**The “DATA.xlsx” file is the preprocessed and oversampled dataset.**

Table 1. Variable name mapping: Dataset (Persian) vs. Paper (English)

|  |  |
| --- | --- |
| Original Name (Dataset, Persian) | Translated Name (Paper, English) |
| ELAT\_TAME | Main cause |
| NOE\_BARKHORD | Crash type |
| NAHVE\_BARKHORD | Collision direction |
| NOE\_GAVAHINAME | License status |
| TAHSILAT | Education |
| NOE\_RAH | Road type |
| HAVA | Weather |
| KHATKESHI | Road marking |
| KARBARI\_MAHAL1 | Land use |
| SHARAYET\_SATH\_RAH | Road surface |
| HENDESE\_MAHAL | Road geometry |
| NOE\_TASADOF | Crash severity |
| GENSIAT | Gender |
| SEN\_RANADEH | Age |
| ROSHANAEI | Light condition |
| MOGHEAT\_TASADOF | Collision location |
| NAGHS\_RAH | Road deficiencies |
| shib | Road slope |
| Time | Time of collision |
| AMEL\_ENSANI | Human factors |

Table 2. Variable categories: Codes and corresponding labels

|  |  |  |
| --- | --- | --- |
| Attribute | Category code | Category name |
| Education | 0 | Primary education or less |
|  | 1 | High school diploma or less |
|  | 2 | Associate degree or higher education |
| Age | 2 | <18 |
|  | 0 | 18-25 |
|  | 1 | >25 |
| Gender | 0 | Man |
|  | 1 | Woman |
| Human factors | 3 | Speeding |
|  | 2 | Fatigue |
|  | 1 | Violations |
|  | 0 | Not involved |
| License status | 0 | Valid |
|  | 1 | Not valid |
| Road deficiencies | 1 | Observed |
|  | 0 | Not observed |
| Road slope | 0 | Level |
|  | 1 | Significant slope |
| Road geometry | 0 | Tangent |
|  | 1 | Curve |
| Road type | 0 | Urban major street |
|  | 1 | Urban minor street |
|  | 2 | Main road |
|  | 3 | Side road |
|  | 4 | Rural road |
|  | 5 | Highway/freeway |
|  | 6 | Other |
| Road marking | 1 | Dashed lane marking |
|  | 0 | Double/continuous marking |
|  | 2 | No marking |
| Road surface | 0 | Dry |
|  | 1 | Wet/road contaminants |
| Weather | 0 | Clear |
|  | 1 | Cloudy/foggy |
|  | 2 | Snow/rain |
| Land use | 0 | Residential |
|  | 1 | Non-residential |
| Light condition | 0 | Night |
|  | 1 | Day |
|  | 3 | Sunrise |
|  | 2 | Sunset |
| Main cause | 0 | Not paying attention to the road |
|  | 1 | Not maintaining lateral distance |
|  | 2 | Not yielding the right of way |
|  | 3 | Inability to control the vehicle |
|  | 4 | Exceeding the prescribed speed limit |
|  | 5 | Speeding |
|  | 6 | Exceeding a safe speed |
|  | 7 | Vehicle’s technical defect leading to an accident |
|  | 8 | Violation of cargo regulations |
|  | 9 | Driving in reverse |
|  | 10 | Physical or vehicle technical defect |
|  | 11 | Left deviance during overtaking |
|  | 12 | Driving in the wrong direction |
|  | 13 | Veering to the right |
|  | 14 | Sudden lane change |
|  | 15 | Suddenly opening a car door |
|  | 16 | Not maintaining a safe longitudinal distance |
|  | 17 | Passing through a prohibited area |
|  | 18 | Improper turning |
|  | 19 | Other causes |
|  | 20 | U-turn in a prohibited area |
|  | 21 | Lack of driving skills |
|  | 22 | Running a red light |
|  | 23 | Improper towing |
|  | 24 | Not maintaining lateral distance (duplicate of code 1) |
|  | 25 | Fatigue and drowsiness |
|  | 26 | Consumption of alcohol, drugs, or psychotropic substances |
|  | 27 | Other violations |
|  | 28 | Driving without valid license/allowing unlicensed drivers |
|  | 29 | Carrying oversized/overweight cargo |
|  | 30 | Using mobile phone/installing dashboard electronic devices |
| Time of collision | 0 | [17.0 - 24.0] |
|  | 1 | [0.0 - 10.0] |
|  | 2 | [10.0 - 17.0] |
| Collision direction | 0 | Sideswipe |
|  | 1 | Angle |
|  | 2 | Head-on |
|  | 3 | Rear-end |
|  | 4 | Other |
| Collision location | 0 | Carriageway lane |
|  | 1 | Roadside |
|  | 2 | Outside right-of-way |
|  | 3 | Shoulder |
| Crash type | 0 | Two vehicles involved |
|  | 1 | Multivehicle involved |
|  | 2 | Run-off-the-road/fixed object |
|  | 3 | Rollover |
|  | 4 | Motorcycle/bicycle involved |
|  | 5 | Pedestrian involved |
| Crash severity | 1 | Fatal/injury |
|  | 0 | Property damage only (PDO) |