# The Impact Scores and the Component Variables

| Impact Score | Component Variable | Dataset | Variable Importance Score | |
|---|---|---|---|---|
| | | | Naive Scoring | Decision Tree Scoring |
| **Vehicle-User Factor** | Movement (speed) | User input | 16081.6 | 2290.8 |
| | Vehicle Age | | | 153.8 |
| | User Age | | 631.4 | 143 |
| | User Sex | | 3938.3 | 673.8 |
| | Vehicle Make | | | 0.9 |
| **VUF - score** | | | | |
| **Road Features and Infrastructure** | Curvature/Geometry | *Satellite data | | |
| | Surface Type | OSM | | 121 |
| | Number of Lanes | OSM | | |
| | Conditions/Smoothness | OSM | | 27.2 |
| | Flood Prone | OSM | | |
| | Lane Markings | OSM | | |
| | Intersection | | 7391 | 120.1 |
| | Road Condition Rate | Data Gov PH | | 27.2 |
| | State highway | National Roads by Classification, Surface Type and Condition by District Engineering Office - https://data.gov.ph/?q=dataset/national-roads-classification-surface-type-and-condition-district-engineering-office | 6306.4 | 3769 |

| | | | | |
|---|---|---|---|---|
| | | ❑ Road Classification - https://psa.gov.ph/content/road-classification<br>❑ Road and Bridge Inventory - https://www.arcgis.com/apps/webappviewer/index.html?id=4bc4f2dc3a5644088c57de02108a8fd3<br>❑ National Roads by Route Numbering and Kilometer Post by District Engineering Office - https://data.gov.ph/?q=dataset/national-roads-route-numbering-and-kilometer-post-district-engineering-office | | |
| | Elevation | *Topographic maps | | |
| | Wetness of Road | *Dependent on weather? | | |
| **RF - score** | | | | |
| **Environmental Factor** | Flood ProneTime of Day | User input | | |
| | Day of Week | User input | 992.3 | |
| | Weather | | | |
| | Traffic | | | |
| | Visibility | *Satellite data | | |
| | Roadside establishments | | | |
| **EF - score** | | | | |
| **Time-Varying Elements Factor** | Average Daily Traffic Density | | | |
| | Average Traffic Speed per road segments | | | |
| | Presence or Crossing Pedestrian Density | | | |

| | | |
|---|---|---|
| **TVE - score** | | |
| **FINAL SYNTHETIC RISK SCORE (SRS)** | | |

# Kalye's Road Safety Scoring Framework

## 1. Generating the individual Impact Scores (IS) for each of the following key factors for each road segment in a road network

- ❏ [VUF] Vehicle + user factor => speed of the car, state of the driver, etc.
- ❏ [RF] Road features (infrastructures) => geometry, surface type, U-turn, intersections, etc.
- ❏ [EF] Environment factor => road wetness, visibility, etc.
- ❏ *Time-varying elements => speed of cars alongside me, presence of crossing pedestrians, etc.

For calculation of each of the 4 Impact Scores, use scoring obtained from:

- Make sure to standardize the scores before multiplying them to each factor.

## 2. Implement the Weighted Combination Modelling to calculate the final Synthetic Risk Score (SRC)

### I. Objective Method of Weights Determination: Entropy Method

$$SRS_{partial} = w_1 \cdot VUF \ + \ w_2 \cdot RF \ + \ w_3 \cdot EF \ + \ w_4 \cdot TVE$$

a) Consolidate the individual values of each Impact Score (4) for all the road segments in a road network
b) Normalize all the performance ratings

$$p_i = \frac{x_i}{\sum_{i=1}^{n} x_i}$$

c) Determine the entropy of each impact score

$$E_j = -\frac{\sum_{i=1}^{n}(p_i \ln p_i)}{\ln n}, \; j = 1, 2, 3, 4$$

d) Decide how to assign the weights based on the entropy of each impact score using

$$w_j = \frac{1-E_j}{\sum_{j=1}^{m}(1-E_j)}, \; j = 1, 2, 3, 4$$

## II. Calculate the Feature Correlation and the Weight to Assign to the Correlation Factor

You account for the interdependency/correlation between the four impact scores (if there is!) using an added variable to be added with a weighted impact.

$$w_c = \frac{1-E_j}{\sum_{j=1}^{m}(1-E_j)}$$

## III. Combine the Weighted Impact Scores and the Weighted Correlation Factor

$$SRS_{full} = w_1 \cdot VUF + w_2 \cdot RF + w_3 \cdot EF + w_4 \cdot TVE + w_5 \cdot CF$$

### IV. Encode the Raw Synthetic Risk Scores (SRS) into User-Friendly Road Safety Classification

The raw scores won't be easily interpreted by the user. For example, if Helen is driving along Road Section A and she sees the hazard risk level is 0.987, she wouldn't easily know what means. As such, the raw scores are encoded as

- ❏ Low Risk
- ❏ Medium Risk
- ❏ High Risk

Note that Kalye's labelling adopted a tripartite system **for ease of interpretation**, making it user-friendly. For example, had we used the five-level categorization of very low risk, low risk, medium risk, high risk, and very high risk, **it would be very hard for users to differentiate how to behave as driver in a High Risk vs. Very High Risk environment.**

## Kalye's Output:

- ❏ All road segments in a city's road network
- ❏ For each segment, Synthetic Risk Score (SRS) per user type
- ❏ Encoded Ris Classification on each road segment for each user type for the entire network
- ❏ Identify KEY FACTORS considered to be impactful in determining the risk/how prone a road segment is to accidents?

# Justification of the Weighted Combination Modelling Approach

In using impact factors such as the vehicle-user factors, road-related factors, environment-factors, and time-varying factors, most models would simply take the product of each to generate the risk score. The rationale is because each impact factor simply scales the overall impact into the final risk.

**However, Kalye opted for a weighted combination modelling approach** and here are the advantages:

- The weights $w_j, j = 1, 2, 3, 4, 5$ for each factor allows determination of feature important. The **feature importance justifies the validity of the model** (i.e. why is the risk score that high/low?). In other words, it aids in the interpretability of the model.
- The feature importance basically helps identify which of the four impact factors is more impactful to road hazards. This leads to **design applications**. For example:
    - If road-related factors are deemed as most impactful factor, **then during urban developments, you can use the model to optimize between features of the road** to minimize road hazards!
    - If vehicle-user factors are seen to be very impactful factors, i**t will aid policy makers to craft public policies that helps control vehicle-user behaviour** to avoid hazards! Examples would be implementation of speed limits, monitoring of driver behaviour, etc.
- Lastly, the model Kalye devised **accounts for the interdependence of the impact factors** which is unavoidable and relevant to account for a better picture of risk assessments!