

Pemodelan Kecelakaan Perlintasan Sebidang Kereta Api dengan Pendekatan Zero-Inflated Poisson dan Zero-Inflated Negative Binomial

Efendhi Parih Raharjo^{1*}, Anisa Mahadita Candrarahayu², Wahyu Tamtama Adi³, Septiana Widi Astuti⁴,
Muhamad Nurhadi⁵, Puspita Dewi⁶
Politeknik Transportasi Darat Indonesia^{1,2},
Politeknik Perkeretaapian Indonesia Madiun^{3,4,5,6}
Corresponding Author : efendhisttd@gmail.com

ABSTRAK

Kecelakaan pada perlintasan sebidang kereta api merupakan permasalahan keselamatan transportasi yang kompleks dan dipengaruhi oleh berbagai faktor geometrik, operasional, sumber daya manusia serta kelengkapan prasarana. Data kecelakaan disetiap perlintasan pada umumnya berbentuk data hitung yang didominasi oleh nilai nol, sehingga sering kali tidak dapat dimodelkan secara optimal menggunakan regresi Poisson atau Binomial Negatif konvensional. Penelitian ini bertujuan untuk memodelkan jumlah kecelakaan pada perlintasan sebidang kereta api di wilayah DAOP 7 dengan menggunakan pendekatan *Zero-Inflated Poisson* (ZIP) dan *Zero-Inflated Negative Binomial* (ZINB). Variabel prediktor yang digunakan meliputi frekuensi perjalanan kereta api, lebar jalan raya, kemiringan persimpangan jalan raya dan rel baik dari sisi kanan maupun sisi kiri, jenis perlintasan, keberadaan *Early Warning System* (EWS), status jalan, serta jenis pintu perlintasan. Sampel penelitian ditentukan sebanyak 150 JPL dari total populasi 215 JPL di wilayah DAOP 7 Madiun menggunakan pendekatan Isaac dan Michael, sedangkan pemilihan sampel dilakukan dengan *simple random sampling*. Analisis dilakukan menggunakan perangkat lunak SPSS dan Stata dengan tahapan pemeriksaan proporsi nol, estimasi model regresi ZIP dan regresi ZINB, serta pemilihan model terbaik berdasarkan kriteria *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), dan *log-likelihood*. Hasil penelitian menunjukkan bahwa model regresi ZINB memberikan kecocokan yang lebih baik dibandingkan model regresi ZIP. Faktor-faktor yang secara signifikan berpengaruh terhadap jumlah kecelakaan pada perlintasan sebidang di wilayah DAOP 7 meliputi frekuensi, kemiringan kanan, kemiringan kiri, jenis perlintasan, keberadaan *Early Warning System* dan jenis pintu perlintasan.

Kata Kunci: Jumlah kecelakaan, Perlintasan Sebidang, Regresi Zero-Inflated Poisson, Regresi Zero-Inflated Negative Binomial.

ABSTRACT

Accidents at railroad crossings are a complex transportation safety issue influenced by various geometric, operational, human resource, and infrastructure factors. Accident data at each crossing is generally count data dominated by zero values, so it often cannot be modeled optimally using conventional Poisson or Negative Binomial regression. This study aims to model the number of accidents at railroad crossings in the DAOP 7 region using the *Zero-Inflated Poisson* (ZIP) and *Zero-Inflated Negative Binomial* (ZINB) approaches. The independent variables used include train frequency, road width, the slope of the road and rail intersection on both the right and left sides, the type of crossing, the presence of an *Early Warning System* (EWS), road status, and the type of crossing gate. The research sample consisted of 150 JPL selected from a total population of 215 JPL in the Daop 7 Madiun using the Isaac and Michael approach, while the sample selection was carried out through *simple random sampling*. The analysis was conducted using SPSS and Stata software with stages of zero proportion testing, ZIP regression and ZINB regression model estimation, and selection of the best model based on the *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), and *log-likelihood* criteria. The results of the study indicate that the ZINB regression model provides a better fit than the ZIP regression model. Factors that significantly influence the number of accidents at level crossings in DAOP 7 include frequency, right slope, left slope, type of crossing, presence of an *Early Warning System*, and type of crossing gate.

Keywords: Number of accidents, Level crossing, Zero-inflated Poisson regression, Zero-inflated negative binomial regression.

1 PENDAHULUAN

Pertumbuhan mobilitas masyarakat dan peningkatan volume lalu lintas jalan raya maupun kereta api meningkatkan potensi konflik pada titik perpotongan keduanya, khususnya pada perlintasan sebidang. Perlintasan sebidang merupakan titik rawan kecelakaan karena melibatkan interaksi langsung antara kendaraan jalan raya dan kereta api dengan perbedaan karakteristik kecepatan dan massa. Perlintasan sebidang juga memiliki tingkat risiko bahaya yang tinggi[1]. Model yang sudah digunakan untuk memodelkan kecelakaan pada perlintasan sebidang yaitu Regresi Poisson dan Regresi Binomial Negatif. Regresi ini banyak digunakan karena kesesuaiannya dengan karakteristik data hitung serta kemudahan interpretasi parameter[2], [3]. Akan tetapi data kecelakaan di perlintasan sebidang menunjukkan bahwa sebagian besar lokasi tidak mengalami kecelakaan dalam periode pengamatan tertentu, sementara sebagian kecil lokasi mengalami satu atau lebih kejadian kecelakaan.

Data kecelakaan di perlintasan sebidang pada umumnya menghasilkan data hitung dengan dominasi nilai nol. Model regresi Poisson mensyaratkan kesetaraan antara nilai rata-rata dan varians, yang dalam praktik sering kali tidak terpenuhi pada data kecelakaan. Regresi Binomial Negatif dikembangkan untuk mengatasi overdispersi, namun tetap belum sepenuhnya mampu menjelaskan keberadaan nol struktural, yaitu nol yang muncul karena suatu lokasi secara sistematis tidak berpotensi mengalami kecelakaan. Berbagai penelitian terdahulu telah menggunakan analisis statistik untuk memodelkan data kecelakaan lalu lintas dan kecelakaan perlintasan sebidang[4], [5], [6]. Salah satu yang banyak digunakan yaitu Regresi Poisson dan Regresi Binomial Negatif. Regresi ini banyak digunakan karena kesesuaiannya dengan karakteristik data hitung serta kemudahan interpretasi parameter[2], [3].

Meskipun regresi Binomial Negatif dan regresi Poisson terbukti dapat memodelkan dengan baik data-data terkait kecelakaan, beberapa penelitian menunjukkan bahwa kedua pendekatan tersebut masih memiliki keterbatasan ketika data kecelakaan didominasi oleh nilai nol [7]. Pendekatan Zero-Inflated Poisson (ZIP) dan Zero-Inflated Negative Binomial (ZINB) dikembangkan untuk mengakomodasi dua proses pembentukan data, yaitu proses yang selalu menghasilkan nol dan proses hitung yang mengikuti distribusi Poisson atau Binomial Negatif[8].

Studi internasional menunjukkan bahwa model zero-inflated efektif untuk data kecelakaan transportasi yang memiliki kelebihan nol (*excess zeros*), di mana Pew et al. melaporkan bahwa *zero-inflated negative binomial* memberikan kinerja terbaik secara keseluruhan dalam pemodelan frekuensi kecelakaan[9], sedangkan Simmachan et al. menunjukkan bahwa pada data fatalitas kecelakaan jalan yang mengandung *underdispersion* dan *excessive zeros*, model *zero-inflated Conway–Maxwell–Poisson* (ZICMP) sedikit lebih unggul dibandingkan model pembanding [10], sehingga pendekatan zero-inflated relevan digunakan untuk memperkuat analisis kecelakaan transportasi secara global. Oleh karena itu, penelitian ini mengusulkan penggunaan model ZIP dan ZINB untuk memodelkan jumlah kecelakaan pada perlintasan sebidang kereta api di wilayah Daop 7 Madiun. Penggunaan model *zero-inflated* menjadi relevan untuk memberikan gambaran yang lebih realistis terhadap mekanisme pembentukan data kecelakaan, sekaligus memperkaya pemahaman terhadap faktor-faktor risiko yang memengaruhi kejadian kecelakaan pada perlintasan sebidang. Tujuan penelitian ini adalah mengidentifikasi model yang paling sesuai dalam memodelkan data kecelakaan perlintasan sebidang dan menentukan faktor-faktor yang berpengaruh signifikan terhadap jumlah kecelakaan.

2 KAJIAN LITERATUR

Pada data perlintasan sebidang, banyak lokasi yang secara empiris tidak mengalami kecelakaan dalam periode pengamatan tertentu, baik karena tingkat paparan risiko yang rendah maupun karena keberadaan sistem pengamanan yang efektif. Kondisi ini mengindikasikan adanya nol struktural yang tidak sepenuhnya dapat dijelaskan oleh Poisson atau Binomial Negatif. *Zero inflation* adalah banyaknya nilai nol yang lebih dari 50% pada peubah respon. Pada variabel respon pada data diskrit mungkin ditemukan data bernilai nol. Akan tetapi dalam banyak kasus, nol memiliki arti penting dalam penelitian[11]. Jika nilai nol memiliki arti penting dalam data diskrit maka harus dimasukkan dalam analisis. Banyaknya nilai nol yang lebih dari 50% berakibat penggunaan regresi Poisson menjadi tidak tepat untuk memodelkan data penelitian. Regresi *Zero-Inflated Poisson* (ZIP) dan Regresi *Zero-Inflated Negative Binomial* (ZINB) digunakan untuk menganalisis data hitung yang memiliki banyak nol berlebihan. Walaupun keduanya mengatasi masalah yang sama, mereka memiliki perbedaan dalam distribusi dasar yang digunakan untuk menggambarkan data yang tidak nol. Model

Zero-Inflated Poisson (ZIP) digunakan ketika data hitung (*count data*) memiliki banyak nilai nol yang lebih banyak dari yang diprediksi oleh distribusi Poisson. Model ZIP mengasumsikan bahwa ada dua proses yang terjadi yaitu proses Bernoulli (logistik) untuk memodelkan probabilitas kehilangan data nol (*zero inflation*). Model ZIP untuk probabilitas distribusi (Y) (jumlah hitung yang diamati) dapat dituliskan sebagai:

$$P(Y = y_i) = \begin{cases} \omega_i + (1 - \omega_i)e^{-\mu_i}, & \text{untuk } y_i = 0 \\ \frac{(1 - \omega_i)e^{-\mu_i}\mu_i^{y_i}}{y_i!}, & \text{untuk } y_i > 0 \end{cases}$$

y_i = Jumlah kejadian

μ_i = Rata-rata jumlah kejadian yang terjadi dalam selang waktu tertentu

dengan $Y_i \sim ZIP(\mu, \omega)$ dan μ adalah parameter dari sebaran Poisson sedangkan ω adalah peluang kejadian bernilai nol[12]. Nilai nol pada data berasal dari komponen *count* yang mengikuti distribusi Poisson dan komponen *zero-inflation* yang memodelkan pembentukan nilai nol struktural[13]. Dalam fungsi $\mu_i = \mu_i(x_i)$ dan $\varphi_i = \varphi_i(z_i)$ memenuhi fungsi *link* sebagai berikut:

$$\log \mu_i = \sum_{j=1}^k x_{ij}\beta_j = x^T\beta$$

dan

$$\begin{aligned} \text{logit}(\varphi_i) &= \log\left(\frac{\varphi_i}{1 - \varphi_i}\right) \\ &= -\tau \sum_{j=1}^k z_{ij}\beta_j \end{aligned}$$

Model ini memiliki beberapa kelebihan antara lain sesuai untuk data yang memiliki banyak nol, yang tidak bisa dijelaskan oleh model Poisson biasa dan dapat memberikan estimasi yang lebih baik untuk data dengan pola distribusi yang tidak sesuai dengan Poisson sederhana. Akan tetapi model ini memiliki kekurangan yaitu tidak cocok untuk data yang sangat terdispersi atau memiliki overdispersi (varian lebih besar dari rata-rata) yang signifikan. Sedangkan Model *Zero-Inflated Negative Binomial (ZINB)* adalah perpanjangan dari model ZIP, yang digunakan ketika data hitung memiliki kelebihan nol dan juga menunjukkan overdispersi dimana variansnya lebih besar dari rata-rata[14]. Distribusi Negative Binomial digunakan dalam model ini untuk menggantikan distribusi Poisson karena distribusi Negative Binomial dapat menangani overdispersi lebih baik. Seperti pada model ZIP, ZINB juga mengasumsikan dua proses yaitu proses logistik untuk memodelkan probabilitas data bernilai nol dan proses *negative binomial* untuk

memodelkan data hitung yang bernilai lebih besar dari nol[15]. Model ZINB untuk probabilitas distribusi Y dapat dituliskan sebagai [16]:

$$P(Y_i = y_i) = \begin{cases} \pi_i + (1 - \pi_i)\left(\frac{1}{1 + K\lambda_i}\right)^{\frac{1}{k}}, & \text{untuk } y_i = 0 \\ (1 - \pi_i)\frac{r\left(y_i + \frac{1}{k}\right)}{r\left(\frac{1}{k}\right)y_i!}\left(\frac{1}{1 + K\lambda_i}\right)^{\frac{1}{k}}\left(\frac{K\lambda_i}{1 + K\lambda_i}\right)^{y_i}, & \text{untuk } y_i > 0 \end{cases}$$

dengan $0 \leq \pi_i \leq 1, \lambda_i \geq 0, k$ adalah parameter disperse untuk $\frac{1}{k} > 0$ dan $\Gamma(\cdot)$ adalah fungsi gamma.

Ketika $\pi_i = 0$, variable acak y_i memiliki distribusi binomial negatif dengan rata-rata λ_i dan parameter dispersi k , sehingga $Y_i \sim NB(\lambda_i, k)$. Model ZINB membentuk data menjadi dua bagian yaitu pembentukan data *count* dan mekanisme yang secara khusus menghasilkan nilai nol[17]. Komponen *count* dimodelkan menggunakan fungsi penghubung $\log \lambda_i$ sebagai berikut :

$$\ln \hat{\lambda}_i = \hat{\beta}_0 + \sum_{j=1}^p \hat{\beta}_j x_{ij}, i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, p$$

Selanjutnya, peluang terjadinya nol struktural dimodelkan melalui regresi logistik dengan fungsi logit untuk *zero state* π_i sebagai berikut:

$$\begin{aligned} \text{logit } \hat{\pi}_i &= \ln\left(\frac{\pi}{1 - \pi}\right) = \hat{\gamma}_0 \\ &+ \sum_{j=1}^p \hat{\gamma}_j x_{ij}, i = 1, 2, \dots, n \text{ dan } j = 1, 2, \dots, p \end{aligned}$$

dengan:

- p : jumlah variable prediktor
- n : jumlah pengamatan
- $\hat{\beta}_0, \hat{\gamma}_0$: intersep
- $\hat{\beta}_j, \hat{\gamma}_j$: parameter ke- j dari regresi ZINB

Beberapa kelebihan ZINB yaitu dapat menangani overdispersi yang terjadi pada data hitung, dimana varians lebih besar daripada rata-rata dan sesuai untuk data yang memiliki banyak nol berlebihan dan juga menunjukkan overdispersi dalam data positif[14].

3 METODE PENELITIAN

3.1 Data dan Variabel Penelitian

Penelitian ini menggunakan data sekunder perlintasan sebidang kereta api di wilayah DAOP 7. Variabel respon adalah jumlah kecelakaan pada masing-masing perlintasan selama periode pengamatan. Variabel prediktor yang digunakan meliputi data dengan jenis numerik dan data kategorik. Variabel frekuensi kereta yang melewati perlintasan, lebar jalan raya, kemiringan persimpangan jalan raya dan rel baik kemiringan kanan maupun kemiringan kiri merupakan data jenis numerik yang digunakan sedangkan variabel

jenis perlintasan, EWS (Early Warning System), status jalan raya dan jenis pintu perlintasan merupakan variabel kategorik. Untuk keterangan variabel kategorik dapat dilihat pada tabel 1.

Tabel 1. Variabel Kategorik

Variabel	Keterangan
Jenis perlintasan	1. Perlintasan dengan pintu 2. Perlintasan tanpa pintu
EWS	1. Berfungsi 2. Tidak berfungsi 3. Tanpa EWS
Status jalan	1. Nasional 2. Provinsi 3. Kab/Kota
Jenis pintu perlintasa	1. Elektrik 2. Mekanik 3. Portal Sederhana 4. Tanpa Pintu

Teknik pengambilan sampel dalam penelitian ini menggunakan *simple random sampling* karena seluruh JPL di wilayah Daerah Operasi 7 Madiun dinilai memiliki karakteristik yang relatif homogen sesuai dengan fokus penelitian. Populasi penelitian berjumlah 215 JPL. Penentuan jumlah sampel mengacu pada rumus Isaac dan Michael sebagaimana dikutip oleh Sugiyono (2014) dengan taraf kesalahan 5% sehingga jumlah sampel minimum yang diperoleh adalah 139 JPL[18]. Namun, untuk meningkatkan kecukupan data dan mengantisipasi kemungkinan data yang tidak lengkap, jumlah sampel penelitian ditetapkan sebanyak 150 JPL.

3.2 Metode Analisis Data

Tahapan analisis data sebagai berikut :

- Melakukan uji multikolinieritas pada variabel prediktor atau variabel penjelas. Uji multikolinieritas dilakukan untuk memastikan tidak adanya hubungan linier yang kuat antar variabel prediktor dalam model regresi. Pendeteksian multikolinieritas dilakukan menggunakan nilai *Variance Inflation Factor* (VIF). Nilai VIF yang lebih dari sepuluh akan mengindikasikan adanya multikolinieritas, sedangkan nilai VIF kurang dari sepuluh menunjukkan bahwa model tidak mengalami masalah multikolinieritas dan layak digunakan untuk analisis selanjutnya[19].

- Melakukan uji kecocokan distribusi Poisson
- Mengidentifikasi proporsi nilai nol atau *excess Zeros* untuk mengetahui apakah data memiliki nol berlebih pada variabel jumlah kecelakaan digunakan.
- Melakukan pemodelan alternatif menggunakan ZIP untuk mengatasi permasalahan *excess zero*.
- Melakukan pemodelan alternatif menggunakan ZINB untuk mengatasi permasalahan *excess zero*.
- Melakukan pengujian signifikansi parameter model regresi ZIP dan model regresi ZINB.
- Pemilihan model yang terbaik dengan membandingkan nilai *Akaike Information Criterion* (AIC), *Bayesian Information Criterion* (BIC), dan *log-likelihood* antara model ZIP dan ZINB. Model dengan nilai AIC dan BIC paling kecil serta *log-likelihood* paling besar dipilih sebagai model yang paling sesuai[20]. Analisis dilakukan menggunakan bantuan perangkat lunak SPSS dan Stata.

4 HASIL PENELITIAN DAN PEMBAHASAN

4.1 Pemeriksaan Multikolinieritas

Pemeriksaan multikolinieritas yang merupakan suatu proses untuk mengetahui adanya hubungan antara variabel prediktor dengan memperhatikan nilai VIF. Jika VIF untuk setiap variabel prediktor memiliki nilai kurang dari sepuluh, maka dapat disimpulkan bahwa tidak ada isu multikolinieritas yang perlu dikhawatirkan[21]. Berdasarkan tabel 2 terlihat bahwa seluruh variabel prediktor memiliki nilai VIF kurang dari sepuluh untuk seluruh variabel prediktor sehingga dapat disimpulkan bahwa model tidak mengalami masalah multikolinieritas dan layak digunakan untuk analisis lanjut.

Tabel 2. Nilai VIF Variabel Prediktor

Variabel Prediktor	VIF
Frekuensi	1.165
Lebar Jalan	1.717
Kemiringan Kanan	1.353
Kemiringan Kiri	1.476
Jenis Perlintasan	4.690
EWS	1.769
Status Jalan	1.397
Jenis pintu perlintasan	6.527

4.2 Uji Kecocokan Distribusi Poisson

Uji kecocokan distribusi Poisson bertujuan untuk mengetahui data berdistribusi Poisson atau tidak. Salah satu uji kecocokan distribusi Poisson yang dapat digunakan adalah uji Kolmogorov-Smirnov[21]. Hasil uji kecocokan distribusi Poisson berdasarkan output SPSS ditunjukkan pada Tabel 3.

Tabel 3. Uji Kecocokan Distribusi Poisson

Variabel Respon	Statistik Uji	P-value
Jumlah Kecelakaan (Y)	0.180	0.100

Tabel 3 menunjukkan nilai p-value sebesar 0.100 yang lebih besar dari tingkat signifikansi yang digunakan pada penelitian ini yaitu $\alpha = 0.05$. Artinya, variabel jumlah kecelakaan berdistribusi Poisson sehingga dapat dianalisis menggunakan regresi Poisson.

4.3 Pemeriksaan *Excess Zeros* pada Variabel Respon

Pemeriksaan *excess zeros* dilakukan dengan menghitung persentase amatan yang bernilai nol pada variabel respon.

Tabel 4. Persentase Excess Zero Variabel Jumlah Kecelakaan

Nilai	Jumlah	Persentase (%)
0	119	79.3
1	26	17.3
2	4	2.7
3	1	0.7

Berdasarkan tabel di atas dapat dilihat bahwa nilai nol yang ada pada variabel respon yaitu jumlah kecelakaan memiliki proporsi tertinggi dibandingkan dengan nilai lainnya. Persentase nilai nol yang ada memiliki nilai sebesar 79.33 % atau lebih dari 50% sehingga variabel jumlah kecelakaan yang didominasi oleh nilai nol ini menandakan perlunya pendekatan *zero-inflated*.

4.4 Pemodelan

Pemodelan yang menggambarkan hubungan jumlah kecelakaan kereta api diperlintasan sebidang ini akan dilakukan dengan regresi ZIP dan regresi ZINB.

a. Pemodelan regresi *Zero Inflated Poisson (ZIP)*
 Model regresi ZIP terbagi menjadi dua yaitu model *count* yang digunakan untuk menentukan peluang dari variabel respon suatu pengamatan bernilai selain nol dan model *zero-inflated* atau model logit yang digunakan untuk menentukan peluang dari variabel respon suatu pengamatan bernilai nol. Pada Tabel 5 disajikan estimasi dan pengujian parameter regresi ZIP.

Tabel 5. Hasil pengujian signifikansi parameter model *count* regresi ZIP

Variabel	Estimasi	P_Value	Keterangan
Intercept	1.995	0.044	Signifikan
X1 (Frekuensi)	0.0726	0.011	Signifikan
X2 (Lebar Jalan)	-0.196	0.046	Signifikan
X3(Kemiringan Kanan)	0.423	0.001	Signifikan
X4 (Kemiringan Kiri)	0.236	0.023	Signifikan
X5 (Jenis Perlintasan)	2.403	0.025	Signifikan
X6 (EWS)	2.339	0.010	Signifikan
X7 (Status Jalan)	1.067	0.021	Signifikan
X8 (Jenis Pintu Perlintasan)	-3.107	0.250	Tidak Signifikan

Pada penelitian ini diambil tingkat signifikansi 0.05 sehingga dari tabel 5 terlihat bahwa pada variabel frekuensi, lebar jalan, kemiringan kanan, kemiringan kiri, jenis perlintasan, keberadaan EWS dan status jalan memiliki nilai *p-value* < 0.05 atau dapat dikatakan bahwa variabel-variabel tersebut signifikan untuk menerangkan kejadian kecelakaan. Hanya terdapat satu variabel yang tidak signifikan yaitu jenis pintu perlintasan.

Tabel 6. Hasil pengujian signifikansi parameter model *zero-inflated* regresi ZIP

Variabel	Estimasi	P_Value	Keterangan
X1	18.129	0.043	Signifikan
X2	-9.456	0.010	Signifikan
X3	-12.369	0.124	Tidak Signifikan
X4	4.105	0.327	Tidak Signifikan
X5	-3.257	0.195	Tidak Signifikan
X6	3,627	0.311	Tidak Signifikan
X7	6.337	0.511	Tidak Signifikan
X8	-30.553	0.137	Tidak Signifikan

Pada bagian model inflation, hanya terdapat dua variabel yang signifikan yaitu X1 (frekuensi) dengan *P_value*=0,043 dan X2 (lebar jalan) dengan *P_value* =0,010. Koefisien X1 yang positif menunjukkan kecenderungan peningkatan probabilitas nol struktural, sedangkan koefisien X2 yang negatif menunjukkan kecenderungan penurunan probabilitas nol struktural atau probabilitas tidak mengalami kecelakaan sama sekali. Berdasarkan hasil pengujian variabel yang signifikan didapatkan model regresi ZIP sebagai berikut :

$$\log(\mu_i) = 1.995 + 0.0726x_1 - 0.196x_2 + 0.423x_3 + 0.236x_4 + 2.403x_5 + 2.339x_6 + 1.067x_7$$

$$\text{logit}(\mu_i) = 181.290x_1 - 9.456x_2$$

b. Pemodelan regresi ZINB

Model regresi ZINB dibagi menjadi dua komponen, yaitu model *count* atau model log dan model *zero-inflated* atau model logit. Berdasarkan pengujian signifikansi parameter pada tabel 7 terlihat bahwa X1, X3, X4, X5, X6, X8 memperoleh nilai *p-value* < 0.05 atau dapat dikatakan bahwa frekuensi, kemiringan kanan, kemiringan kiri, jenis perlintasan, status jalan, dan jenis pintu perlintasan berpengaruh signifikan secara statistik terhadap jumlah kecelakaan. Terdapat dua variabel yang tidak signifikan yaitu lebar jalan dan status jalan.

Tabel 7. Hasil pengujian signifikansi parameter model *count* regresi ZINB

Variabel	Estimasi	<i>P_Value</i>	Keterangan
Konstanta	1.643	0.399	Tidak Signifikan
X1	0.067	0.020	Signifikan
X2	-0.187	0.110	Tidak Signifikan
X3	0.435	0.002	Signifikan
X4	0.138	0.014	Signifikan
X5	1.795	0.121	Signifikan
X6	2.849	0.013	Signifikan
X7	0.837	0.141	Tidak Signifikan
X8	2.743	0.002	Signifikan

Sedangkan pada bagian model *zero-inflated* hanya terdapat satu variabel yang signifikan yaitu variabel lebar jalan (X2) dengan *p_value* 0.039. Koefisien X2 yang negatif menunjukkan kecenderungan penurunan probabilitas tidak mengalami kecelakaan sama sekali.

Tabel 8. Hasil pengujian signifikansi parameter model *zero-inflated* regresi ZINB

Variabel	Estimasi	<i>P_Value</i>	Keterangan
Konstanta	14.445	0.738	Tidak Signifikan
X1	18.129	0.435	Tidak Signifikan
X2	-9.456	0.039	Signifikan
X3	-12.369	0.442	Tidak Signifikan
X4	4.105	0.484	Tidak Signifikan
X5	-3.257	0.468	Tidak Signifikan
X6	3,627	0.603	Tidak Signifikan
X7	6.337	0.421	Tidak Signifikan
X8	-30.553	0.466	Tidak Signifikan

Berdasarkan hasil pengujian variabel yang signifikan didapatkan model regresi ZINB untuk model logaritma natural dan logit sebagai berikut :

$$\ln(\lambda_i) = -0.067x_1 + 0.435x_3 + 0.138x_4 + 1.795x_5 + 2.849x_6 + 2.743x_8$$

$$\text{logit}(\pi_i) = -12.369x_2$$

4.5 Model Terbaik

Pemilihan model terbaik dilakukan dengan membandingkan nilai Akaike Information Criterion (AIC), Bayesian Information Criterion (BIC), dan log-likelihood antara model ZIP dan ZINB. Model dengan nilai AIC dan BIC paling kecil serta log-likelihood paling besar dipilih sebagai model yang paling sesuai. Dari tabel 9 terlihat bahwa nilai AIC dan BIC untuk model ZINB lebih kecil dibandingkan ZIP, selain itu dari nilai *Log-likelihood* untuk model ZINB juga lebih besar dibandingkan model ZIP. Berdasarkan nilai-nilai tersebut dapat disimpulkan bahwa model regresi *zero-inflated Poisson* merupakan pilihan yang lebih baik dalam memodelkan jumlah kecelakaan di perlintasan sebidang. Hasil ini selaras dengan beberapa penelitian sebelumnya yang menunjukkan bahwa regresi ZINB lebih baik digunakan untuk memodelkan data yang mengandung *excess zero* dan overdispersi [7], [17].

Tabel 9. Nilai AIC, BIC dan *Log Likelihood*

Kriteria	Regresi <i>Zero-Inflated Poisson</i> (ZIP)	dan Regresi <i>Zero-Inflated Negative Binomial</i> (ZINB)
Akaike's Information Criterion (AIC)	134.0704	121.0989
Bayesian Information Criterion (BIC)	191.2724	183.2797
<i>Log Likelihood</i>	48,54947	48,03518

4.6 Interpretasi Model Terbaik

Berdasarkan hasil analisis model yang terbaik yaitu model regresi *Zero-Inflated Negative Binomial* (ZINB) sehingga faktor-faktor yang secara signifikan berpengaruh terhadap jumlah kecelakaan adalah frekuensi, lebar jalan, kemiringan kanan, kemiringan kiri, jenis perlintasan, keberadaan EWS, dan jenis pintu perlintasan.

Variabel frekuensi mempunyai koefisien 0.067 dan bertanda positif memberikan makna bahwa semakin padat jumlah perjalanan kereta yang melintas suatu perlintasan maka jumlah kecelakaan semakin bertambah. Koefisien ini bertanda positif, maka pada setiap peningkatan satu perjalanan

kereta api akan menaikkan jumlah kecelakaan sebesar $100(e^{0.067}-1) = 100(1.0693-1) = 6.93\%$. Variabel kemiringan kanan dengan koefisien 0.435 dapat diartikan bahwa setiap peningkatan satu satuan gradien pada sisi kanan perlintasan akan meningkatkan jumlah kecelakaan sebesar $100(e^{0.435}-1) = 100(1,544-1)=54,4\%$. Sedangkan untuk kemiringan kiri memiliki makna bahwa setiap peningkatan satu satuan gradien pada sisi kiri perlintasan akan menaikkan jumlah kecelakaan sebesar $100(e^{0.138}-1) = 100(1,148-1)=14,8\%$

Pada variabel kategorik, variabel yang signifikan yaitu jenis perlintasan, keberadaan EWS dan jenis pintu perlintasan. Variabel jenis perlintasan mempunyai koefisien 1.795 dan bertanda positif menandakan bahwa pada perlintasan kereta api yang tidak dilengkapi dengan pintu perlintasan akan mengakibatkan meningkatnya jumlah kecelakaan. Sedangkan variabel EWS bertanda positif mempunyai koefisien 2.849 memberikan makna bahwa pada perlintasan yang tidak memiliki EWS mengakibatkan meningkatnya jumlah kecelakaan. Variabel jenis pintu perlintasan mempunyai koefisien sebesar 2.743 dapat diartikan bahwa apabila perlintasan kereta api yang tidak dilengkapi dengan pintu perlintasan akan mengakibatkan meningkatnya jumlah kecelakaan.

Pada komponen *inflated*, variabel frekuensi dan status jalan berkontribusi signifikan dalam meningkatkan probabilitas suatu perlintasan tidak mengalami kecelakaan sama sekali, sedangkan variabel lebar jalan menurunkan probabilitas suatu perlintasan tidak mengalami kecelakaan sama sekali.

5 KESIMPULAN

Berdasarkan hasil pemodelan, regresi *Zero-Inflated Negative Binomial* (ZINB) merupakan model yang paling tepat untuk data kecelakaan di perlintasan sebidang karena mampu mengakomodasi karakter data yang didominasi nilai nol. Hasil uji signifikansi menunjukkan bahwa frekuensi perjalanan, lebar jalan, kemiringan kanan, kemiringan kiri, jenis perlintasan, keberadaan EWS dan jenis pintu perlintasan merupakan faktor yang mempengaruhi kejadian jumlah kecelakaan di perlintasan sebidang. Hal ini menegaskan bahwa risiko kecelakaan tidak hanya dipengaruhi oleh faktor intensitas operasi kereta, tetapi juga oleh karakteristik geometrik jalan serta tingkat perlindungan infrastruktur perlintasan. Dengan demikian, pendekatan penanganan kecelakaan perlu melibatkan rekayasa infrastruktur, pengaturan lalu lintas, dan peningkatan standar keselamatan operasional secara simultan.

6. DAFTAR PUSTAKA

- [1] S. W. Astuti, P. Dewi, M. Nurhadi, W. Nopryanto, and M. K. Cesarasyid, "Identifikasi Risiko Kecelakaan Perlintasan Sebidang di Kota Madiun," *Jurnal Teknik Industri Terintegrasi*, vol. 7, no. 1, pp. 109–118, Jan. 2024, doi: 10.31004/jutin.v7i1.22484.
- [2] M. Al Haris and P. R. Arum, "Negative Binomial Regression and Generalized Poisson Regression Models on the Number of Traffic Accidents in Central Java," *BAREKENG: Jurnal Ilmu Matematika dan Terapan*, vol. 16, no. 2, pp. 471–482, 2022, doi: 10.30598/barekengvol16iss2pp471-482.
- [3] P. S. Matematika, K. B. Jimbaran, and P. Korespondesi, "Regresi Poisson Dalam Memodelkan Jumlah".
- [4] A. Abdulhafedh, "Road Crash Prediction Models: Different Statistical Modeling Approaches," *J. Transp. Technol.*, vol. 07, no. 02, pp. 190–205, 2017, doi: 10.4236/jtts.2017.72014.
- [5] S. R. Hu, C. S. Li, and C. K. Lee, "Investigation of key factors for accident severity at railroad grade crossings by using a logit model," *Saf. Sci.*, vol. 48, no. 2, pp. 186–194, 2010, doi: 10.1016/j.ssci.2009.07.010.
- [6] D. Konur, M. M. Golias, and B. Darks, "A mathematical modeling approach to resource allocation for railroad-highway crossing safety upgrades," *Accid. Anal. Prev.*, vol. 51, pp. 192–201, 2013, doi: 10.1016/j.aap.2012.11.011.
- [7] A. Nuraeni, S. Martha, and S. Aprizkiyandari, "Penerapan Regresi Zero-Inflated Negative Binomial (Zinb) Pada Data Kecelakaan Lalu Lintas Di Kota Pontianak," vol. 11, no. 1, pp. 89–96, 2022.
- [8] J. Matematika, F. Matematika, D. A. N. Ilmu, and P. Alam, "Model Regresi Poisson Terbaik Menggunakan Zero-Inflated Poisson (Zip) Dan Zero-Inflated Negative Binomial (Zinb)," 2017.
- [9] T. Pew, R. L. Warr, G. G. Schultz, and M. Heaton, "Justification for considering zero-inflated models in crash frequency analysis,"

- Transp. Res. Interdiscip. Perspect.*, vol. 8, Nov. 2020, doi: 10.1016/j.trip.2020.100249.
- [10] T. Simmachan, N. Wongsai, S. Wongsai, and R. Lerdsuwansri, "Modeling road accident fatalities with underdispersion and zero-inflated counts," *PLoS One*, vol. 17, no. 11 November, Nov. 2022, doi: 10.1371/journal.pone.0269022.
- [11] M. Salman Al Farisyi, "PERBANDINGAN REGRESI ZERO INFLATED POISSON (ZIP) DAN REGRESI ZERO-INFLATED NEGATIVE BINOMIAL (ZINB) PADA KASUS TANAH LONGSOR DI KOTA SEMARANG JURNAL ILMIAH Diajukan sebagai salah satu syarat memperoleh gelar Sarjana Statistika." [Online]. Available: <http://repository.unimus.ac.id>
- [12] L. H. Hashim, K. H. Hashim, and M. A. K. Shiker, "An Application Comparison of Two Poisson Models on Zero Count Data," in *Journal of Physics: Conference Series*, IOP Publishing Ltd, Mar. 2021. doi: 10.1088/1742-6596/1818/1/012165.
- [13] C. Feng, "Zero-inflated models for adjusting varying exposures: a cautionary note on the pitfalls of using offset," *J. Appl. Stat.*, vol. 49, no. 1, pp. 1–23, 2022, doi: 10.1080/02664763.2020.1796943.
- [14] D. C. Jhwueng and C. Y. Wu, "A Novel Phylogenetic Negative Binomial Regression Model for Count-Dependent Variables," *Biology (Basel)*, vol. 12, no. 8, Aug. 2023, doi: 10.3390/biology12081148.
- [15] Y. N. Phang and E. F. Loh, "Statistical Analysis for Overdispersed Medical Count Data," 2014.
- [16] A. Fajri *et al.*, "Pemodelan Regresi Zero Inflated Negative Binomial pada Data yang Mengalami Overdispersi," *ESTIMASI: Journal of Statistics and Its Application*, vol. 6, no. 1, pp. 2721–379, 2025, doi: 10.20956/ejsa.v6i1.21037.
- [17] D. Nada Fitriani and yunin Sofro, "PEMODELAN KASUS PNEUMONIA BERAT PADA BALITA DI KOTA SURABAYA MENGGUNAKAN ZERO-INFLATED NEGATIVE BINOMIAL," *Jurnal Matematika Januari*, vol. 6, 2026, [Online]. Available: <https://ejurnal.unisap.ac.id/leibniz/index>
- [18] Sugiyono, *Metode Penelitian Manajemen*, vol. 1. Bandung: Alfabeta, 2014.
- [19] J. H. Kim, "Multicollinearity and misleading statistical results," *Korean J. Anesthesiol.*, vol. 72, no. 6, pp. 558–569, Dec. 2019, doi: 10.4097/kja.19087.
- [20] O. G. A, O. C. N, and O.-I. H. O, "Effect of Akaike Information Criterion on Model Selection in Analyzing Auto-crash Variables," *International Journal of Sciences: Basic and Applied Research (IJSBAR) International Journal of Sciences: Basic and Applied Research*, vol. 26, no. 1, pp. 98–109, 2016, [Online]. Available: <http://gssrr.org/index.php?journal=JournalOfBasicAndApplied>
- [21] M. Selviana, A. Atti, J. Rosalina Pahnuel, and D. Program Studi Matematika, "Tahun 2025 MODEL REGRESI POISSON UNTUK MENGANALISIS FAKTOR-FAKTOR YANG BERPENGARUH TERHADAP JUMLAH PENGANGGURAN DI NUSA TENGGARA TIMUR Robertus Dole Guntur".