**Diccionario de datos**

**Dataset**: Canadian Car Accidents

Práctica 1 – Machine Learning

*Máster Universitario en Ciencia de Datos*

Álvaro Serrano del Rincón (*a.serranodelricon@cunef.edu*)

Carlos Viñals Guitart(*carlos.vinals@cunef.edu*)

**Descripción General**

Este documento pretende aportar información descriptiva del contenido y forma del *dataset* a utilizar en la práctica de la asignatura de Aprendizaje Automático del Máster Universitario en Ciencia de Datos de CUNEF Universidad.

El *dataset* a utilizar es **Canadian Car Accidents** (NCDB\_1999\_to\_2014.csv). Este proporciona información relativa a accidentes por colisiones de vehículos en Canada. En este diccionario nos disponemos a describir cada una de las variables que posee, su significado, posibles valores que puede tomar y otros aspectos técnicos relevantes para el manejo de este *dataset* y que permitan la elaboración de modelos.

A prior, podemos comentar que este *dataset* posee un total de 22 variables de estudio o columnas. Que se agrupan en tres grupos: Datos relativos al accidente, Datos relativos al vehículo y Datos relativos a la persona(s) involucradas.

**Listado de Variables General**

**Datos del accidente:**

C\_YEAR, C\_MONTH, C\_WDAY, C\_HOUR, C\_SEV, C\_VEHS, C\_CONF, C\_RCFG, C\_WTHR, C\_RSUR, C\_RALN, C\_TRAF

**Datos del vehículo:**

V\_ID, V\_TYPE, V\_YEAR

**Datos de la persona:**

P\_ID, P\_SEX, P\_AGE, P\_PSN, P\_ISEV, P\_SAFE, P\_USER

# **Datos relativos al accidente**

## **C\_YEAR**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_YEAR* | *Año cuando se produjo el accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** |
| *Numérica* | *Numeric (Int)* | *Valores comprendidos entre los años 19XX y 20XX. Siendo XX los dos últimos dígitos del año en cuestión.* |

## **C\_MNTH**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_MNTH* | *Código del mes cuando se produjo el accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica (Categórica)* | *Numeric (Int)* | |  |  |  | | --- | --- | --- | | *01:Enero*  *02:Febrero*  *03:Marzo*  *04:Abril*  *05:Mayo*  *06:Junio* | *07:Julio*  *08:Agosto*  *09:Septiembre*  *10:Octubre*  *11:Noviembre*  *12:Diciembre* | *UU:Desconocido*  *XX:Clasificado* | |

## **C\_WDAY**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_WDAY* | *Día de la semana cuando se produjo el accidente* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica (Categórica)* | *Numeric (Int)* | |  |  |  | | --- | --- | --- | | *1:Lunes*  *2:Martes*  *3:Miércoles*  *4:Jueves* | *5:Viernes*  *6:Sábado*  *7:Domingo* | *U:Desconocido*  *X:Clasificado* | |

## **C\_HOUR**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_HOUR* | *Intervalo horario cuando se produjo el accidente* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica (Categórica)* | *Numeric (Int)* | |  |  | | --- | --- | | *00:(00:00 a 00:59)*  *01:(01:00 a 01:59)*  *02:(02:00 a 02:59)*  *03:(03:00 a 03:59)*  *04:(04:00 a 04:59)*  *05:(05:00 a 05:59)*  *06:(06:00 a 06:59)*  *07:(07:00 a 07:59)*  *08:(08:00 a 08:59)*  *09:(09:00 a 09:59)*  *10:(10:00 a 10:59)*  *11:(11:00 a 11:59)*  *12:(12:00 a 12:59)* | *13:(13:00 a 13:59)*  *14:(14:00 a 14:59)*  *15:(15:00 a 15:59)*  *16:(16:00 a 16:59)*  *17:(17:00 a 17:59)*  *18:(18:00 a 18:59)*  *19:(19:00 a 19:59)*  *20:(20:00 a 20:59)*  *21:(21:00 a 21:59)*  *22:(22:00 a 22:59)*  *23:(23:00 a 23:59)*  *UU:Desconocido*  *XX:Clasificado* | |

## **C\_SEV**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_SEV* | *Código que indica la severidad del accidente, es decir, si hubo fallecimientos o no.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica (Categórica)* | *Numeric (Int)* | |  |  | | --- | --- | | *1:Mínimo un fallecido*  *2:Sin fallecidos* | *U:Desconocido*  *X:Clasificado* | |

## **C\_VEHS**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_VEHS* | *Número de vehículos involucrados en el accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | |  |  | | --- | --- | | *Entre [01 – 98]*  *99:99 o más vehículos* | *UU:Desconocido*  *XX:Clasificado* | |

## **C\_CONF**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_CONF* | *Tipología del accidente o descripción del mismo.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Categórica* | *Texto* | *01:Un solo vehículo – Colisión con objeto en movimiento.*  *02:Un solo vehículo – Colisión con objeto estacionario.*  *03:Un solo vehículo – Cuneta izquierda (vuelta de campana)*  *04:Un solo vehículo – Cuneta derecha (vuelta de campana)*  *05:Un solo vehículo – Vuelta campana en asfalto.*  *06:Un solo vehículo – Otro tipo de colisión.*  *21:Dos vehículos en mismo sentido – Colisión trasera.*  *22:Dos vehículos en mismo sentido – Colisión lateral.*  *23:Dos vehículos en mismo sentido – Adelantamiento por la izquierda.*  *24:Dos vehículos en mismo sentido – Adelantamiento por la derecha.*  *25:Dos vehículos en mismo sentido – Otro tipo de colisión.*  *31:Dos vehículos en sentido opuesto – Colisión frontal.*  *32:Dos vehículos en sentido opuesto – Colisión lateral.*  *33:Dos vehículos en sentido opuesto – Giro izquierda en cruce.*  *34:Dos vehículos en sentido opuesto – Giro derecha en cruce.*  *35:Dos vehículos en sentido opuesto – Colisión ángulo derecho.*  *36:Dos vehículos en sentido opuesto – Otro tipo de colisión.*  *41:Dos vehículos – Colisión con vehículo aparcado.*  *QQ:Otro tipo de colisión*  *UU:Desconocido*  *XX:Clasificado* |

## **C\_RCFG**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_RCFG* | *Tipología de la carretera o situación del lugar donde ocurrió el accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Categórica* | *Texto* | *01:Normal (no es un cruce)*  *02:Intersección entre dos carreteras públicas.*  *03:Intersección con salida de parking/carretera privada.*  *04:Cruce de vías de tren.*  *05:Viaducto.*  *06:Tunel.*  *07:Carril de adelantamiento o carril lento.*  *08:Rampa.*  *09:Rotonda.*  *10:Carril rápido en autovía/autopista*  *11:Carril de traspaso en autovía/autopista*  *12:Carril de salida en autovía/autopista*  *QQ:Otro tipo de carretera*  *UU:Desconocido.*  *XX:Clasificado.* |

## **C\_WTHR**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_WTHR* | *Condiciones climatológicas cuando ocurrió el accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Categórica* | *Texto* | *01:Despejado y soleado.*  *02:Nublado, sin precipitaciones.*  *03:Lluvía.*  *04:Nevando, sin posibilidad de deslizamientos.*  *05:Lluvía congelada, aguanieve, granizo.*  *06:Visibilidad limitada.*  *07:Viento fuerte.*  *QQ:Otra situación climatológica.*  *UU:Desconocido.*  *XX:Clasificado.* |

## **C\_RSUR**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_RSUR* | *Condiciones de la superficie del asfalto.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Categórica* | *Texto* | |  |  | | --- | --- | | *1:Seco (normal).*  *2:Mojado.*  *3:Nevado (reciente).*  *4:Fango, nieve “húmeda”*  *5:Helado.*  *6:Tierra, grava, barro seco.* | *7:Lodo*  *8:Aceite (vertido).*  *9:Inundado*  *Q:Otra condición del asfalto.*  *U:Desconocido.*  *X:Clasificado.* | |

## **C\_RALN**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_RALN* | *Alineación de la carretera, pendiente, …* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Categórica* | *Texto* | |  |  | | --- | --- | | *1:Recta y nivelada.*  *2:Recta y pendiente.*  *3:Curva y nivelada.*  *4:Curva y pendiente.*  *5:Cima de colina o pendiente.*  *6:Principio de colina o pendiente.* | *Q:Otra situación.*  *U:Desconocido.*  *X:Clasificado.* | |

## **C\_TRAF**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *C\_TRAF* | *Estado de las medidas de señalización y seguridad presentes en la vía.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Categórica* | *Texto* | *01:Señales totalmente operativas.*  *02:Señales en estado intermitente.*  *03:Señal de stop.*  *04:Señal de ceda el paso.*  *05:Señal de advertencia (triángulo amarillo).*  *06:Paso de peatones.*  *07:Oficial de policía.*  *08:Voluntario de escuela.*  *09:Paso escolar(peatones)*  *10:Zona de velocidad reducida.*  *11:Señal de prohibido el paso.*  *12:Marcas en el asfalto.*  *13:Bus escolar parado con señales intermitentes.*  *14:Bus escolar parado sin señales intermitentes.*  *15:Cruce de vías con semáforos y/o barreras de seguridad.*  *16:Cruce de vías con señales.*  *17:Señalización o sistema de seguridad no especificado.*  *18:Señalización o sistema de seguridad inexistente.*  *QQ:Otro tipo de situación.*  *UU:Desconocido.*  *XX:Clasificado.* |

# **Datos relativos al vehículo**

## **V\_ID**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *V\_ID* | *Identificador del vehículo respecto a la persona.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | *Valor comprendido entre [1-98]*  *99:Peatones*  *UU:Desconocido.* |

## **V\_TYPE**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *V\_TYPE* | *Código del tipo de vehículo.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica*  *(Categórica)* | *Numeric (Int)* | *01:Vehículo de servicio (pasajeros, servicios auxiliares, emergencias, …)*  *02:Camioneta (<=4536 KG GVWR)*  *03:Otros vehículos de carga (<=4536 KG GVWR)*  *04:Camiones (>4536 HG GVWR)*  *08:Tractor (con o sin remolque)*  *09:Bus escolar*  *10:Bus escolar pequeño (<25 pasajeros)*  *11:Bus urbano*  *14:Motocicleta*  *16:Vehículo ajeno a la carretera*  *17:Bicicleta.*  *18:Vehículos de propósito específico.*  *19:Remolques o herramientas de arado.*  *20:Remolques o herramientas de construcción.*  *21:Camión de bomberos.*  *22:Quitanieves.*  *23:Coche de ciudad.*  *NN:No aplicable (para el vehículo con id:99 aka peatón)*  *QQ:Otro tipo de vehículo*  *UU:Desconocido.*  *XX:Clasificado.* |

## **V\_YEAR**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *V\_YEAR* | *Año de fabricación del modelo de vehículo.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | *Valores comprendidos entre 19YY y 20YY*  *NNNN:No aplicable*  *UUUU:Desconocido*  *XXXX:Clasificado* |

# **Datos relativos a la persona**

## **P\_ID**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_ID* | *Identificador de la persona.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | *Valores comprendidos entre [1-99]*  *NN:No aplicable (dummy?)*  *UU:Desconocido.* |

## **P\_SEX**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_SEX* | *Sexo de la persona.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | *F:Mujer*  *M:Hombre*  *N:No aplicable (dummy?)*  *U:Desconocido*  *X:Clasificado* |

## **P\_AGE**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_AGE* | *Edad de la persona.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | *00:Menos de 1 año.*  *Valores comprendidos entre [1-98]*  *99:99 o más años.*  *NN:No aplicable (dummy?)*  *UU:Desconocido*  *XX:Clasificado* |

## **P\_PSN**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_PSN* | *Posición respecto al vehículo en el momento del accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica* | *Numeric (Int)* | *11:Conductor*  *VALOR(X)VALOR(Y):Fila del coche(X) y posición(Y)*  *Ejemplo: 21:Segunda fila , posición izquierda*  *96:Desconocida, pero dentro del vehículo*  *97:Sobre el regazo de otro pasajero.*  *98:Fuera del vehículo, pero en él.*  *99:Pasajero*  *QQ:Otro valor*  *NN:No aplicable (dummy?)*  *UU:Desconocido.*  *XX:Clasificado.* |

## **P\_ISEV**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_ISEV* | *Nivel de severidad en la persona del accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica*  *(Categórica)* | *Numeric (Int)* | *1:Ilesa*  *2:Herida*  *3:Fallecida*  *N:No aplicable (dummy?)*  *U:Desconocido*  *X:Clasificado* |

## **P\_SAFE**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_SAFE* | *Elemento de seguridad presente en el momento del accidente.* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica*  *(Categórica)* | *Numeric (Int)* | *01:Ningún sistema de seguridad utilizado.*  *02:Se utiliza un sistema de seguridad.*  *09:Casco*  *10:Chaleco reflectante.*  *11:Casco y chaleco reflectante.*  *12:Otro tipo de sistema de seguridad.*  *13:No tiene sistemas de seguridad.*  *NN:No aplicable.*  *QQ:Otra opción*  *UU:Desconocido.*  *XX:Clasificado.* |

## **P\_USER**

|  |  |  |
| --- | --- | --- |
| **Variable** | **Descripción** | |
| *P\_SAFE* | *Tipo de usuario* | |
| **Tipo** | **Tipo de Dato** | **Valores** (Valor:Referencia) |
| *Numérica*  *(Categórica)* | *Numeric (Int)* | *1:Conductor*  *2:Pasajero*  *3:Peatón*  *4:Ciclista*  *5:Conductor de motocicleta.*  *U:Desconocido* |