**Predicting Injury Severity in Road Accidents: A Real-Time Classification Approach**

# Column Description and Feature Engineering

David Cebulla (1922129)
Gabriel Himmelein (1649181)
Lukas Ott (1842341)
Artur Loreit (2268917)
Aaron Niemesch (1836924)

October 25, 2025

Submitted to
Data and Web Science Group
Dr. Sven Hertling
University of Mannheim

# 1 General

The columns can be divided into two main groups:

1. **Original Columns:** Data originating from the original French datasets and translated into English by `a_rename.py`.
2. **Engineered Features:** New columns calculated and added in `c_feature_engineering.py` to enrich the dataset.

# 2 Original Columns

These columns describe the core information about the accident, location, vehicles, and users.

## 2.1 General Accident Characteristics

- `year`: Year of the accident. *(Original: `an`)*
- `hour_minute`: Exact hour and minute of the accident. *(Original: `hrmn`)*
- `lighting_condition`: Lighting conditions under which the accident occurred. *(Original: `lum`)*
  - `1`: Full day
  - `2`: Twilight or dawn
  - `3`: Night without public lighting
  - `4`: Night with public lighting not on
  - `5`: Night with public lighting on
- `location`: Location of the accident. *(Original: `agg`)*
  - `1`: Outside built-up area
  - `2`: In built-up area
- `intersection`: Type of intersection. *(Original: `int`)*
  - `1`: Outside intersection
  - `2`: X-intersection
  - `3`: T-intersection
  - `4`: Y-intersection
  - `5`: Intersection with ¿4 branches
  - `6`: Roundabout
  - `7`: Place / Square

  - 8: Level crossing
  - 9: Other intersection

- `weather_condition`: Atmospheric conditions. *(Original: `atm`)*
  - -1: Not specified
  - 1: Normal
  - 2: Light rain
  - 3: Heavy rain
  - 4: Snow - hail
  - 5: Fog - smoke
  - 6: Strong wind - storm
  - 7: Dazzling weather
  - 8: Overcast
  - 9: Other

- `type_of_collision`: Type of collision. *(Original: `col`)*
  - -1: Not specified
  - 1: Two vehicles - frontal
  - 2: Two vehicles - from behind
  - 3: Two vehicles - from the side
  - 4: Three vehicles and more - in chain
  - 5: Three vehicles and more - multiple collisions
  - 6: Other collision
  - 7: Without collision

- `latitude`: Latitude (WGS84). *(Original: `lat`)*
- `longitude`: Longitude (WGS84). *(Original: `long`)*

## 2.2 Location Characteristics

- `road_category`: Category of the road. *(Original: `catr`)*
  - 1: Motorway
  - 2: National Road
  - 3: Departmental Road
  - 4: Communal Way
  - 5: Off public network

  - `6`: Public parking lot
  - `7`: Urban metropolis roads
  - `9`: Other

- `traffic_regime`: Traffic regime. *(Original: `circ`)*
  - `-1`: Not specified
  - `1`: One-way
  - `2`: Two-way
  - `3`: Separated carriageways
  - `4`: With variable assignment lanes

- `number_of_traffic_lanes`: Total number of traffic lanes. *(Original: `nbv`)*

- `reserved_lane_present`: Indicates the existence of a reserved lane. *(Original: `vosp`)*
  - `-1`: Not specified
  - `0`: Not applicable
  - `1`: Bicycle path
  - `2`: Bicycle lane
  - `3`: Reserved lane

- `longitudinal_profile`: Longitudinal profile (slope) of the road. *(Original: `prof`)*
  - `-1`: Not specified
  - `1`: Flat
  - `2`: Slope
  - `3`: Top of hill
  - `4`: Bottom of hill

- `horizontal_alignment`: Horizontal alignment (plan) of the road. *(Original: `plan`)*
  - `-1`: Not specified
  - `1`: Straight section
  - `2`: Left curve
  - `3`: Right curve
  - `4`: In "S" shape

- `carriageway_width`: Width of the carriageway (in meters), excluding central reservations or parking. *(Original: `larrout`)*

- `pavement_condition`: Condition of the road surface. *(Original: `surf`)*

  - `-1`: Not specified
  - `1`: Normal
  - `2`: Wet
  - `3`: Puddles
  - `4`: Flooded
  - `5`: Snowy
  - `6`: Mud
  - `7`: Icy
  - `8`: Grease - oil
  - `9`: Other

- `infrastructure`: Special infrastructure or facilities at the accident site. *(Original: `infra`)*
  - `-1`: Not specified
  - `0`: None
  - `1`: Underground - tunnel
  - `2`: Bridge - overpass
  - `3`: Interchange ramp or connection
  - `4`: Railway
  - `5`: Equipped junction
  - `6`: Pedestrian zone
  - `7`: Toll zone
  - `8`: Construction site
  - `9`: Others

- `accident_situation`: Situation of the accident. *(Original: `situ`)*
  - `-1`: Not specified
  - `0`: None
  - `1`: On roadway
  - `2`: On emergency lane
  - `3`: On shoulder
  - `4`: On sidewalk
  - `5`: On bicycle path
  - `6`: On other special lane
  - `8`: Others

- `speed_limit`: Authorized speed limit at the time and place of the accident. *(Original: `vma`)*

## 2.3 User Characteristics (Occupants/Pedestrians)

- `position`: Position occupied by the user in the vehicle (e.g., `1`: Driver, `2-9`: Passenger seats). *(Original: `place`)*
- `user_category`: Category of the user. *(Original: `catu`)*
  - `1`: Driver
  - `2`: Passenger
  - `3`: Pedestrian
- `role`: User's role, derived from `user_category`. *(Created in `b_table_merge.py`)*
  - `'driver'`: (From `user_category` = 1)
  - `'passenger'`: (From `user_category` = 2)
  - `'pedestrian'`: (From `user_category` = 3)
  - `'other'`: (For any other or null value)
- `injury_severity`: Severity of the user's injury. *(Original: `grav`)*
  - `1`: Uninjured
  - `2`: Killed
  - `3`: Hospitalized injured
  - `4`: Lightly injured
- `sex`: Sex of the user. *(Original: `sexe`)*
  - `1`: Male
  - `2`: Female
- `pedestrian_location`: Location of the pedestrian at the time of the accident. *(Original: `locp`)*
  - `-1`: Not specified
  - `0`: Not applicable
  - `1`: On roadway, ¿50m from pedestrian crossing
  - `2`: On roadway, ¡50m from pedestrian crossing
  - `3`: On pedestrian crossing, without traffic light
  - `4`: On pedestrian crossing, with traffic light
  - `5`: On sidewalk
  - `6`: On shoulder

  - 7: On refuge or emergency lane
  - 8: On parallel lane
  - 9: Unknown

- pedestrian_action: Action of the pedestrian. *(Original: actp)*
  - -1: Not specified
  - 0: Not specified or not applicable
  - 1: Moving in the same direction as the striking vehicle
  - 2: Moving in the opposite direction of the vehicle
  - 3: Crossing
  - 4: Masked / Hidden
  - 5: Playing - running
  - 6: With animal
  - 9: Other
  - A: Getting on/off vehicle
  - B: Unknown

## 2.4 Vehicle Characteristics

- number_vehicle_x / number_vehicle_y: Vehicle identification number within the accident (e.g., A01, B01). *(Original: num_veh)*. The _x and _y suffixes typically result from merging tables.

- direction_of_travel: Direction of travel. *(Original: senc)*
  - -1: Not specified
  - 0: Unknown
  - 1: Increasing PK/PR or postal address number
  - 2: Decreasing PK/PR or postal address number
  - 3: No landmark

- vehicle_category: Detailed vehicle category. *(Original: catv)*
  - 00: Indeterminable
  - 01: Bicycle
  - 02: Moped ¡50cm3
  - 03: Light quadricycle (car-bodied)
  - 04: Unused reference since 2006 (registered scooter)
  - 05: Unused reference since 2006 (motorcycle)

- 06: Unused reference since 2006 (side-car)
- 07: Light vehicle alone
- 08: Unused reference since 2006 (LV + caravan)
- 09: Unused reference since 2006 (LV + trailer)
- 10: Utility vehicle (1.5T ¡= GVW ¡= 3.5T)
- 11: Unused reference since 2006 (UV (10) + caravan)
- 12: Unused reference since 2006 (UV (10) + trailer)
- 13: HGV alone (3.5T ¡ GVW ¡= 7.5T)
- 14: HGV alone ¿ 7.5T
- 15: HGV ¿ 3.5T + trailer
- 16: Road tractor alone
- 17: Road tractor + semi-trailer
- 18: Unused reference since 2006 (public transport)
- 19: Unused reference since 2006 (tramway)
- 20: Special machine
- 21: Agricultural tractor
- 30: Scooter ¡ 50 cm3
- 31: Motorcycle ¿ 50 cm3 and ¡= 125 cm3
- 32: Scooter ¿ 50 cm3 and ¡= 125 cm3
- 33: Motorcycle ¿ 125 cm3
- 34: Scooter ¿ 125 cm3
- 35: Light quad ¡= 50 cm3 (non-car-bodied)
- 36: Heavy quad ¿ 50 cm3 (non-car-bodied)
- 37: Bus
- 38: Coach
- 39: Train
- 40: Tramway
- 41: 3-wheeler ¡= 50 cm3
- 42: 3-wheeler 50 cm3 ¡= 125 cm3
- 43: 3-wheeler ¿ 125 cm3
- 50: Motorized personal mobility device
- 60: Non-motorized personal mobility device
- 80: E-bike

  - `99`: Other vehicle

- `fixed_obstacle_struck`: Fixed obstacle struck. *(Original: `obs`)*
  - `-1`: Not specified
  - `0`: Not applicable
  - `1`: Parked vehicle
  - `2`: Tree
  - `3`: Metal guard rail
  - `4`: Concrete guard rail
  - `5`: Other guard rail
  - `6`: Building, wall, bridge pier
  - `7`: Road sign support or emergency call box
  - `8`: Post
  - `9`: Street furniture
  - `10`: Parapet
  - `11`: Island, refuge, high bollard
  - `12`: Curb
  - `13`: Ditch, embankment, rock wall
  - `14`: Other fixed obstacle on roadway
  - `15`: Other fixed obstacle on sidewalk or shoulder
  - `16`: Road exit without obstacle
  - `17`: Culvert

- `mobile_obstacle_struck`: Mobile obstacle struck. *(Original: `obsm`)*
  - `-1`: Not specified
  - `0`: None
  - `1`: Pedestrian
  - `2`: Vehicle
  - `4`: Vehicle on rail
  - `5`: Domestic animal
  - `6`: Wild animal
  - `9`: Other

- `initial_point_of_impact`: Initial point of impact on the vehicle. *(Original: `choc`)*
  - `-1`: Not specified

  - `0`: None
  - `1`: Front
  - `2`: Front right
  - `3`: Front left
  - `4`: Rear
  - `5`: Rear right
  - `6`: Rear left
  - `7`: Right side
  - `8`: Left side
  - `9`: Multiple impacts (rollover)

  *Note: The source document contained warnings indicating that typo values (`44` for profile, `44` for pavement, `11` for mobile obstacle) might exist in older data. The correct values are used here.*

- `main_maneuver_before_accident`: Main maneuver before the accident. *(Original: `manv`)*
  - `-1`: Not specified
  - `0`: Unknown
  - `1`: Without change of direction
  - `2`: Same direction, same lane
  - `3`: Between 2 lanes
  - `4`: Reversing
  - `5`: Wrong way
  - `6`: Crossing central reservation
  - `7`: In bus lane, same direction
  - `8`: In bus lane, opposite direction
  - `9`: Merging
  - `10`: Making a U-turn on the roadway
  - `11`: Changing lane: Left
  - `12`: Changing lane: Right
  - `13`: Veering: Left
  - `14`: Veering: Right
  - `15`: Turning: Left
  - `16`: Turning: Right
  - `17`: Overtaking: Left

  - 18: Overtaking: Right
  - 19: Crossing the roadway
  - 20: Parking maneuver
  - 21: Evasive maneuver
  - 22: Opening door
  - 23: Stopped (not parked)
  - 24: Parked (with occupants)
  - 25: Driving on sidewalk
  - 26: Other maneuvers
- motor_type: Type of vehicle motorization. *(Original: motor)*
  - -1: Not specified
  - 0: Unknown
  - 1: Hydrocarbon (Gasoline/Diesel)
  - 2: Hybrid electric
  - 3: Electric
  - 4: Hydrogen
  - 5: Human (e.g., bicycle)
  - 6: Other

## 3 Engineered Features (From c_feature_engineering.py)

These columns were newly created to model more complex relationships.

- vehicle_category_simplified: A simplified grouping of the detailed vehicle_category. Possible values: bicycle, powered_2_3_wheeler, light_motor_vehicle, hgv_truck, bus_coach, other, unknown.
- impact_score: A calculated "impact score" (weight) based on the vehicle_category_simplified (e.g., hgv_truck: 6, light_motor_vehicle: 4, bicycle: 2).
- age: The user's age at the time of the accident, calculated as year - year_of_birth.
- age_group: An age bracket based on the age: child_teen (0-17), young_adult (18-24), adult (25-39), middle_aged (40-64), senior (65+).
- used_belt: A binary indicator (1 or 0) specifying if the user used a seatbelt (based on safety_equipment_1, _2, _3 columns, where the value 1 means "Belt").
- used_helmet: A binary indicator (1 or 0) specifying if the user used a helmet (based on safety_equipment_1, _2, _3 columns, where the value 2 means "Helmet").

- `day_of_week_sin` / `day_of_week_cos`: Cyclical features (Sine/Cosine) representing the day of the week to model weekly seasonality.
- `month_sin` / `month_cos`: Cyclical features (Sine/Cosine) representing the month to model annual seasonality.
- `day_of_year_sin` / `day_of_year_cos`: Cyclical features (Sine/Cosine) representing the day of the year (more granular annual seasonality).
- `road_complexity_index`: An index (scaled from 0-10) that assesses the complexity of the road. It is based on a weighted sum of `intersection`, `road_category`, `traffic_regime`, and `number_of_traffic_lanes`.
- `surface_quality_indicator`: A binary indicator (`1`=good, `0`=bad). It is only `1` (good) if the `pavement_condition` is "Normal" (1) AND the `longitudinal_profile` is "Flat" (1).