

Diccionario de datos

Dataset: Canadian Car Accidents

Práctica 1 – Machine Learning

Máster Universitario en Ciencia de Datos

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 (Categoría)	Numeric (Int)	01:Enero 07:Julio UU:Desconocido 02:Febrero 08:Agosto XX:Clasificado 03:Marzo 09:Septiembre 04:Abril 10:Octubre 05:Mayo 11:Noviembre 06:Junio 12:Diciembre

C_WDAY

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

C_HOUR

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

C_SEV

Variable	Descripción		
<i>C_SEV</i>	<i>Código que indica la severidad del accidente, es decir, si hubo fallecimientos o no.</i>		
Tipo	Tipo de Dato	Valores (Valor:Referencia)	
<i>Númerica (Categorica)</i>	<i>Numeric (Int)</i>	<i>1:Mínimo un fallecido</i> <i>2:Sin fallecidos</i>	<i>U:Desconocido</i> <i>X:Clasificado</i>

C_VEHS

Variable	Descripción		
<i>C_VEHS</i>	<i>Número de vehículos involucrados en el accidente.</i>		
Tipo	Tipo de Dato	Valores (Valor:Referencia)	
<i>Númerica</i>	<i>Numeric (Int)</i>	<i>Entre [01 – 98]</i> <i>99:99 o más vehículos</i>	<i>UU:Desconocido</i> <i>XX:Clasificado</i>

C_CONF

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

C_RCFG

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

C_WTHR

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

C_RSUR

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

C_RALN

Variable	Descripción	
<i>C_RALN</i>	<i>Alineación de la carretera, pendiente, ...</i>	
Tipo	Tipo de Dato	Valores (Valor:Referencia)
<i>Catagórica</i>	<i>Texto</i>	<i>1:Recta y nivelada.</i> <i>2:Recta y pendiente.</i> <i>3:Curva y nivelada.</i> <i>4:Curva y pendiente.</i> <i>5:Cima de colina o pendiente.</i> <i>6:Principio de colina o pendiente.</i> <i>Q:Otra situación.</i> <i>U:Desconocido.</i> <i>X:Clasificado.</i>

C_TRAF

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

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	
<i>P_ID</i>	<i>Identificador de la persona.</i>	
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	
<i>P_SEX</i>	<i>Sexo de la persona.</i>	
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	
<i>P_AGE</i>	<i>Edad de la persona.</i>	
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	
<i>P_PSN</i>	<i>Posición respecto al vehículo en el momento del accidente.</i>	
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	
<i>P_ISEV</i>	<i>Nivel de severidad en la persona del accidente.</i>	
Tipo	Tipo de Dato	Valores (Valor:Referencia)
<i>Numérica (Categorica)</i>	<i>Numeric (Int)</i>	<i>1:Illesa 2:Herida 3:Fallecida N:No aplicable (dummy?) U:Desconocido X:Clasificado</i>

P_SAFE

Variable	Descripción	
<i>P_SAFE</i>	<i>Elemento de seguridad presente en el momento del accidente.</i>	
Tipo	Tipo de Dato	Valores (Valor:Referencia)
<i>Numérica (Categorica)</i>	<i>Numeric (Int)</i>	<i>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.</i>

P_USER

Variable	Descripción	
<i>P_SAFE</i>	<i>Tipo de usuario</i>	
Tipo	Tipo de Dato	Valores (Valor:Referencia)
<i>Numérica (Categorica)</i>	<i>Numeric (Int)</i>	<i>1:Conductor 2:Pasajero 3:Peatón 4:Ciclista 5:Conductor de motocicleta. U:Desconocido</i>