Ing. Carlos Ormella Meyer y Asoc.
Gestión y Auditoría de Riesgos y Seguridad de la Información
Tel: +54-11-4371-4638 – Cel: +54-911-6513-2751
E-mail: ciprio@ingcomyasoc.com.ar

BAYES Y EL CASO DE LOS FALSOS POSITIVOS

En la Web sobre todo abundan ejemplos de cómo las probabilidades de datos estadísticos se modifican a partir de nuevas evidencias.

Un caso bastante común y llamativo se refiere a las enfermedades y sus síntomas, especialmente con respecto a los llamados “falsos positivos”, es decir donde la prueba o síntoma señala una enfermedad pero que en realidad no existe.

Estas publicaciones en muchos casos dicen que es un error común entre estudiantes y aún profesionales de la salud, tomar el porcentaje de personas enfermas con cierta enfermedad respecto de un total de personas como igual a la posibilidad de tener esa enfermedad entre quienes dieron positivo en una prueba al respecto. Efectivamente, la aplicación del **teorema o regla de Bayes** (un ministro presbiteriano inglés del siglo XVIII) señala que los resultados reales pueden variar significativamente e incluso en ocasiones ser mucho menores.

Demos un ejemplo con un enfoque para los no iniciados en Bayes (para los entendidos el planteo se complementa con el cuadro de la derecha).

1) Supongamos que en personas de ciertas características (por ejemplo mujer, edad, etc.) las posibilidades de tener un cierto tipo de cáncer es de 1 cada 10.000 personas, que todas las personas enfermas dan positivo en las pruebas pero que adicionalmente hay un 2% que también dan positivo sin estar enfermas.

2) A su vez, sea que tenemos datos de una muestra de por ejemplo 100.000 casos. Conforme la información anterior, de este total habrá 10 enfermos cuyas pruebas darán positivas, por lo que decimos que tenemos **10 positivos verdaderos**. Y, por lo tanto, no estarán enfermas las restantes personas, es decir, **99.990**.

3) De esa cantidad (99.990) habrá un 2% de **falsos positivos**, lo que totaliza prácticamente unas **2000** personas.

4) Por lo tanto el **total de positivos** encontrados será **2010** (10 positivos verdaderos más 2000 falsos positivos).

5) En consecuencia, la relación de positivos verdaderos (o sea 10) respecto del total de positivos encontrados (o sea 2010) será igual a 10/2010, Esto, que equivale a algo menos del 0,5%, implica que ***sólo tendrá la enfermedad en cuestión 1 de cada 201 personas cuyas pruebas dieron positivo!!!***

Bayes para los entendidos

Para el ejemplo dado la nomenclatura y la asignación serán:

P(A): Probabilidad *previa* de que una persona tenga cierta enfermedad.
P(B|A): Probabilidad condicional de que la prueba resulte positiva en una persona enferma.
P(B): Probabilidad total de pruebas positivas obtenidas en personas enfermas (verdaderos positivos) y no enfermas (falsos positivos).
P(A|B): Probabilidad *posterior* de que una persona esté enferma respecto del total de pruebas positivas obtenidas.

La regla de Bayes establece que:
:

$$P(A|B) = \frac{P(B|A) * P(A)}{P(B)}$$, sustituyendo será:

$$P(A|B) = \frac{1 * 0{,}0001}{1 * 0{,}0001 + 0{,}02 * 0{,}9999} = \frac{0{,}0001}{0{,}020098} =$$

$$= \frac{1}{200{,}98} = 0{,}497562\%$$

O sea que el 2% *previo* se reduce a menos del 0,5% como probabilidad *posterior* de que una persona esté realmente enferma habiéndole dado positiva la prueba.

Tel: +54-11-4371-4638 – Cel: +54-911-6513-2751
E-mail: ciprio@ingcomyasoc.com.ar

Se puede decir también que la probabilidad *previa* inicial del 2% se reajusta a la probabilidad *posterior* de menos del 0,5% que resulta de la relación de positivos verdaderos sobre el total de positivos (verdaderos y falsos),

Una observación sobre lo visto. Se puede comprobar fácilmente que la cantidad de la muestra disponible no modifica el resultado, así sean menos o más de los 100.000 casos considerados.

¿Por qué Bayes?
En muchos casos la primera vez que un especialista en seguridad de la información se encuentra con la regla de Bayes es cuando se busca justificar las inversiones en seguridad de una empresa para poder "vender" el proyecto en cuestión a un gerente de administración y finanzas. A personas en este tipo de cargos, en lugar de hacer consideraciones de hackers, virus, etc. se les puede hablar con ventajas de por ejemplo ***indicadores financieros*** (es decir un tema básicamente de su área), especialmente del ROI (Retorno Sobre la Inversión) y de ROSI como extensión para el caso de la seguridad de la información.

Pero para determinar los riesgos existentes muchas veces la empresa no tiene suficientes datos de incidentes registrados. Ahí es donde se puede aplicar Bayes porque justamente trabaja combinando esos datos cuantitativos con datos cualitativos surgidos de opiniones de expertos para producir, de esta manera, información *posterior* a partir de las estimaciones *previas* condicionadas a la evidencia de los datos históricos disponibles.

A partir de la regla de Bayes las aplicaciones se van ampliando. Hoy día por ejemplo es la base de las **redes bayesianas** de uso por ejemplo para determinar los *riesgos operacionales* especialmente en los bancos que deben cumplir con las normativas de Basilea II/III. Pero no es el único caso. También se aplica en la detección de correo spam, ingeniería de software, economía, educación, medicina, hidráulica, para diagnóstico y prevención de fallas de equipamiento, etc.