

Sesgo, varianza y ética

Oscar E. Velandia G. CSPA, director de Actuaría
Fasecolda

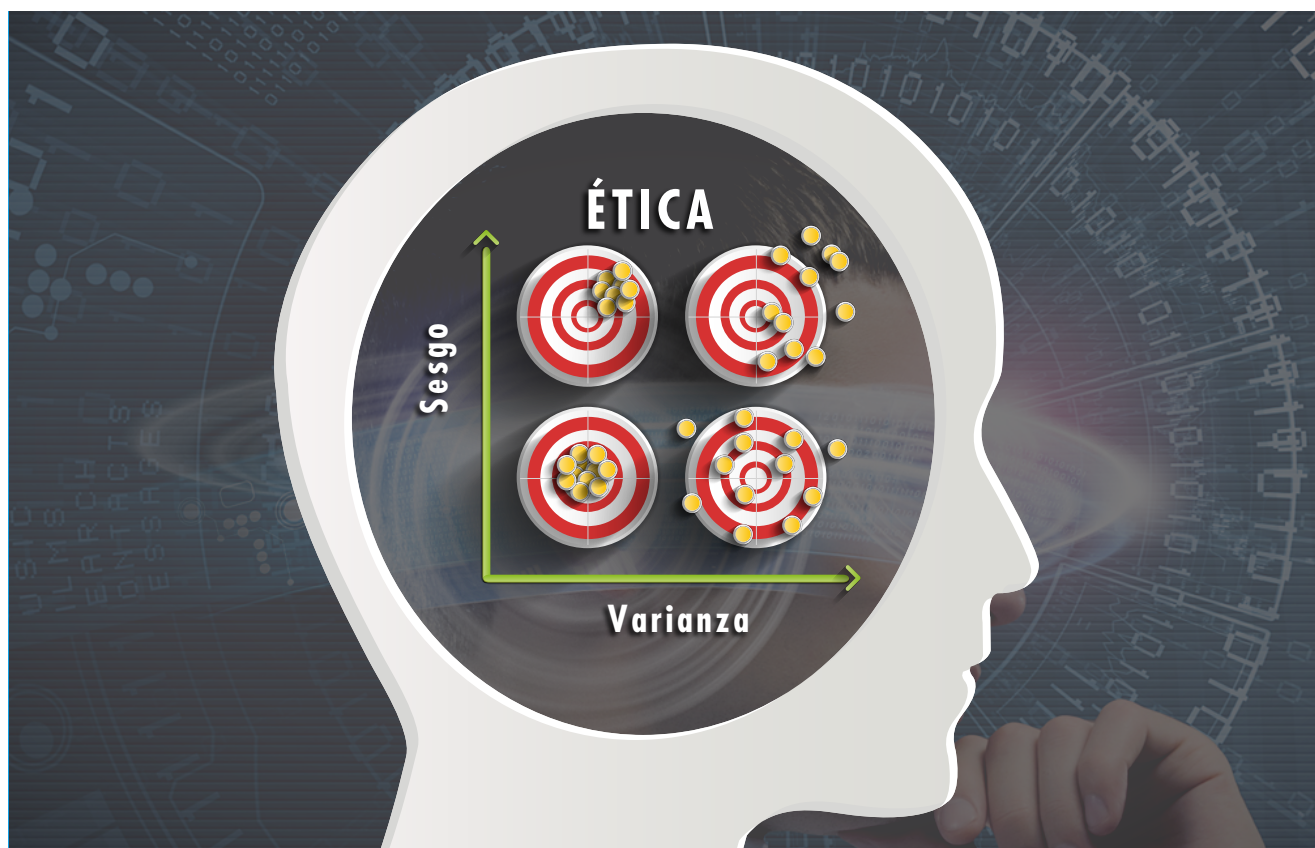
Con el auge de la ciencia de datos y el aprendizaje profundo, los conceptos de sesgo, varianza y ética cobran mayor relevancia y tienen nuevas interpretaciones. No obstante, el no entendimiento o mal manejo de estos conceptos generan limitantes en la adopción y aplicación de técnicas de aprendizaje automático o profundo al interior de las industrias.

En estadística, el sesgo es la diferencia entre el valor esperado de un estimador y su estimado (estimado se entiende como el verdadero valor del parámetro a estimar). Algo suficientemente técnico, además la definición de sesgo puede variar dependiendo del contexto. Tal como lo cita Cassie Kozyrkov¹ «El problema del sesgo en inteligencia artificial empieza -pero no termina- con la definición misma del término “sesgo”. Este término está sobrecargado y tiene un significado drásticamente distinto bajo diferentes contextos.», entonces no es único y hay de distintas clases. En

el ámbito actuarial, se trabaja constantemente con el sesgo de información, que corresponde al uso de información que inicialmente fue construida con un fin específico y no para un estudio actuarial.

Para ilustrar este concepto, considere el caso en que una entidad aseguradora tenga únicamente las bases de datos de administración de relaciones con el cliente (CRM), cartera, pólizas y siniestros. Cada una de ellas por sí sola no aporta suficientes datos para tareas generalizadas de pronóstico, ya que fueron diseñadas para tareas

1. Cassie Kozyrkov, La Verdad Sobre el Sesgo en Inteligencia Artificial, <https://medium.com/datos-y-ciencia/la-verdad-sobre-el-sesgo-en-inteligencia-artificial-5e228be3aee7>.



específicas, por lo cual, para la creación e implementación de algoritmos de aprendizaje automático o profundo, los equipos de Actuaría o Ciencia de Datos trabajan continuamente en la construcción de bases adecuadas, que sean lo más pobladas, confiables y limpias, pero condicionadas al entendimiento² de los datos disponibles. Esto implica que la información siempre estará sesgada al tratamiento, limpieza y el entendimiento que se posea, asociado siempre a la madurez analítica de los equipos técnicos de cada compañía³.

Por otra parte, la varianza se puede entender como la dispersión en la estimación de un estimador, de nuevo otro concepto técnico. Y en la literatura se habla del *trade-off* o compensación entre sesgo y varianza⁴, también de modelos flexibles y medidas de rendimiento o del ajuste de los algoritmos⁵, lo cual es distinto a las tradicionales pruebas de hipótesis de significancia de un modelo o de los parámetros, los cuales son ampliamente utilizados en modelos de tarificación. Todos sabemos que dentro de una

2. Los algoritmos sesgados están por todas partes, y parece que a nadie le importa, MIT technology review <https://www.technologyreview.es/s/8344/los-algoritmos-sesgados-estan-por-todas-partes-y-parece-que-nadie-le-importa>
3. Cómo se produce el sesgo algorítmico y por qué es tan difícil detenerlo, MIT technology review <https://www.technologyreview.es/s/10924/como-se-produce-el-sesgo-algoritmico-y-por-que-es-tan-dificil-detenerlo>
4. <https://aprendeia.com/bias-y-varianza-en-machine-learning/>
5. Utilización del machine learning en la industria 4.0, <https://1library.co/document/y9gl78jq-utilizacion-del-machine-learning-en-la-industria.html>

compañía la «magia» que hacen los actuarios está principalmente basada en el entendimiento tradicional de la estadística y matemática actuarial, que abarca muchos de estos conceptos; generalmente no se sabe qué es, pero se confía y se les da la responsabilidad de hacerlo bien.

Sin embargo, al presentar cualquier resultado siempre se pide una explicación de estos y si dicha explicación no es intuitiva o natural al sentir de los tomadores de decisiones, se puede generar resistencia o un problema de comunicación. Siempre teniendo presente que los modelos tradicionales son mecanismos generalmente analíticos, que pretenden la explicación de los fenómenos y que son de más fácil interpretación a diferencia de los de aprendizaje automático.

➔ Dado que el Gobierno colombiano quiere que el marco ético propuesto aplique al sector público y privado, es necesario considerar las realidades y particularidades de cada sector.

Con el aprendizaje profundo pasa igual, incluso peor, por ser algoritmos más flexibles son menos interpretables y más explicables⁶, al punto de ser llamados «cajas negras». Por naturaleza nos disgusta lo que no podemos explicar, por tal razón es normal que cuando se construya un modelo de este tipo, el público objetivo pida explicaciones de qué hace el algoritmo y cómo toma la decisión.

Lo anterior genera la necesidad de pasar de «cajas negras» a «cajas de cristal».

Para esta tarea se han creado métodos de interpretación agnóstica de modelos⁷, lo que permite dar explicaciones independientes de la técnica usada, porque el mismo método de interpretación se puede usar para cualquiera. Una alternativa es solo usar algoritmos interpretables, lo que a menudo tiene una gran desventaja, pues se pierde el rendimiento predictivo de los algoritmos más complejos.

Los métodos de interpretación agnóstica han permitido llegar a las llamadas «cajas de cristal» para evitar daños, perfilamientos o discriminaciones implícitas en los resultados de los modelos usados⁸. Algunos de los algoritmos más complejos de inteligencia artificial, basados en reconocimiento facial y procesamiento de lenguaje natural, han estado y permanecen en constante vigilancia, dados los hallazgos de discriminación en los resultados de esta tecnología, al punto de ser prohibido su uso o implementación.

Sonados casos como el de Amazon o la prohibición del uso del reconocimiento facial en ciudades de Estados Unidos⁹, ponen en evidencia el peligro de una mala interpretación de los datos, el contexto y el resultado de un modelo sofisticado.

Lo anterior no quiere decir que no es deseable el uso de esta tecnología en los seguros, lo que muestra son las consideraciones que hay que tener al poner en producción un algoritmo de aprendizaje automático.

Por ahora, en el sector asegurador es más fácil sofisticar modelos de ventas cruzadas, de retención de clientes, de cobro de cartera, predicción de costos

6. Interpretable vs Explainable Machine Learning, <https://towardsdatascience.com/interpretable-vs-explainable-machine-learning-1fa525e12f48>

7. Interpretable ML book; <https://christophm.github.io/interpretable-ml-book/>

8. <https://theconversation.com/como-evitar-que-la-inteligencia-artificial-aprenda-nuestros-prejuicios-142141>

9. <https://www.nytimes.com/2019/05/14/us/facial-recognition-ban-san-francisco.html>

de siniestros y un sinnúmero de prácticas que ya se realizan y que a la necesidad de mejores pronósticos se pueden volver más complejos con el nuevo conjunto de herramientas que ofrece el aprendizaje automático.

Para concluir, hay que tener un último ingrediente que suele atacarse y enseñarse al final, pero que debería ser el primero: la ética, para esto distintas organizaciones han venido creando y comunicando sus propios códigos de ética, muchos de ellos en organizaciones de actuarios o seguros; la Academia Americana de Actuarios (AAA)¹⁰, la Sociedad de Actuarios (SOA)¹¹, el Instituto Facultad de Actuarios (IFoA)¹², la Asociación Nacional de Comisionados de Seguros (NAIC)¹³ por nombrar algunos, sin desconocer el marco ético para la inteligencia artificial en Colombia¹⁴.

En general, se parte de un contexto y definiciones preliminares, mediante códigos de conducta y estándares, también se habla de la regulación que es escasa, se enuncian los principios éticos con los que se definió el documento de trabajo y se termina por hacer un llamado al profesionalismo de los integrantes de cada equipo.

De esta forma comenzar por el marco ético colombiano resulta en un buen ejercicio, por ser el más general y el que hace el siguiente llamado: «e insta a que «este marco sirva como un insumo general que permita después a cada entida» Dado que el Gobierno colombiano quiere que el marco ético propuesto aplique al sector público

➡ Las «cajas de cristal» evitan daños, perfilamientos o discriminaciones implícitas en los resultados de los modelos usados¹⁵.

y privado, es necesario considerar las realidades y particularidades de cada sector d dentro de alguno de los sectores desarrollar su propio marco.»

Los principios del marco ético colombiano son una mayor cantidad y son más amplios en su definición que los de otras organizaciones, donde los comunes son: responsabilidad, transparencia, predictibilidad, auditabilidad e incorruptibilidad y siempre se debe tener en cuenta que se implementan mediante la aplicación en paralelo a la ética de datos, ética de algoritmos y ética de las prácticas.

Como se vio, sesgo, varianza y ética son tres conceptos aparentemente simples, pero que para llevar a la práctica en la implementación de aprendizaje automático o profundo acarrea un permanente trabajo, que solo las compañías que lo comienzan a transitar darán cuenta de tal responsabilidad. 

10. Big data and the role of the actuary, <https://www.actuary.org/sites/default/files/files/publications/BigDataAndTheRoleOfTheActuary.pdf>

11. Ethical Use of Artificial Intelligence for Actuaries; SOA; <https://www.soa.org/globalassets/assets/files/resources/research-report/2019/ethics-ai.pdf>

12. Ethical and professional guidance on data sciences: A guide for members, IFoA; https://www.actuaries.org.uk/system/files/field/document/IFoA_Ethical_Professional_Guidance_Data_Science_Feb_2021.pdf

13. National Association of Insurance Commissioners (NAIC) Principles on Artificial Intelligence (AI); https://content.naic.org/sites/default/files/inline-files/AI%20principles%20as%20Adopted%20by%20the%20TF_0807.pdf

14. <https://dapre.presidencia.gov.co/AtencionCiudadana/DocumentosConsulta/consulta-marco-etico-IA-Colombia-200813.pdf>

15. <https://theconversation.com/como-evitar-que-la-inteligencia-artificial-aprenda-nuestros-prejuicios-142141>