Uno de los principales desafíos en el modelado predictivo a través de técnicas de aprendizaje automático es la selección de características. En muchos casos, las variables relevantes para desarrollar un modelo predictivo de elevada exactitud pueden ser no solo las recomendadas por la literatura y la bibliografía, pudiendo existir otras variables cuya inclusión conjunta con las sustentadas por bibliografía potencien la exactitud de los modelos predictivos desarrollados. Por este motivo un abordaje utilizado en el modelado predictivo es la selección de características a partir de grandes bases de datos, con el objetivo de poder discernir las más relevantes a la hora de desarrollar el modelo predictivo.

En este caso particular se trata del modelado predictivo de la concentración de material particulado fino a partir de variables satelitales. Para abordar el problema se creó una amplia base de datos de variables satelitales, que involucra no solo aquellas variables recomendadas por bibliografía, sino que numerosas que no se hallan explícitamente sustentadas.

En este contexto, las técnicas de selección de características son una herramienta útil para reducir la complejidad y seleccionar las variables de mayor relevancia para desarrollar el modelo predictivo. Es importante destacar que, en algunos casos, las variables seleccionadas pueden no tener un respaldo explícito en la literatura existente, sin embargo, esto no significa que estas sean irrelevantes o que el modelo sea incorrecto. En muchos casos, estas pueden estar relacionadas de manera compleja e indirecta con el fenómeno que se está modelando. De esta forma su inclusión al modelo conjuntamente con las variables sustentadas por bibliografía permite aumentar la exactitud de las predicciones otorgadas. Gracias a la implementación de técnicas de aprendizaje automático, es que muchas de estas vinculaciones pueden ser desenmascaradas y estudiadas, permitiendo derivar una interpretación teórica de su influencia.

De esta forma los modelos predictivos desarrollados para la concentración de material particulado fino, no solo incluyen las variables destacadas como relevantes según la bibliografía, sino que también incorporan otras variables satelitales que permiten aumentar la exactitud de las predicciones obtenidas. En conclusión, la aplicación de técnicas de selección de características y aprendizaje automático puede proporcionar una forma efectiva de modelar los niveles de concentración de material particulado fino en función de variables satelitales. Si bien es importante tener en cuenta las limitaciones y el contexto específico del problema, estas técnicas pueden permitir la identificación de relaciones entre variables previamente ocultas, lo que puede ser de gran importancia para la comprensión y la predicción del fenómeno.