*Seminario de Proyectos*

*3C - Proyecto integrador - Pregunta general y específicas de la investigación*

*Nombre: David Aaron Ramirez Olmeda*

*Programa: Maestría en Ciencia de Datos e Información*

**Introducción**

En esta unidad nos enfocaremos en definir las preguntas específicas para nuestro proyecto "Predicción de la tasa de accidentes viales en Sinaloa, México, mediante modelos avanzados de ciencia de datos".

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
|  | | Pregunta | Objetivo | Hipótesis |
| General |  | ¿Cómo se puede desarrollar un modelo de ciencia de datos que permita predecir de manera efectiva la tasa de accidentes viales en el estado de Sinaloa, México, considerando variables relevantes y aplicando modelos avanzados de ciencia de datos? | *Desarrollar un modelo de ciencia de datos que permita predecir la tasa de accidentes viales en Sinaloa, México.* |  |
| Específico | 1 | ¿Cuáles son los pasos esenciales para desarrollar la base de datos que sustentará el modelo de predicción de accidentes viales en Sinaloa, México? | Desarrollar la base de datos y seleccionar variables relevantes de los datos de INEGI sobre accidentes viales, realizar el preprocesamiento, limpiar, unir y filtar otros datos relevantes con el fin de utilizarlos como base de datos para predecir accidentes viales en Sinaloa, México |  |
| 2 | ¿Qué criterios específicos justificarán la elección de algoritmos para la predicción de la tasa de accidentes viales en Sinaloa? | Seleccionar y entrenar el algoritmo para predecir accidentes viales en Sinaloa, México, considerando su rendimiento, ajustarlo y justificar su elección. |  |
| 3 | ¿Qué diferencias significativas se esperan al contrastar los resultados obtenidos utilizando diferentes algoritmos, considerando aspectos como interpretación, eficiencia y calidad predictiva de nuestras predicciones sobre la tasa de accidentes viales en Sinaloa, México? | Analizar y contrastar el modelo ajustado evaluando su rendimiento, revisar los valores del número de accidentes viales generados para Sinaloa, Mexico, para evaluar las predicciones y explorarlas visualmente |  |

|  |  |
| --- | --- |
| Variable dependiente | **Número de accidentes** (número de accidentes que se contabiliza de acuerdo a nuestras variables independientes.  Obtenido de la cantidad histórica de accidentes [lo que ocurrió] como de la ejecución del modelo [lo que ocurrirá]. ) |
| Variables independientes | * Municipio y estado * Latitud y longitud * Hora, día, mes, año * Horario (día, noche, atardecer, etc.) * Lugar (zona, colonia) * Clima (condición atmosférica) * Densidad de población * Eventos especiales (carnavales, ferias, etc.) * Factores socioecnonómicos (nivel socioeconómico de la zona) * Entre otras relevantes que podamos identifcar en el proceso |

*Nota: El número de accidentes es una variable predecible por nuestro(s) modelo(s) y nos permite calcular la tasa anual, mensual, etc., tomando como caso de estudio la zona particular de Sinaloa, México*