

Penerapan Algoritma K-Nearest Neighbor dalam Pengklasifikasian Ge Burst pada Penambangan Batu Bara

Anasthashya Rachman
121450013

Kirana Ratu Malhanny
121450082

Revaldo Dafa Fahmindo
121450085

Dara Cantika Dewi
121450127

Mayada
121450145

Abstrak—Penerapan algoritma K-NN dan seismic-bumps dataset membahas mengenai dampak pertambangan batubara terhadap ekosistem dan pentingnya model geologi dalam eksploitasi sumber daya batubara yang berhubungan dengan ge-burst. Ge-burst dapat terjadi dengan menimbulkan ledakan geologi yang dapat menyebabkan konsekuensi berat pada infrastruktur pertambangan. Oleh karena itu, mencegah dan mengendalikan ledakan geologi merupakan tantangan besar bagi industri pertambangan. Laporan ini mengusulkan penggunaan analisis data seismik untuk mengidentifikasi area aktivitas seismik tinggi dan memprediksi pelepasan energi potensial, sehingga mengurangi risiko ledakan tanah. Studi ini menggunakan algoritma K-Nearest Neighbor (K-NN) untuk mengklasifikasikan *ge burst* berdasarkan dataset seismic-bumps dari UCI Machine Learning Repository. Tujuan dari penelitian ini adalah untuk menerapkan algoritma K-NN dalam mengidentifikasi faktor yang berkontribusi pada seismic-bumps, meningkatkan akurasi klasifikasi, dan menginterpretasikan proses pengambilan keputusan metode K-NN. Hasil dari laporan ini menyoroti korelasi antara seismic-bumps, ledakan geologi, dan algoritma K-NN. Pada laporan ini menyimpulkan bahwa metode yang diusulkan dapat membantu mengidentifikasi dan mengurangi risiko ledakan tanah dalam pertambangan batubara, yang mengarah pada peningkatan keselamatan dan pelestarian infrastruktur.

Kata kunci—K-Nearest Neighbors (K-NN), *Ge burst*, *Seismic bumps*, mitigasi risiko, pertambangan, batubara

I. PENDAHULUAN

Kegiatan pertambangan batu bara merupakan kegiatan eksploitasi sumber daya alam yang tidak dapat diperbaharui, dimana di dalam kegiatan penambangan dapat berdampak pada rusaknya ekosistem. Mekanisasi peralatan dan teknologi pertambangan telah menyebabkan skala pertambangan semakin besar dan untuk dapat memanfaatkan potensi sumber daya batubara yang ada di bawah permukaan, perlu diketahui model geologi batu baranya, yakni dengan melakukan kegiatan eksploitasi pengeboran. Struktur geologi, metode penambangan, dan parameter desain merupakan faktor kritis dalam mengontrol gerakan tanah dan mencegah terjadinya *ge burst* dalam penambangan batubara bawah tanah. Jika operasi penambangan tidak dilakukan dengan aman, dapat menyebabkan kegagalan lapisan tanah dan kemudian terjadinya *ground burst*, yang dapat menyebabkan kerusakan pada peralatan, infrastruktur, dan luka pada personil.

Terjadinya *ge burst* dapat berdampak serius pada keselamatan para pekerja tambang, karena dapat menyebabkan cedera, kematian, dan kerusakan pada

infrastruktur tambang. Oleh karena itu, pencegahan dan pengendalian *ge burst* merupakan tantangan besar bagi industri pertambangan. Data seismik dapat digunakan untuk menganalisis risiko *ground burst* dengan mengidentifikasi daerah dengan aktivitas seismik yang tinggi dan memperkirakan potensi pelepasan energi.

Dalam penelitian ini, kami menggunakan algoritma *K-Nearest Neighbor* (K-NN) untuk mengklasifikasikan *Ge Burst* berdasarkan dataset seismic-bumps dari UCI Machine Learning Repository [1]. Algoritma ini dapat menangani klasifikasi multi kelas dan tidak membuat asumsi tentang distribusi data. Hal tersebut yang membuat sangat relevannya dataset dan algoritma yang digunakan pada laporan ini.

Pelaksanaan tugas besar statistika sains data ini memiliki tujuan dengan penerapan algoritma K-NN, yaitu: dapat mengetahui faktor terjadinya *seismic bumps*, metode K-NN dalam memprediksi terjadinya *seismic-bumps*, penggunaan *data preprocessing* dalam meningkatkan akurasi hasil klasifikasi, interpretasi metode K-NN dalam pengambilan keputusan, dan bagaimana hasil klasifikasi dengan peningkatan keamanan di tambang batu bara.

II. METODE

Metode pada penelitian ini menggunakan metode kuantitatif berbasis analisis data yang melibatkan pengolahan data, pemilihan variabel yang relevan, serta penggunaan algoritma *KNN* (*K-Nearest Neighbors*) untuk melakukan klasifikasi pada data seismic bumps. Metode analisis data Seismic-bumps ini merupakan metode yang digunakan untuk mengolah data aktivitas seismik guna mengidentifikasi kejadian *seismic bumps* (*ge burst*) pada tambang batubara. Dengan menggunakan metode analisis data *Seismic-bumps*, informasi yang diperoleh akan lebih detail dan akurat tentang kejadian *seismic bumps* khususnya *ge burst* pada tambang batubara, serta menganalisis hubungannya dengan faktor-faktor lain dalam dunia pertambangan.

III. HASIL DAN PEMBAHASAN

Dalam penerapannya, laporan ini memiliki hal yang sangat penting dalam keterkaitan antara *seismic-bumps*, *ge burst*, dan algoritma KNN yang akan ditinjau lebih lanjut.

A. Seismic-bumps

Seismic-bumps adalah suatu peristiwa yang terjadi dalam dunia pertambangan yang berupa runtuhnya batuan-batuan

di sekitar area penambangan batu bara. Hal ini disebabkan oleh peningkatan tekanan pada tiang-tiang batu bara. *Seismic-bumps* menjadi suatu hal buruk yang dapat merusak infrastruktur dan dapat membahayakan karyawan pertambangan. *Seismic-bumps* sangat dikaitkan dengan *ge burst* karena *ge burst* merupakan faktor terjadinya *seismic-bumps* [2].

B. Ge-burst

Ge burst seringkali dikaitkan dengan dunia pertambangan yang sangat berkaitan erat dan termasuk dalam seismik penambangan batubara. *Ge burst* ini merupakan salah satu fenomena terjadinya pelepasan gas metana secara tiba-tiba dan dalam jumlah besar dari lapisan batubara di bawah tanah. Pelepasan gas ini dapat menyebabkan tekanan dalam tambang meningkat secara dramatis dan dapat menyebabkan retakan dan deformasi pada lapisan batubara dan batuan di sekitarnya.

Ge burst atau *ground burst* dapat terjadi akibat dari *outburst* atau pelepasan energi yang tiba-tiba dari massa batuan di sekitar wajah ekskavasi. Hal ini menyebabkan eksesi keras berupa pelepasan batuan, arang, gas, atau kombinasi dari bahan-bahan tersebut. Fenomena ini sangat kompleks dan dipengaruhi oleh banyak faktor, seperti geologi, desain tambang, urutan penambangan, metode penambangan, rezim tekanan, dan tekanan gas. Mekanisme pasti dari *ground burst* diyakini terkait dengan akumulasi dan pelepasan energi elastis di massa batuan, yang disebabkan oleh perubahan dalam medan tegangan atau oleh pembukaan tiba-tiba retakan di massa batuan [3].

Ge burst sangat berisiko, sehingga butuh upaya dalam menganalisis risiko seismik dari *ge burst* atau *ground burst* pada tambang batu bara dengan berbagai metode. Metode-metode ini meliputi :

1. Pemantauan seismik

Pemantauan seismik dapat digunakan untuk mendeteksi dan menentukan lokasi dari peristiwa mikroseismik, yang sering kali merupakan pendahulu dari *ge burst* atau *ground burst*.

2. Tomografi seismik

Tomografi seismik dapat digunakan untuk membuat gambar struktur di bawah permukaan dan mendeteksi sumber potensial dari konsentrasi tegangan.

3. Inversi seismik.

Inversi seismik dapat digunakan untuk memperkirakan sifat mekanik dari massa batuan, seperti elastisitas dan kekuatannya, yang merupakan faktor penting dalam menentukan risiko *ground burst*.

Selain itu, terdapat juga beberapa faktor yang dapat mempengaruhi risiko *ge burst* atau *ground burst* meliputi :

1. Kondisi geologi dan struktural dari seam batu bara dan massa batuan sekitarnya
2. Metode dan urutan penambangan
3. Kedalaman penambangan
4. Keadaan tegangan massa batuan
5. Kandungan dan tekanan gas dalam seam batu bara.

C. Algoritma K-NN (K-Nearest Neighbors)

Algoritma *K-Nearest Neighbor* (*K-NN*) adalah algoritma machine learning yang sederhana dan populer untuk klasifikasi data. Algoritma ini didasarkan pada gagasan mencari k closest training di ruang fitur dan menggunakan label kelas untuk menentukan label kelas contoh baru. Dengan kata lain, algoritma ini memperkirakan distribusi probabilitas bersyarat variabel kelas yang diberikan nilai fiturnya. Algoritma *K-NN* memiliki syarat dengan nilai k yang tidak boleh lebih besar dari jumlah data latih, diharuskan untuk bernilai ganjil dan lebih dari satu [4].

Pada penerapannya, algoritma *K-NN* memiliki beberapa langkah dalam perhitungan metode *K-Nearest Neighbor*, yaitu:

1. Menentukan parameter K
2. Lakukan pemisahan antara data latih dan data test dengan ketentuan data latih lebih banyak daripada data tes
3. Melakukan perhitungan jarak antara data training dan data testing

Perhitungan jarak dengan metode *K-NN* dapat dilakukan dengan perhitungan euclidean. Rumus euclidean sebagai berikut :

$$euc = \sqrt{\sum_{i=1}^n (p_i - q_i)^2} \quad (1)$$

Pada rumus tersebut memiliki keterangan:

p_i = sample data pada data training

q_i = sample data pada data testing

i = variabel data

n = dimensi data

4. Lakukan pengurutan jarak yang telah didapatkan pada rumus euclidean untuk mendapatkan jarak terdekat sampai urutan K
5. Kelompokkan data berdasarkan kelas yang bersesuaian
6. Lakukan pencarian jumlah kelas dari tetangga yang terdekat dan tetapkan kelas tersebut sebagai kelas data yang akan dievaluasi

Selain itu, dalam penerapannya, algoritma *K-NN* juga mampu memprediksi dan melihat akurasi pemodelan yang dilakukan dengan confusion matrix. Confusion matrix merupakan pembuatan ringkasan data berdasarkan tabel yang terdiri atas banyaknya baris data uji yang diprediksi dengan benar dan tidak benar oleh model pengklasifikasian yang dapat menentukan kinerja dari pemodelan data. Rumus yang digunakan dalam confusion matrix adalah:

$$AKURASI = ((TP+TN))/((TP+TN+FP+FN)) \quad (2)$$

$$PRESISI = TP/(TP+FP) \quad (3)$$

$$RECALL = TP/(TP+FN) \quad (4)$$

$$F\text{-Measure} = 2 (Presisi \times Recall)/(Presisi+Recall) \quad (5)$$

Keterangan variabel:

TP : True Positive

TN : True Negative

FP : False Positive

FN : False Negative.

Algoritma K-NN memiliki kelebihan dalam melakukan penyederhanaan dan interpretabilitas. Tetapi, performa algoritma K-NN dapat dipengaruhi oleh pemilihan nilai k, metrik jarak, dan skalabilitas fitur. Selain itu, algoritma dapat menjadi mahal secara komputasional untuk dataset yang besar [5].

D. Penerapan antara seismic-bumps dataset dan algoritma K-NN

Berdasarkan penjelasan yang telah dijelaskan lebih rinci tentang algoritma K-NN didapatkan bahwa, algoritma K-NN digunakan untuk menganalisis data yang berbentuk klasifikasi. Hal ini sangat berkaitan dengan seismic-bumps dataset yang merupakan bentuk kumpulan data berbasis data klasifikasi. Dalam pengklasifikasian pada data ini, terdapat variabel class yang berisi data klasifikasi 0 dan 1. Data 0 menandakan keadaan aman dan 1 menandakan keadaan bahaya.

Pada dataset seismic-bumps terdiri dari 2584 sampel dengan 19 atribut. Atribut yang terdapat pada dataset ini antara lain:

1. seismic: hasil penilaian bahaya seismik pada lokasi tambang yang diperoleh melalui metode seismik. Nilai yang mungkin adalah a (tidak ada bahaya), b (bahaya rendah), c (bahaya tinggi), atau d (keadaan berbahaya).
2. seismoacoustic: hasil penilaian bahaya seismik pada lokasi tambang yang diperoleh melalui metode seismoakustik. Nilai yang mungkin adalah a (tidak ada bahaya), b (bahaya rendah), c (bahaya tinggi), atau d (keadaan berbahaya).
3. shift: informasi tentang jenis shift, apakah shift untuk pengambilan batu bara (W) atau shift persiapan (N).
4. genenergy: energi seismik yang direkam pada shift sebelumnya oleh geophone paling aktif (GMax) dari geophone yang memantau longwall.
5. gpuls: jumlah pulse seismik yang direkam pada shift sebelumnya oleh GMax.
6. gdenergy: deviasi energi yang direkam pada shift sebelumnya oleh GMax dari rata-rata energi yang direkam selama delapan shift sebelumnya.
7. gdpuls: deviasi jumlah pulse yang direkam pada shift sebelumnya oleh GMax dari jumlah pulse yang direkam selama delapan shift sebelumnya.
8. ghazard: hasil penilaian bahaya seismik pada lokasi tambang yang diperoleh melalui metode seismoakustik berdasarkan data yang berasal dari GMax saja. Nilai yang mungkin adalah a (tidak ada bahaya), b (bahaya rendah), c (bahaya tinggi), atau d (keadaan berbahaya).
9. nbumps: jumlah bump seismik yang direkam pada shift sebelumnya.
10. nbumps2: jumlah bump seismik (dalam rentang energi $[10^2, 10^3]$) yang terdaftar pada shift sebelumnya.
11. nbumps3: jumlah bump seismik (dalam rentang energi $[10^3, 10^4]$) yang terdaftar pada shift sebelumnya.

12. nbumps4: jumlah bump seismik (dalam rentang energi $[10^4, 10^5]$) yang terdaftar pada shift sebelumnya.
13. nbumps5: jumlah bump seismik (dalam rentang energi $[10^5, 10^6]$) yang terdaftar pada shift sebelumnya.
14. nbumps6: jumlah bump seismik (dalam rentang energi $[10^6, 10^7]$) yang terdaftar pada shift sebelumnya.
15. nbumps7: jumlah bump seismik (dalam rentang energi $[10^7, 10^8]$) yang terdaftar pada shift sebelumnya.
16. nbumps8: Jumlah bump seismik (dalam rentang energi $[10^8, 10^{10}]$) yang terdaftar pada shift sebelumnya.
17. energy: total energi dari bump seismik yang terdaftar pada shift sebelumnya.
18. maxenergy: Energi maksimum dari bump seismik yang terdaftar pada shift sebelumnya.
19. class: atribut keputusan - '1' berarti terjadi guncangan seismik dengan energi tinggi pada shift berikutnya ('keadaan berbahaya'), sedangkan '0' berarti tidak terjadi guncangan seismik dengan energi tinggi pada shift berikutnya ('keadaan aman').

Dalam penerapannya, algoritma K-NN dapat diterapkan melalui software R-Studio dengan eksekusi kode sebagai berikut:

1. Import library untuk dapat digunakan dalam analisis seismic-bumps dataset

library(readr) : untuk membaca file data dalam format CSV atau TSV ke dalam R

library(dplyr) : untuk memanipulasi dan menyaring data

library(caret) : untuk memfasilitasi pemodelan data dan pengujian

library(class) : menyediakan fungsi klasifikasi untuk penggunaan dalam model machine learning

library(ggplot2) : untuk membuat visualisasi data

library(scales) : untuk mengelola dan memformat data numerik

2. Melakukan import data dan menampilkan 6 data teratas

```
seismic=read_csv("C:/Users/User/Downloads/seismic-bumps.csv")
```

```
head(seismic)
```

```
seismic = read_csv("C:/Users/User/Downloads/seismic-bumps.csv")

## Rows: 2584 Columns: 19
## — Column specification —
## Delimiter: ","
## chr (4): seismic, seismoacoustic, shift, ghazard
## dbl (15): genergy, gpuls, gdenery, gdpuls, nbumps, nbumps2, nbumps3, nt
##
## 1 Use `spec()` to retrieve the full column specification for this data.
## 2 Specify the column types or set `show_col_types = FALSE` to quiet thi

head(seismic)

## # A tibble: 6 x 19
##   seismic seismoacou.1 shift genergy gpuls gdenery2 gdpuls ghazard nbumps
##   <chr>   <chr>      <chr>   <dbl> <dbl>   <dbl> <dbl> <chr>   <dbl>
## 1 a      a          N      15180  48     -72   -72 a      6
## 2 a      a          N      14720  33     -70   -79 a      1
## 3 a      a          N      8050   30     -81   -78 a      6
## 4 a      a          N      28820  171    -23    40 a      1
## 5 a      a          N      12640  57     -63   -52 a      6
## 6 a      a          W      63760  195    -73   -65 a      6
## # ... with 9 more variables: nbumps3 <dbl>, nbumps4 <dbl>, nbumps5 <dbl>,
## #   nbumps6 <dbl>, nbumps7 <dbl>, nbumps89 <dbl>, energy <dbl>,
## #   maxenergy <dbl>, class <dbl>, and abbreviated variable names
## #   *seismoacoustic, *gdenery
```

Gambar 1. Hasil output import data dan data teratas dari dataset

3. Melakukan eksplorasi data

`str(seismic)` : untuk melihat struktur dari data dari seismic-bumps dataset

`summary(seismic)` : untuk menghasilkan ringkasan statistik dari data untuk memberikan gambaran umum tentang nilai-nilai dalam data.

`sum(is.na(seismic))` : untuk menghitung jumlah missing value yang terdapat pada dataset seismic dan terdapat 0 missing value

4. Preprocessing Data

`seismic <- seismic[, -c(1, 2, 3, 8)]` : menghapus kolom yang tidak diperlukan

`seismic$class <- as.factor(seismic$class)` : mengubah kolom "class" dalam dataset "seismic" menjadi tipe data faktor. Faktor digunakan untuk memperlakukan kolom "class" sebagai variabel kategori.

```
normalize <- function(x) {
  return ((x - min(x)) / (max(x) - min(x))) }
```

```
seismic_norm <- as.data.frame(lapply(seismic[, -ncol(
  seismic)], normalize))
```

`seismic_norm$class <- seismic$class` : untuk melakukan normalisasi pada data numerik. Kolom "class" tidak dinormalisasi karena merupakan kolom yang berisi label atau kategori, normalisasi tidak diperlukan karena nilainya mewakili kategori atau kelas yang diskrit.

```
set.seed(123)
```

```
train.index <- createDataPartition(seismic$class, p
= 0.7, list = FALSE)
```

```
train.data <- seismic[train.index, ]
```

`test.data <- seismic[-train.index,]` : untuk membagi data menjadi data pelatihan dan data uji. Data pelatihan sebanyak 70% dan data uji sebanyak 30%.

```
train.features <- train.data[, -15]
```

```
train.target <- train.data$class
```

```
test.features <- test.data[, -15]
```

`test.target <- test.data$class` : Memisahkan fitur dan target dari data pelatihan dan data uji

5. Pemodelan KNN

`accuracy_values <- data.frame(k = integer(), accuracy = numeric())` : Membuat dataframe kosong dengan dua kolom, "k" untuk nilai k dalam KNN dan "accuracy" untuk nilai akurasi yang akan ditampung.

```
for (i in 1:20) {
```

```
  model_knn <- knn(train = train.features,
```

```
    test = test.features,
```

```
    cl = train.target,
```

```
    k = i)
```

```
  prediction_knn <- as.factor(model_knn)
```

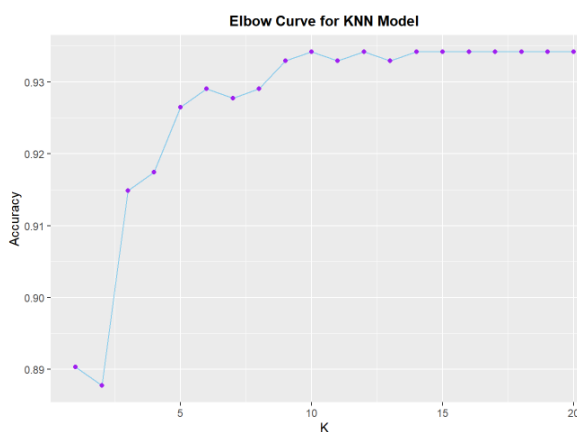
```
  accuracy_values[i, 1] <- i
```

```
  accuracy_values[i, 2] <- sum(prediction_knn ==
test.target)/nrow(test.data)
```

} : Dilakukan iterasi dari $k = 1$ hingga $k = 20$ menggunakan nilai k pada data pelatihan dan data uji.

```
ggplot(accuracy_values, aes(x = k, y = accuracy)) +
+
geom_line(color = "skyblue") +
geom_point(color = "purple") +
xlab("K") + ylab("Accuracy") +
ggtitle("Elbow Curve for KNN Model") +
```

theme(plot.title=element_text(face='bold',color='black',hjust=0.5,size=12)) : didapatkan grafik tertinggi pada angka 10 dan nilai tersebut yang menandakan k terbaik



Gambar 2. Penerapan elbow curve dalam pemodelan K-NN

Membuat model berdasarkan akurasi:

```
k_best <- which.max(accuracy_values$accuracy)

cat("K terbaik:", k_best, "\n")

cat("Akurasi tertinggi:", max(
accuracy_values$accuracy), "\n") : untuk mencari
indeks baris dengan akurasi tertinggi. Didapatkan k
terbaik = 10 dan akurasi tertinggi = 0.9341935

model_knn_final <- knn(train = train.features,
test = test.features,
cl = train.target,
```

$k = k_best$)

Model ini akan digunakan untuk memprediksi kelas pada data uji.

6. Evaluasi Model

prediction_knn <- as.factor(model_knn_final) :
Memprediksi data uji dengan model KNN yang telah dilatih

confusion_matrix<-confusionMatrix(prediction_knn, test.target)

confusion_matrix : Menampilkan hasil prediksi pada data uji

```
## Confusion Matrix and Statistics
##
##           Reference
## Prediction  0    1
##           0 724  50
##           1   0   1
##
##           Accuracy : 0.9355
##           95% CI : (0.9158, 0.9517)
##           No Information Rate : 0.9342
##           P-Value [Acc > NIR] : 0.4795
##
##           Kappa : 0.036
##
##           Mcnemar's Test P-Value : 4.219e-12
##
##           Sensitivity : 1.00000
##           Specificity : 0.01961
##           Pos Pred Value : 0.93540
##           Neg Pred Value : 1.00000
##           Prevalence : 0.93419
##           Detection Rate : 0.93419
##           Detection Prevalence : 0.99871
##           Balanced Accuracy : 0.50980
##
##           'Positive' Class : 0
##
```

Gambar 3. Hasil confusion matrix dan akurasi

```
accuracy.knn <- round(mean(test.target ==
prediction_knn), digits = 2)*100
```

```
print(paste('Akurasi model: ',accuracy.knn,"%"))
```

Akurasi model = 94% (sudah dibulatkan) dan tes error = 6% (sudah dibulatkan). Hal ini menandakan

bahwa model yang digunakan sangat cocok dan sangat baik.

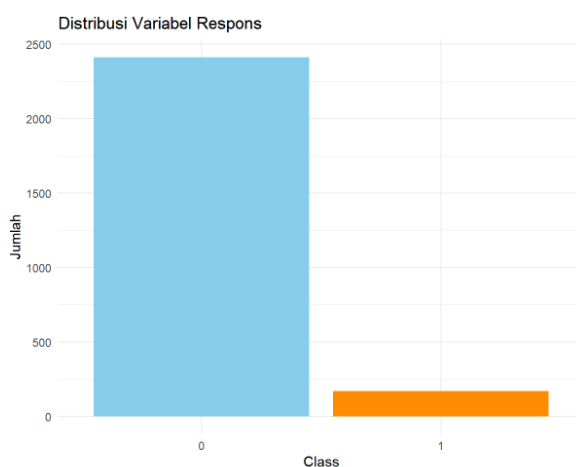
7. Visualisasi

```
ggplot(data = seismic, aes(x = class, fill = class)) +  
  geom_bar() +
```

```
  scale_fill_manual(values = c("skyblue",  
  "darkorange")) +
```

```
  labs(title = "Distribusi Variabel Respon", x =  
  "Class", y = "Jumlah") +
```

```
  theme_minimal()
```



Gambar 4. Visualisasi variabel respon (class)

Pada distribusi variabel respon didapatkan bahwa, tidak terjadi guncangan seismik dengan energi tinggi pada shift berikutnya ('keadaan aman').

Selain dari visualisasi tersebut, terdapat juga visualisasi antara variabel respon (class) dengan variabel prediktor yang paling mempengaruhi, seperti variabel genergy, gpuls, nbumps, nbumps2, nbumps3, nbumps4 yang diketahui dari analisis regresi dan didapatkan hasilnya sebagai berikut:

```
call:
lm(formula = class ~ ., data = seismic)

Residuals:
    Min       1Q   Median       3Q      Max
-0.57635 -0.07796 -0.03754 -0.00939  1.02708

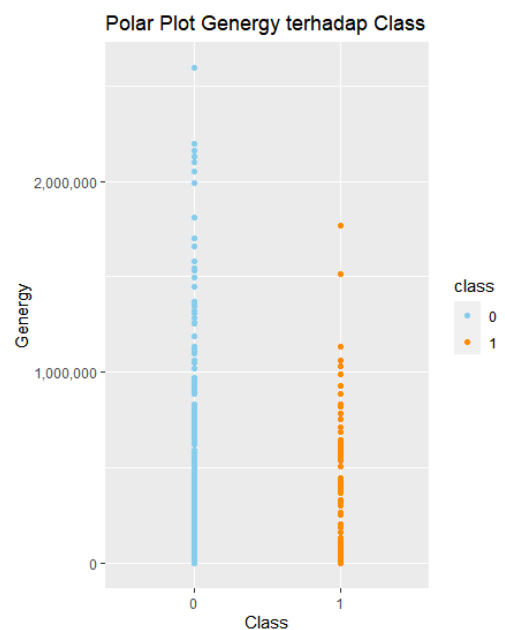
Coefficients: (3 not defined because of singularities)
(Intercept)   -1.580e-02  9.264e-03 -1.705
seismicb       1.725e-02  1.087e-02  1.587
seismoacousticb 1.737e-03  1.043e-02  0.167
seismoacousticc 2.165e-02  4.617e-02  0.469
shifw         -2.745e-04  1.157e-02  0.024
genergy       -8.806e-08  3.463e-08 -2.543
gpuls         1.025e-04  1.673e-05  6.123
gdenergy      -6.589e-05  1.007e-04 -0.655
gpuls        -1.987e-04  1.372e-04 -1.448
ghazardb     -2.983e-03  1.930e-02 -0.155
ghazardc     -7.924e-02  5.803e-02 -1.366
nbumps       4.679e-01  1.680e-01  2.785
nbumps2      -4.287e-01  1.682e-01 -2.548
nbumps3      -4.266e-01  1.681e-01 -2.538
nbumps4      -4.640e-01  1.708e-01 -2.716
nbumps5      -3.036e-01  2.334e-01 -1.301
nbumps6      NA
nbumps7      NA
nbumps89     NA
energy       3.388e-07  2.397e-06  0.141
maxenergy    -1.115e-06  2.335e-06 -0.478

Pr(>|t|)
(Intercept)   0.08826 .
seismicb      0.11273
seismoacousticb 0.86769
seismoacousticc 0.63920
shifw         0.98108
genergy       0.01104 *
gpuls        1.06e-09 ***
gdenergy      0.51275
gpuls         0.14766
ghazardb      0.87717
ghazardc      0.17220
nbumps        0.00538 **
nbumps2       0.01089 *
nbumps3       0.01120 *
nbumps4       0.00664 **
nbumps5       0.19345
nbumps6      NA
nbumps7      NA
nbumps89     NA
energy        0.88760
maxenergy     0.63290
---
Signif. codes:
  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.2371 on 2566 degrees of freedom
Multiple R-squared:  0.09178,    Adjusted R-squared:  0.08576
F-statistic: 15.25 on 17 and 2566 DF,  p-value: < 2.2e-16
```

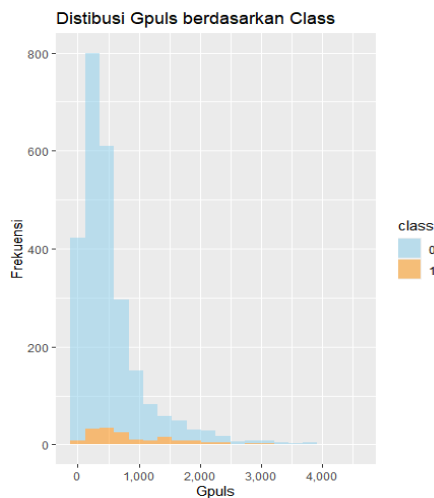
Gambar 5. Hasil analisis regresi untuk menentukan variabel faktor yang paling berpengaruh

Bentuk visualisasinya variabel yang paling berpengaruh adalah :



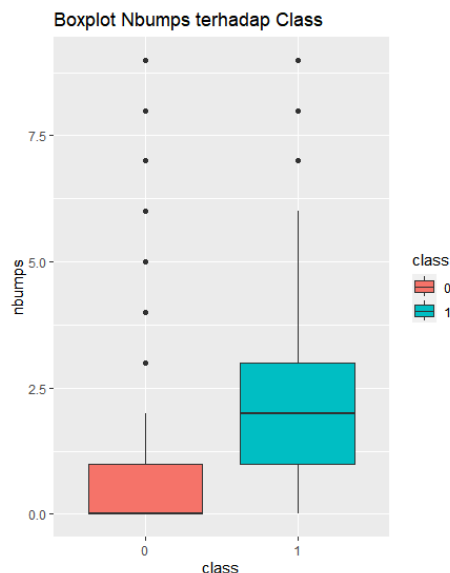
Gambar 6. Visualisasi variabel prediktor (genergy) terhadap variabel respon (class)

Pada visualisasi antara variabel respon (class) dan variabel prediktor (genergy) didapatkan bahwa, energi seismik yang direkam pada shift sebelumnya banyak menandakan keadaan aman.



Gambar 7. Visualisasi variabel prediktor (gpuls) terhadap variabel respon (class)

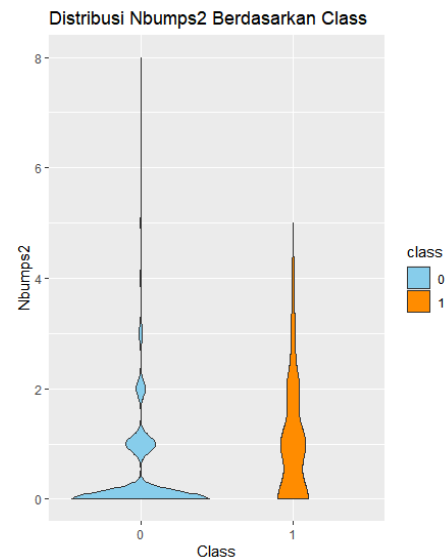
Pada visualisasi antara variabel respon (class) dan variabel prediktor (gpuls) didapatkan bahwa, jumlah pulse seismik yang direkam pada shift sebelumnya oleh GMax lebih banyak menandakan keadaan aman.



Gambar 8. Visualisasi variabel prediktor (Nbumps) terhadap variabel respon (class)

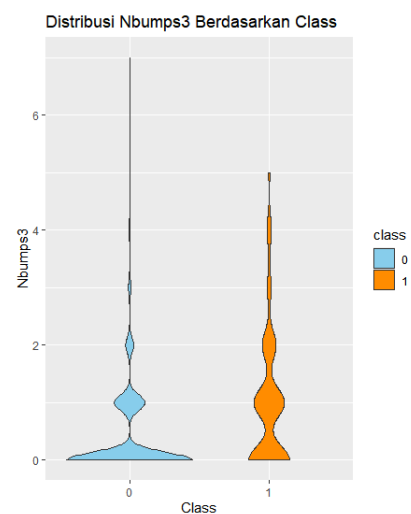
Pada visualisasi antara variabel respon (class) dan variabel prediktor (Nbumps) didapatkan bahwa,

jumlah bump seismik yang direkam pada shift sebelumnya lebih condong menunjukkan gejala tidak aman yang dapat menyebabkan keruntuhan atau longsor yang mengancam keselamatan para pekerja tambang.



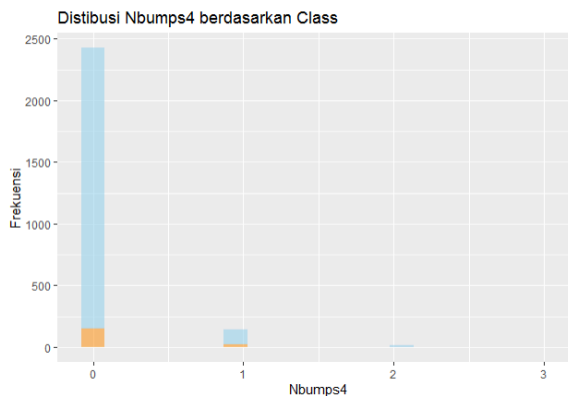
Gambar 9. Visualisasi variabel prediktor (Nbumps2) terhadap variabel respon (class)

Pada visualisasi antara variabel respon (class) dan variabel prediktor (Nbumps2) didapatkan bahwa, jumlah bump seismik dalam rentang energi $[10^2, 10^3]$ yang terdaftar pada shift sebelumnya lebih banyak berdistribusi pada class yang menunjukkan keadaan aman.



Gambar 10. Visualisasi variabel prediktor (Nbumps3) terhadap variabel respon (class)

Pada Nbums3 juga memiliki visualisasi yang hampir mirip dengan Nbums2. Hal ini menandakan bahwa, jumlah bump seismik dalam rentang energi $[10^2, 10^4]$ yang terdaftar pada shift sebelumnya lebih banyak berdistribusi pada class yang menunjukkan keadaan aman.



Gambar 11. Visualisasi variabel prediktor (Nbums4) terhadap variabel respon (class)

Pada Nbums4 juga memiliki visualisasi yang hampir mirip dengan Nbums2. Hal ini menandakan bahwa, jumlah bump seismik dalam rentang energi $[10^2, 10^5]$ yang terdaftar pada shift sebelumnya lebih banyak berdistribusi pada class yang menunjukkan keadaan aman.

IV. KESIMPULAN

Hasil evaluasi menunjukkan bahwa model K-NN memiliki akurasi sebesar 93.55% dengan $k = 10$. Akurasi ini menunjukkan kemampuan model dalam memprediksi *ge burst* dengan tepat dalam konteks penambangan batubara. Confusion matrix menunjukkan bahwa model berhasil mengklasifikasikan sebagian besar *ge burst* dengan benar (724 true positives) dengan tingkat sensitivitas yang tinggi (1000). Dalam hal ini, dibuktikan bahwa model mampu mengenali sebagian besar kejadian *ge burst* yang sebenarnya. Hal ini sangat penting karena *ge burst* dapat menjadi indikasi adanya potensi bahaya yang dapat membahayakan pekerja dan infrastruktur tambang. Namun, model memiliki tingkat kesalahan yang rendah (hanya 50 false positives) atau spesifisitas yang rendah (0,01961), yang berarti model cenderung memberikan banyak false positives (misclassifying *non-ge burst* sebagai *ge burst*).

Tingkat spesifisitas yang rendah menunjukkan bahwa model memiliki kecenderungan untuk memberikan prediksi false positive, yaitu mengklasifikasikan kejadian *non-ge burst* sebagai *ge burst*. Hal ini dapat mengakibatkan peningkatan alarm palsu dan mengganggu operasional tambang. Untuk mencapai tingkat keamanan yang optimal,

perlu dilakukan penyesuaian pada metode K-NN atau menerapkan strategi tambahan untuk meningkatkan spesifisitas tanpa mengorbankan sensitivitas. Pengembangan model yang lebih canggih atau penambahan atribut yang relevan dalam proses pemodelan dapat membantu meningkatkan kinerja dan keandalan prediksi.

Faktor terjadinya seismic bumps berdasarkan hasil analisis yang dilakukan yaitu, atribut-atribut seperti energi gelombang seismik (*genergy*), lebar gelombang puncak (*gpuls*), dan energi gelombang seismik yang didistribusikan (*gdenrgy*) memiliki pengaruh terhadap terjadinya *ge burst*. Jumlah *ge burst* dari tipe 2 hingga 9 (nbumps2 hingga nbumps89) juga memberikan kontribusi dalam memprediksi *ge burst*. Dengan memperhatikan faktor-faktor ini, metode K-NN dapat digunakan untuk memprediksi *ge burst* berdasarkan pola dan karakteristik dari atribut-atribut yang terkait.

Hasil klasifikasi yang diperoleh dengan peningkatan keamanan di tambang batu bara dapat memberikan manfaat yang signifikan. Dengan mengidentifikasi *ge burst* secara akurat, langkah-langkah pencegahan dan mitigasi risiko yang tepat dapat diambil untuk mengurangi potensi bahaya dan kerugian yang disebabkan oleh *ge burst*. Ini membantu meningkatkan keamanan di tambang batu bara dengan memberikan informasi penting kepada operator tambang untuk mengambil tindakan yang diperlukan dengan cepat dan efektif. Penerapan metode K-NN untuk dataset seismic bumps dalam pengklasifikasian *ge burst* terhadap penambangan batubara memberikan hasil yang menjanjikan.

Penggunaan data preprocessing dapat meningkatkan akurasi hasil klasifikasi, sementara interpretasi metode K-NN mengandalkan mayoritas *voting* dari kategori tetangga terdekat untuk menentukan kelas dari sampel yang akan diprediksi. Interpretasi metode ini adalah bahwa *ge burst* akan diklasifikasikan berdasarkan mayoritas dari *ge burst* pada tetangga terdekatnya dalam ruang atribut yang diukur. Dengan menggunakan metode ini, kita dapat mendapatkan prediksi klasifikasi berdasarkan pola yang teramati dari data pelatihan yang ada. Peningkatan keamanan yang dihasilkan dari hasil klasifikasi yang akurat, risiko kecelakaan dan kerugian di tambang batu bara dapat dikurangi, sehingga meningkatkan keamanan secara keseluruhan.

DAFTAR PUSTAKA

- [1] UCI Machine Learning Repository, "Seismic Bumps Data Set." [Online]. Available: <https://archive.ics.uci.edu/ml/datasets/seismic-bumps>
- [2] Harinda, R. R. (2017). "Seismic Bumps pada Tambang Batubara di Indonesia." *Jurnal Sains dan Teknologi Indonesia*, vol. 19, no. 2, hlm. 89-94.
- [3] Scott, B., Ranjith, P. G., Choi, X. S. K., & Khandelwal, M. (2010). "Geological and geotechnical aspects of underground coal mining methods within Australia." *Environmental Earth Sciences*, vol. 60(5), pp. 1007-1019, May 2010.
- [4] Rivki, M., & Bachtar, A. M. (2017). "Implementasi Algoritma K-Nearest Neighbor dalam Pengklasifikasian Follower Twitter yang Menggunakan Bahasa Indonesia." *Journal of Information Systems*, vol. 11, no. 1, pp. 31-37.
- [5] Cahyantia, D., Rahmayania, A., & Husniara, S. A. (2020). "Analisis Performa Metode KNN pada Dataset Pasien Pengidap Kanker Payudara." *Indonesian Journal of Data and Science*, vol. 1, no. 2, pp. 39-43, Juli 2020. ISSN: 2715-9930.