

## INTRODUCTION

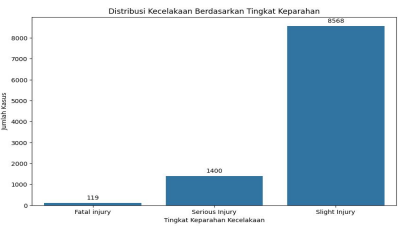
Dari laporan pihak Korlantas Polri, sepanjang periode Januari - September 2022 tercatat **94.617 kasus luka lintas** di wilayah Indonesia<sup>1</sup>. Persentase terbesar penyebab kecelakaan lalu lintas adalah **human error** sebesar 61%<sup>1</sup>. Dataset “*Road Traffic Accidents*” dianalisis lebih lanjut untuk diidentifikasi faktor-faktor penyebab kecelakaan lalu lintas. Catatan informasi atau data yang diambil memiliki **32 fitur** dan **12.316 kejadian** kecelakaan lalu lintas di Sub-Kota Addis Ababa.

## METHODOLOGY

- Preprocessing Data
  - Data Cleaning: Menghapus nilai yang memiliki *missing value*
  - Data Reduction: Menggunakan *features selection* untuk mengurangi jumlah fitur
  - Data Transformation: Mengubah bentuk data kategorik ke numerik
  - Data Normalization: Mengubah nilai data tertentu ke dalam rentang [0, 1] dengan rumus *Mix-Max Scaling*, yaitu

$$X_{norm} = \frac{X - X_{min}}{X_{max} - X_{min}}$$

- Target Class
  - Dengan menggunakan Classification, mampu mempermudah dalam mengklasifikasikan kecelakaan dalam kategori *Slight Injury*, *Serious Injury*, dan *Fatal Injury*. Berikut distribusinya



- Algoritma Klasifikasi
  - Menggunakan *Gradient Boosting Classifier* untuk membangun model klasifikasi dalam memprediksikan fitur-fitur yang diberikan sesuai dengan kategori yang telah ditentukan

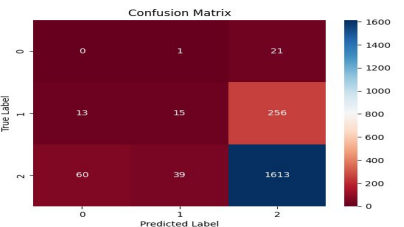
## RESULT

Penggunaan Algoritma *Gradient Boosting* menggunakan metode *SMOTEENN* dengan model *Hyperparameter Tuning* menunjukkan hasil akurasi yang tinggi, yaitu sebesar 81%. Berikut tabel hasil akurasinya:

	precision	recall	f1-score	support
0	0.00	0.00	0.00	22
1	0.27	0.05	0.09	284
2	0.85	0.94	0.9	1712

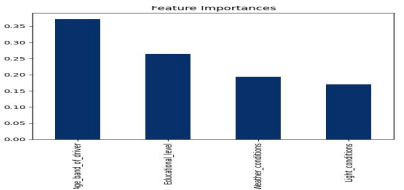
accuracy			0.81	2018
macro avg	0.38	0.33	0.33	2018
weighted avg	0.76	0.81	0.77	2018

Berikut visualisasi dengan Confusion Matrix:



Hasil Tabel Akurasi dan Confusion Matrix::

- Kelas 0 (Ringan) : memiliki masalah dalam mengidentifikasi kelas dengan benar.
- Kelas 1 (Sedang) : hanya 15 dari 284 kasus yang diprediksi benar.
- Kelas 2 (Fatal) : berhasil memprediksi 1613 dari 2018 kasus yang diprediksi benar.



Hasil dari grafik **Feature Importances**, menunjukkan fitur dengan pengaruh terbesar adalah **Age\_band\_of\_driver**. Selain menggunakan *Gradient Boosting*, percobaan dilakukan dengan 5 algoritma lainnya, berikut hasil akurasi:

Algoritma	accuracy	precision	recall	f1-score
Naive Bayes	0.54	0.75	0.54	0.62
Decision Tree	0.72	0.75	0.72	0.73
Logistic Regression	0.75	0.74	0.75	0.74
Random Forest	0.59	0.75	0.59	0.66
XGBoost	0.78	0.76	0.78	0.76

## CONCLUSION

Penyebab kecelakaan lalu lintas, diakibatkan oleh **human error**. Fitur yang memiliki pengaruh besar dalam analisis dengan model *Gradient Boosting Classifier* adalah **Age\_band\_of\_driver** dan **Educational\_level**. Pada Tahun 2023, tercatat telah terjadi 148.307 kasus kecelakaan lalu lintas yang mayoritas disebabkan oleh masyarakat dengan rentang usia 17-29 tahun<sup>2</sup>.

## REFERENCES

- Biro Komunikasi dan Informasi Publik. (2023, September 19). Tekan Angka Kecelakaan Lalu Lintas, Kemenhub Ajak Masyarakat Beralih ke Transportasi Umum dan Utamakan Keselamatan Berkendara Kementerian Perhubungan Republik Indonesia. Kementerian Perhubungan.
- Muhammad, D. A., & Maulana, A. (2024, January 23). Alasan Kenapa Remaja Sering Jadi Korban Kecelakaan Lalu Lintas