakukan pengecekan missing value. Hasil menunjukkan tidak ada missing value pada data subdistrict 2.
5. Melakukan split data dengan proporsi 80% data training dan 20% data testing.
6. Melakukan scaling data pada data numerik karena untuk menyamakan rentang data antara variabel numerik.

Pada data seluruh subdistrict dengan variabel respon total pendapatan, dilakukan pemodelan menggunakan Generalized Linear Model Elastic Net. Berikut merupakan hasil pemodelan yang diperoleh.

|  |  |  |
| --- | --- | --- |
| **RMSE data training** | **RMSE data testing** | **Koefisien Determinasi (R2)** |
| 331010.7 | 330426.9 | 0.9753987 |

Nilai RMSE pada data testing tidak cukup jauh dibandingkan dengan RMSE data training. Sedangkan untuk nilai koefisien determinasi menunjukkan angka 97,54% variabel yang digunakan dapat menjelaskan model yang dihasilkan dan sisanya 2.46% dijelaskan variabel lain yang tidak dimasukkan pada model.

Dari model yang dihasilkan, dapat dilihat variabel yang berpengaruh terhadap model yaitu sebagai berikut.



Dari gambar dapat dilihat bahwa variabel yang paling berpengaruh adalah data\_traffic\_gb, promotion3, dan good\_Video\_Mes.

Berdasarkan pemodelan yang dilakukan pada Subdistrict 2 dan seluruh subdistrict, dapat disimpulkan bahwa variabel yang berpengaruh terhadap total pendapatan perusahaan telekomunikasi adalah variabel data\_traffic\_gb dan good\_Video\_Mes, artinya bahwa untuk meningkatkan pendapatan perusahaan telekomunikasi maka perlu meningkatkan data traffic BTS dan indeks kualitas pesan video.