



Aportes de la Teoría de la Decisión Estadística a Situaciones de Riesgo en el Ámbito Universitario

Myriam Herrera, Susana Ruiz, Simón Gonzalez, Marita Romagnano, Magali Sosa, Benemérito Ignacio, Ma Luisa Gordillo

Departamento de Matemática Física y Química de la Fac. Ciencias Exactas, F. y Naturales -UNSJ
Instituto de Informática de la Fac. Ciencias Exactas, F. y Naturales -UNSJ

mherreram, maritaroma @ [info.unsj.edu.ar] , sbruiz@yahoo.com.ar

Resumen

En la actualidad, en todos los estratos sociales los datos, la información y el conocimiento se han convertido en uno de los recursos más valiosos para la toma de decisiones. En particular, considerando que, cualquiera sea la situación, siempre están presentes riesgos. Es decir, no solo se hace necesario atender a los datos, la información y las decisiones que sobre ellos se vayan a tomar, sino también los riesgos que estas decisiones puedan implicar.

La teoría de la decisión estadística juega un rol importante al momento de asistir situaciones relacionadas a diferentes ámbitos u escenarios sociales, por ejemplo, en la salud, educación, entre otros. Esta teoría propone un método para tomar decisiones, basado en principios básicos sobre la elección coherente entre opciones alternativas. La evaluación de riesgos, al considerar diversos escenarios, es un aspecto primordial.

Este proyecto propone brindar aportes en la toma de decisiones al aplicar técnicas, metodologías y herramientas estadísticas apropiadas en distintos escenarios del ámbito universitario, especialmente relacionados con alumnos y egresados.

¿Qué se hace?

En el proceso de toma de decisiones resulta imprescindible la disponibilidad de información relevante, particularmente cuando las consecuencias de dichas decisiones tienen el potencial de impactar.

Los riesgos son parte inevitable de nuestra vida cotidiana, y han aumentado de manera exponencial en las últimas décadas. Ante estas amenazas, se hacen dos preguntas: ¿cómo podemos evitar esos riesgos?, y en caso de no poder hacerlo, ¿cómo minimizar sus efectos (costos)? El análisis de riesgos se nutre de ciencias y/o disciplinas como la Probabilidad, la Estadística, la Optimización entre otras.

Lógicamente los datos proporcionan información relevante siempre que se utilicen técnicas y análisis apropiados frente a las problemáticas que se plantean y se quieran resolver.

Los datos se han convertido en un recurso crítico en muchas organizaciones. El acceso eficiente a ellos, compartirlos, extraer información de interés y hacer uso de la información se transforma en una urgente necesidad. Es de suma importancia la habilidad para extraer conocimiento útil escondido en esos datos y actuar sobre ese conocimiento. Como resultado hay muchos esfuerzos no sólo para integrar varias fuentes de datos dispersas a través de sitios diferentes, sino también es importante la información extraída de esas bases de datos en la forma de patrones y tendencias.

Independientemente de las distintas realidades que van surgiendo en cualquier ámbito social, hoy en día resulta imprescindible identificar características asociadas a las problemáticas, para así poder realizar predicciones futuras minimizando riesgos. Con los aportes metodológicos y herramientas que brinda la disciplina estadística, entre otros, se logrará establecer el riesgo al contribuir en la toma de decisiones en diferentes realidades.

¿Cómo se hace?

La estadística está basada en la teoría de probabilidades. Formalmente la probabilidad es una función de medida que cumple con ciertas condiciones (axiomas de probabilidad) que cuantifica una incertidumbre.

Las técnicas y metodologías de la estadística, tales como de la Estadística Bayesiana, dentro del Análisis Multivariado, se utilizan para realizar análisis adecuados a la información con la que se cuenta (análisis de datos) y decidir de manera razonable sobre la mejor forma de actuar.

Aunque la definición de función de probabilidad es una, existen varias interpretaciones de la probabilidad: clásica, frecuentista y subjetiva. La Metodología Bayesiana está basada en la interpretación subjetiva de la probabilidad y tiene como punto central el Teorema de Bayes.

La Decisión Bayesiana es un enfoque estadístico fundamental en clasificación de patrones. Se estudian probabilidades de tomar decisiones incorrectas para cuantificar los compromisos de esas decisiones y diseñar las estrategias de menor costo y riesgo.

En un principio utilizamos el modelo Logit, que consiste fundamentalmente en regresión lineal con una transformación logística, resulta útil en este caso debido a que es aplicable a conjuntos de datos multivariantes no normales con variables categóricas.

En general, al aplicar regresión logística, se busca obtener el modelo más simple que se ajuste correctamente a los datos, para una fácil interpretación. Existen distintos criterios de simplificación, siendo uno de los más utilizados el criterio de información de Akaike.

A modo de ejemplo de la metodología se utilizaron:

• Datos: encuesta SIU-Kolla al recién graduado

• Objetivo: determinar relaciones entre los observables y dos variables respuesta relacionadas con:

- el futuro laboral inmediato de los egresados
- el tiempo que les tomó obtener el título

Metodología: modelo Logit

$$L(\beta) = \log P(\beta) = \sum_{i=1}^n (y_i \log p_i + (1 - y_i) \log (1 - p_i))$$

Modelos implementados

MAIC: Modelo Logit con reducción por criterio de Akaike.

MD: Modelo Logit con reducción por criterio de Akaike y por criterio de la Deviance con heurística de eliminación hacia atrás.

MNN: Modelo de redes neuronales.

Variable Respuesta: Tiempo de cursado

$$M_D: \ln \frac{p}{1-p} = 0.244 - 0.130 x_{B2} - 0.626 x_{P1,1} + 1.767 x_{P1,2} + 19.616 x_{P2,1} - 0.018 x_{P2,2} + 5.142 x_{P4,1} + 3.415 x_{P4,2}$$

Variable Respuesta: Condición Actual

$$M_D: \ln \frac{p}{1-p} = -4.052 + 0.111 x_{B2} - 0.895 x_{B1,1} - 1.305 x_{B4,1} + 1.442 x_{B5,1} + 2.359 x_{B5,2} - 1.053 x_{H1,1} + 0.695 x_{H2,1} + 1.599 x_{H2,2} - 0.928 x_{E1,1} - 0.552 x_{E1,2} + 0.805 x_{E1,3} + 0.362 x_{E1,4} - 1.634 x_{P2,1} + 0.918 x_{P2,2} + 0.786 x_{C2,1}$$

La metodología empleada permite obtener modelos comprensibles y que pueden brindar valiosa información en el contexto del análisis de datos de egresados para la toma de decisiones institucionales.

Se implementó en R un método de reducción por eliminación hacia atrás basado en la desviación que será prontamente publicado en formato de librería de código abierto junto con funciones de evaluación de modelos.

Formación de Recursos Humanos

El equipo de investigación está formado por:

- Docentes investigadores de la FCFEN de la UNSJ, entre los que se cuentan una Doctora en Matemática, Magister en Informática, Magister en Estadística y Estudiante en Maestría en Inteligencia Artificial.

- Un alumno del Departamento de Informática.

Se espera sumar alumnos tesis de grado y posgrado (maestría y doctorado), interesados en esta línea de investigación.