

# Penentuan Premi GLM

## 1. Input Data

```
library(readxl)
Data_TA <- read_excel("Data-TA.xlsx", col_types = c("numeric",
  "numeric", "numeric", "numeric",
  "numeric", "numeric"))

## New names:
## • `` -> `...7`

head(Data_TA)

## # A tibble: 6 × 7
##   `Besar Klaim` `Frekuensi Klaim` penyebab  komisi    TSI `Daf
tar Okupasi`   ...7
##           <dbl>           <dbl>    <dbl>    <dbl> <dbl>
<dbl> <dbl>
## 1    92668277             3         4 500000  6.5 e9
3    NA
## 2    90324713             1         4 30890  6.5 e9
3    NA
## 3    64290115             1         2 31987  6    e9
3    NA
## 4    63548250             5         1 36876  5    e9
3    NA
## 5    58467707.             3         3 31372  4.50e9
2    NA
## 6    55396000             1         4 44215. 3.54e9
2    NA
```

## 2. Mengubah Data Kategorik Menjadi Numerik

```
DATAT = Data_TA

Y1 =(DATAT$`Besar Klaim`)
X1 <- as.factor(DATAT$`Daftar Okupasi`)
X2 <- as.factor(DATAT$penyebab)
X3 <- (DATAT$komisi)
X4 <- (DATAT$TSI)

str(DATAT)

## tibble [50 × 7] (S3: tbl_df/tbl/data.frame)
##  $ Besar Klaim      : num [1:50] 92668277 90324713 64290115 6354
8250 58467707 ...
```

```
## $ Frekuensi Klaim: num [1:50] 3 1 1 5 3 1 1 1 2 3 ...
## $ penyebab      : num [1:50] 4 4 2 1 3 4 4 4 3 3 ...
## $ komisi         : num [1:50] 500000 30890 31987 36876 31372
...
## $ TSI            : num [1:50] 6.5e+09 6.5e+09 6.0e+09 5.0e+09
4.5e+09 ...
## $ Daftar Okupasi : num [1:50] 3 3 3 3 2 2 3 3 2 3 ...
## $ ...7           : num [1:50] NA NA NA NA NA NA NA NA NA .
..
```

### 3. Statistika Deskriptif

```
summary(DATAT)

##      Besar Klaim      Frekuensi Klaim      penyebab      komisi
## Min.      :1377060   Min.      :1.000   Min.      :1.00   Min.      :
0
## 1st Qu.:10906326   1st Qu.:1.000   1st Qu.:1.00   1st Qu.: 3
1987
## Median :16970310   Median :2.000   Median :2.00   Median : 5
5043
## Mean    :24210663   Mean    :2.381   Mean    :2.46   Mean    : 13
2284
## 3rd Qu.:29124277   3rd Qu.:3.000   3rd Qu.:4.00   3rd Qu.: 9
5731
## Max.    :92668277   Max.    :9.000   Max.    :4.00   Max.    :117
0000
##
##      TSI      Daftar Okupasi      ...7
## Min.    :1.000e+08   Min.    :1.00   Min.    : NA
## 1st Qu.:3.107e+08   1st Qu.:1.00   1st Qu.: NA
## Median :5.500e+08   Median :2.00   Median : NA
## Mean    :1.634e+09   Mean    :1.76   Mean    :NaN
## 3rd Qu.:2.485e+09   3rd Qu.:2.00   3rd Qu.: NA
## Max.    :6.500e+09   Max.    :3.00   Max.    : NA
##
##      NA's      :29
##      NA's      :50
```

### 4. Identifikasi Multikolinearitas

```
library(rcompanion)

## Warning: package 'rcompanion' was built under R version 4.3.3

#Uji Korelasi Cramer
Cramer= table(X1,X2)
cramerV(Cramer)

## Cramer V
## 0.02483

#Nilai VIF
VIF = lm(Y1 ~ X3 + X4, data = DATAT)
```

```

Nilai_V = car::vif(VIF)
Nilai_V

```

```

##          X3          X4
## 1.024912 1.024912

```

Tidak terdapat multikol dalam kategorik maupun numerik

## 5. Pemodelan GLM

```

#Pemodelan Gamma GLM
gammaglm = glm(Y1 ~ X1 + X2 + X3 + X4, data = DATAT, family = Gamma,
link = "log", method = "glm.fit")
summary(gammaglm)

```

```

##
## Call:
## glm(formula = Y1 ~ X1 + X2 + X3 + X4, family = Gamma(link = "log"),
##      data = DATAT, method = "glm.fit")
##
## Coefficients:
##              Estimate Std. Error t value Pr(>|t|)
## (Intercept) 1.573e+01  1.203e-01 130.779  < 2e-16 ***
## X12         5.584e-01  1.531e-01   3.646 0.000727 ***
## X13         1.058e+00  2.526e-01   4.187 0.000141 ***
## X22         3.963e-01  1.457e-01   2.721 0.009435 **
## X23         6.069e-01  1.848e-01   3.283 0.002072 **
## X24         4.191e-01  1.436e-01   2.919 0.005621 **
## X3          9.378e-08  2.688e-07   0.349 0.728946
## X4          1.661e-10  4.608e-11   3.605 0.000821 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Gamma family taken to be 0.1370671)
##
##      Null deviance: 33.3993  on 49  degrees of freedom
## Residual deviance:  7.5151  on 42  degrees of freedom
## AIC: 1733.2
##
## Number of Fisher Scoring iterations: 6

```

Data berdistribusi Gamma dengan 7 variabel yang betpengaruh dan memiliki iterasi fisher scoring yang baik

```

#Pemodelan BN GLM
bnglm = glm.nb(Y2 ~ z1 + z2 + z3 + z4, link = "log")

## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, t
## race =
## control$trace > : iteration limit reached

```

```
## Warning in theta.ml(Y, mu, sum(w), w, limit = control$maxit, t
race =
## control$trace > : iteration limit reached

summary(bnglm)

##
## Call:
## glm.nb(formula = Y2 ~ z1 + z2 + z3 + z4, link = "log", init.th
eta = 162631.5515)
##
## Coefficients:
##              Estimate Std. Error z value Pr(>|z|)
## (Intercept)  2.064e+00  5.234e-01   3.943 8.04e-05 ***
## z12          -1.098e+00  4.352e-01  -2.524  0.01162 *
## z13          -1.951e+00  6.540e-01  -2.983  0.00285 **
## z22          -8.693e-01  6.263e-01  -1.388  0.16517
## z23          -2.719e-01  5.660e-01  -0.480  0.63093
## z24           2.398e-02  4.399e-01   0.055  0.95654
## z3           -1.599e-06  1.752e-06  -0.912  0.36151
## z4            2.616e-11  1.189e-10   0.220  0.82584
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## (Dispersion parameter for Negative Binomial(162631.6) family t
aken to be 1)
##
##      Null deviance: 25.2179  on 20  degrees of freedom
## Residual deviance:  4.3148  on 13  degrees of freedom
## (7 observations deleted due to missingness)
## AIC: 76.366
##
## Number of Fisher Scoring iterations: 1
##
##
##              Theta: 162632
##              Std. Err.: 5693776
## Warning while fitting theta: iteration limit reached
##
## 2 x log-likelihood: -58.366
```

Data berdistribusi binomial negatif dengan variabel yang berpenga  
ruh 3 dan memiliki perulangan iterasi yang baik

## 6. Uji Kebaikan Model

```
#Uji Kebaikan Besar Klaim
pearsongof = sum(gammaglm$weights*gammaglm$residuals^2)
tabb = data.frame(gofstat=c(gammaglm$deviance,pearsongof))
tabb$DF = rep(gammaglm$df.residual,2)
tabb$pvalue = pchisq(tabb$gofstat, df = tabb$DF, lower.tail = FAL
SE)
```

```

row.names(tabb) = c("deviance", "pearson"); print(tabb, digits =
3)

##           gofstat DF pvalue
## deviance    7.52 42      1
## pearson     5.76 42      1

#Uji Kebaikan Frekuensi Klaim
pearsongofd = sum(bnglm$weights*bnglm$residuals^2)
tabbf = data.frame(gofstat=c(bnglm$deviance,pearsongofd))
tabbf$DF = rep(bnglm$df.residual,2)
tabbf$pvalue = pchisq(tabbf$gofstat, df = tabbf$DF, lower.tail =
FALSE)
row.names(tabbf) = c("deviance", "pearson"); print(tabbf, digits
= 3)

##           gofstat DF pvalue
## deviance    4.31 13  0.987
## pearson     4.01 13  0.991

```

**Kedua model yang didapatkan sudah cocok dan dapat menghasilkan pe ramalan yang baik**

Nilai koefisien determinasi GLM besar klaim yang diperoleh sebagai berikut:

Hasil Koefisien Determinasi GLM Besar Klaim

<i>R- Squared</i>	<i>Adjusted R-Squared</i>
0,912	0,905

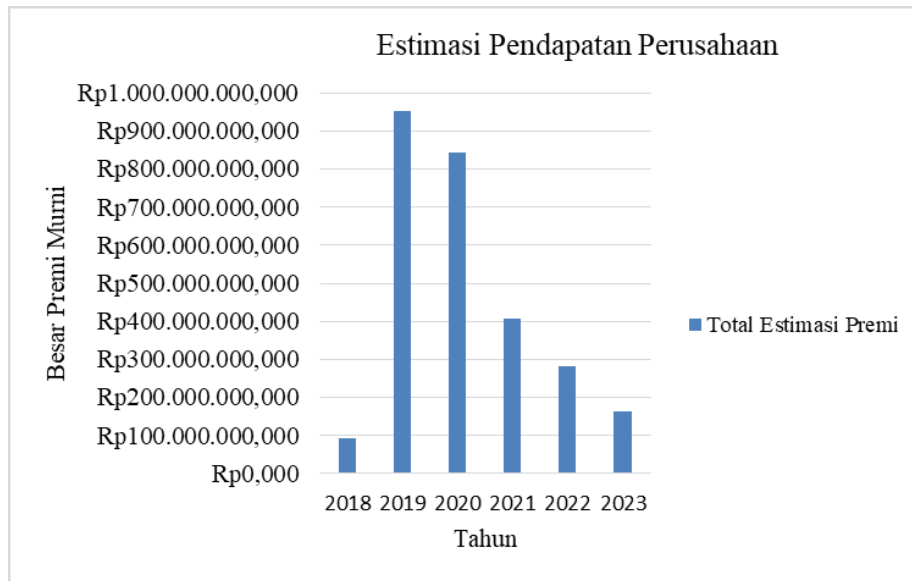
Diperoleh nilai *adjusted R-squared* sebesar 0,905 yang memiliki arti bahwa 90,5% dari besar klaim dapat dijelaskan dengan daftar okupasi kategori rendah ( $X_{11}$ ), daftar okupasi kategori sedang ( $X_{12}$ ), daftar okupasi kategori tinggi ( $X_{13}$ ), penyebab kejadian kerusakan lainnya ( $X_{21}$ ), penyebab kejadian kerusakan akibat air ( $X_{22}$ ), penyebab kejadian kebakaran ( $X_{23}$ ), penyebab kejadian gempa bumi ( $X_{24}$ ), komisi ( $X_3$ ), dan harga pertanggungan ( $X_4$ ).

Nilai koefisien determinasi GLM frekuensi klaim dengan diperoleh hasil sebagai berikut:

**Tabel 4. 1** Hasil Koefisien Determinasi GLM Frekuensi Klaim

<i>R- Squared</i>	<i>Adjusted R-Squared</i>
0,733	0,704

Berdasarkan Tabel 4.18 diperoleh nilai *adjusted R-squared* sebesar 0,704 yang memiliki arti bahwa 70,4 % dari frekuensi klaim dapat dijelaskan dengan daftar okupasi kategori rendah ( $X_{11}$ ), daftar okupasi kategori sedang ( $X_{12}$ ), daftar okupasi kategori tinggi ( $X_{13}$ ), penyebab kejadian kerusakan lainnya ( $X_{21}$ ), penyebab kejadian kerusakan akibat air ( $X_{22}$ ), penyebab kejadian kebakaran ( $X_{23}$ ), penyebab kejadian gempa bumi ( $X_{24}$ ).



Berdasarkan grafik tersebut diperoleh bahwa nilai total estimasi premi menurun seiring berjalannya tahun namun tidak separah pada tahun 2018. Estimasi premi dapat menjadi acuan bagi perusahaan dalam melihat estimasi pendapatan yang dapat digunakan sebagai tolak ukur kesehatan keuangan perusahaan untuk menghindari dari risiko kebangkrutan.