

MODELO DIFUSO PARA LA DETERMINACIÓN DE UN ÍNDICE DE RIESGO DE ABANDONO EN LA ESCUELA POLITÉCNICA NACIONAL

Línea Temática: Factores asociados. Tipos y perfiles de abandono

Sandoval, Iván
Salazar, Diego
Sánchez, Ximena
Prócel, Galo
Guevara, Vanessa
Escuela Politécnica Nacional, Ecuador
e-mail: ximena.sanchez@epn.edu.ec

Resumen. Dentro de las metodologías usualmente utilizadas para calcular indicadores de rendimiento, abandono y aprobación estudiantil se suelen utilizar técnicas estadísticas basadas en regresiones y modelos lineales generalizados, todas éstas técnicas como es normal adolecen de un determinado nivel de error y limitaciones propias. En este sentido se ha efectuado un estudio alternativo utilizando técnicas de lógica difusa que pretenden cubrir de una manera más extensa las particularidades propias en la creación de los indicadores, se ha hecho énfasis en la creación del modelo así como en la explotación de las potencialidades de ésta técnica y se emplea información del último período académico en la Escuela Politécnica Nacional (EPN). Es así, que se realizó un modelo para el cálculo de un Índice de Riesgo de Abandono (IRA) con base en la información de ingreso al curso de nivelación de la EPN del último semestre académico, para verificar el uso eficiente de recursos del Estado, así como el aprovechamiento estudiantil para de esta manera contribuir en la reducción del Abandono en la Educación Superior. Para determinar el IRA se analizaron las posibles causas de abandono a partir de la base de datos socioeconómicos del departamento de bienestar estudiantil de la EPN, que permitió clasificar a los alumnos en situación de riesgo, cuyas notas se utilizaron en el cálculo del mencionado índice con la ayuda de herramientas matemáticas y estadísticas, acorde a las mejores prácticas internacionales. Este IRA refleja la probabilidad de que un determinado estudiante se encuentre en situación de riesgo de abandonar la Universidad, debido a alguno de los factores asociados al abandono y será de utilidad para la Universidad por cuanto permitirá determinar con bastante aproximación a los valores reales el porcentaje de alumnos que se podrían inscribir en los subsiguientes semestres, lo que redundará en una planificación más ágil y eficiente de los cursos, aulas, recursos y profesores necesarios, en beneficio del presupuesto asignado a las Universidades. El IRA, a futuro, permitirá predecir el porcentaje de estudiantes que no continuarían, cuyos resultados servirán como un instrumento en la toma de decisiones de las autoridades. Finalmente, es necesario resaltar que se trata de un enfoque puramente pedagógico acerca de las implicaciones socio-económicas y académicas del fenómeno del abandono, y se ha realizado poniendo énfasis en la obtención de un modelo matemático, de tal manera que se sugieren una mejora en cuanto a la metodología y al contraste estadístico clásico.

Descriptor o Palabras Clave Abandono, Lógica Difusa, Indicadores de Rendimiento, Estadística Inferencial, Escuela Politécnica Nacional

1. Introducción

El abandono estudiantil en los primeros años de Educación Superior ha constituido en la última década un fenómeno de alta relevancia en la universidad ecuatoriana, puesto que no solo influye en la calidad académica de las universidades e institutos tecnológicos y en su posterior ranking, sino que además establece una problemática de orden social y político.

Usualmente, el impacto directo que se observa es referente al tiempo promedio de finalización de los estudios de la población universitaria efectiva, es decir, de los estudiantes que pese a desertar o suspender en una o dos carreras finalmente se titulan. Además, los costos indirectos y directos que advienen para el estado, de mantener durante un determinado tiempo a un estudiante que es beneficiario de la educación superior puesto que en el caso ecuatoriano la educación superior es gratuita.

Hay varios factores explicativos a considerar en el estudio del fenómeno del abandono estudiantil, muchos de estos factores se recopilan de la experiencia metodológica, de estudios pedagógicos en las áreas pertinentes y de estudios generales de factores asociados, según (Cabrera L., Bethencourt J., Pérez P. y González M., 2006) y en (Escandell, O. y Marrero, G., 1999).

En este artículo se propone la creación de un Índice de Riesgo de Abandono (IRA) mediante la utilización de lógica difusa, la intención de esta metodología es recopilar en un modelo más robusto que el estadístico clásico en cuanto a la atención a los detalles pertinentes a la evaluación del Abandono Estudiantil como tal.

2. Metodologías de Lógica Difusa

En las metodologías tradicionales en las cuales un fenómeno está siendo modelado, es muy común asumir el valor de verdad, verdadero o falso para una determinada proposición, es decir el tipo de funciones matemáticas comúnmente utilizadas son del tipo $f_A(x) = \begin{cases} 1 & \text{si } x \in A \\ 0 & \text{si } x \notin A \end{cases}$ es decir, funciones que comprometen un único valor de verdad, mediante f para la pertenencia o no a un determinado conjunto A . Un ejemplo de esto lo constituye, en la variable estado civil (ec), si el estudiante es soltero $f_A(ec) = 1$ ó no lo es cuando $f_A(ec) = 0$, en donde el conjunto A es alguno en donde se encuentren variables descriptivas para el estudiante; en este caso concreto dicho estudiante no puede ser soltero y no soltero a la vez, es decir la *soltería* en principio, no es algo ambiguo sino está perfectamente determinado por una única condición, se ha casado o no.

Supongamos por el contrario un caso en el cual la variable descriptiva del estudiante no esté perfectamente determinada, por ejemplo su altura (al), si se intenta modelar dicha propiedad con la función $f_A(al) = \begin{cases} 1 & \text{si } al \geq 1.75 \\ 0 & \text{si } al < 1.75 \end{cases}$ estableciendo que 1.75[m] es la longitud requerida para considerar a alguien alto o no, nos queda una ambigüedad, puesto que alguien con una estatura de 1.745[m] o de 1.40[m] son considerados igualmente bajos, y así mismo hay igual diferenciación entre el estudiante con 1.745[m] y 1.76[m] que entre el estudiante de 1.50[m] y de 1.76[m]; podemos concluir entonces que dicha función de altura no discrimina de forma satisfactoria cuan alto es un estudiante. Precisamente, para salvar este tipo de ambigüedades es que en este artículo se ha optado por las técnicas de lógica difusa, para variables en las cuales haya más de un criterio para poder determinar la pertenencia o no de un determinado estudiante a un determinado clúster o conjunto de características académicas, sociales o económicas; veamos entonces, como la lógica difusa podría resolver este problema de la *altitud* de un estudiante.

Un modelo difuso, según lo indicado en indica (Sivanandam S., Sumathi S. y Deepa S., 2007) utiliza una función de distribución para la variable, es decir asigna un número en el intervalo [0; 1] para asignar el valor de verdad, teniendo distintos niveles de verdad, es decir distintos grados de altitud en la variable dependiente, de hecho para el ejemplo se ha utilizado una forma de ajuste

clásica dada por $f(al) = \frac{1}{1+e^{-8(al-1.6)}}$, en donde se nota claramente una forma suave y mejor distribuida de la curva que representa la altitud de un estudiante, a este tipo de funciones se las conoce como *funciones de pertenencia* o su equivalente *membership function*.

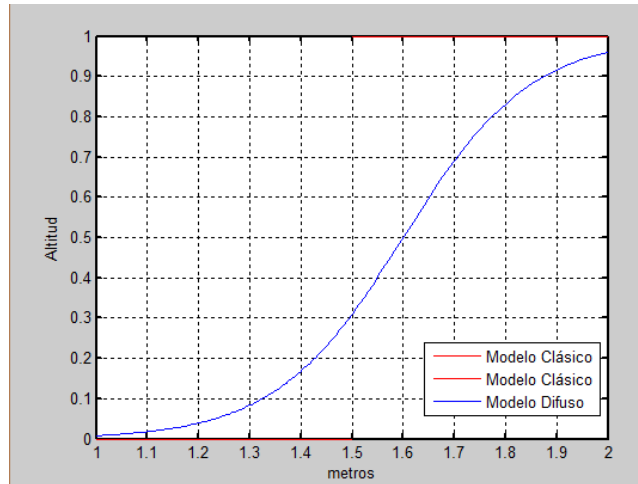


Figura 1. Modelo Clásico vs Modelo Difuso

De esta manera, y siguiendo lo indicado en (*Espinosa J., Vandewalle J. y Wertz V., 2005*) se puede difusificar algunas variables de carácter socio-económico o académico, es decir generar un conjunto de funciones miembro y finalmente combinar todas en un modelo difuso general, cuya puntuación podría ser difusa o se podrían escoger otras técnicas como el centroide, el centro de masa, el promedio ponderado o el método de densidades, según indica (*Sivanandam S. et al, 2007*).

En el contexto de nuestro estudio se pretende desarrollar un modelo para el estudio del Abandono Académico, desde una perspectiva multivariante, para este fin se propone la creación del **Índice de Riesgo de Abandono (IRA)**, el mismo que contendrá variables difusificadas y dicho IRA en sí mismo será una variable difusa. El IRA estará en una escala del [0; 1] y podrá tomar cualquier valor real en dicho intervalo, intenta cuantificar el nivel de riesgo de abandonar o desertar que posee un estudiante al momento del ingreso a la Escuela Politécnica Nacional, este índice fue diseñado no solamente acorde a criterios matemáticos, sino que además y según indica en (*L. Cabrera, J. Bethencourt, P. Pérez y M. González, 2006*) factores pedagógicos y didácticos fueron considerados, así como el conocimiento y la experiencia de los autores.

En este aspecto, podría imaginarse que se deja la evaluación de un índice de acuerdo a criterios subjetivos, pero debe entenderse precisamente lo que la lógica difusa intenta hacer es cuantificar dicha ambigüedad subjetiva y transformarla en una medición estándar y objetiva.

Selección de Variables Difusas

La selección de las variables se hizo en base al criterio de *difusificación posible*, es decir, si es factible modelar usando funciones miembro a aquellas variables sujetas a ser posibles de difusificar, por ejemplo, para las variables cuya pertenencia a un determinado conjunto sea ambigua o pueda estar en dos o más conjuntos simultáneamente. Así, para este modelo se han selecciona 4 variables: 2 académicas y 2 socioeconómicas. Las variables a continuación son:

Académicas

1. NotaENES

Esta variable indica la nota de ingreso de los estudiantes que acabaron la educación media y rinden el Examen Nacional para la Educación Superior (ENES), está en un rango de 0 a 1000.

2. PromedioColegio

El promedio de bachillerato es la nota final de los 6 años de educación media, está en un rango de 1 a 20.

Socioeconómicas

1. Ayuda\$

Es una pregunta tomada de una encuesta de contexto socioeconómica realizada a todos los bachilleres, indica cual es el promedio de dinero (USD) mensual que percibe cada uno de los estudiantes de la EPN. Se ha encontrado que el rango está entre \$0 y \$370.

2. ViolenciaColegio

Es una variable que indica, mediante una encuesta, como cada estudiante considera que se ha vivido en su colegio la problemática de la violencia, concretamente la pregunta es:

Evalúe conforme a su criterio el ambiente sin violencia (la violencia puede incluir violencia física, verbal, psicológica, bullying, racismo, discriminación, etc.) entre estudiantes y docentes.

Los valores para cada una de estas categorías están dados acorde a las siguientes tablas:

Tabla 1. NotaENES

Categoría	Rango
Baja	0 – 600
Media	500 – 900
Alta	800 – 1 000

Tabla 2. PromedioBachillerato

Categoría	Rango
Regular	0 – 12,9
Bueno	13 – 15,9
Muy Bueno	16 – 18,9
Sobresaliente	19 - 20

Tabla 3. Ayuda\$

Categoría	Rango
1	Mínimo 0
2	Máximo 350

Tabla 4. ViolenciaColegio

Categoría	Rango
1	Totalmente Acuerdo 90 – 100
2	Acuerdo 70 – 89
3	Medianamente Acuerdo 50 – 69
4	Desacuerdo 30 – 49
5	Totalmente Desacuerdo 1 - 29

2.1 Modelo Clásico vs Modelo Difuso

En general, los modelos de regresión multivariantes constan de un plano o hiperplano de ajuste el cual minimiza el cuadrado de la distancia entre la nube de puntos y el hiperplano interpolante, para obtener una estimación a un cierto nivel de efectividad, en medio de la dispersión de datos, como se muestra a continuación en un ejemplo ilustrativo.

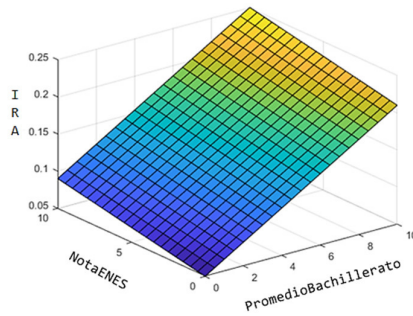


Figura 2. Modelo Multivariante Clásico

Como antes se había indicado, los modelos difusos en donde intervienen varias variables, al combinar tanto modelos lineales como no lineales, se permite obtener una superficie que ajusta o estima de mejor manera, dicha dispersión de datos, como el ejemplo de la superficie a continuación:

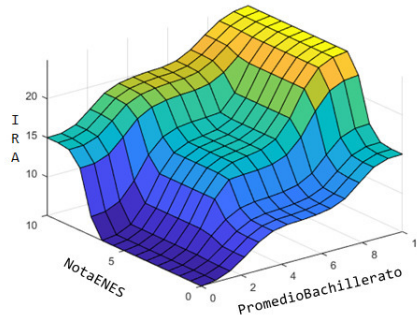


Figura 3. Modelo Difuso

2.2 Modelo Difuso General

El modelo consta de tres secciones, el ingreso de las variables difusas, su procesamiento mediante alguna o varias técnicas, hemos elegido *mamdani*, que utiliza reglas de composición del máximo y mínimo de cada una de las correspondientes funciones, y finalmente el cálculo del IRA, el cual contiene en sí mismo funciones difusas.

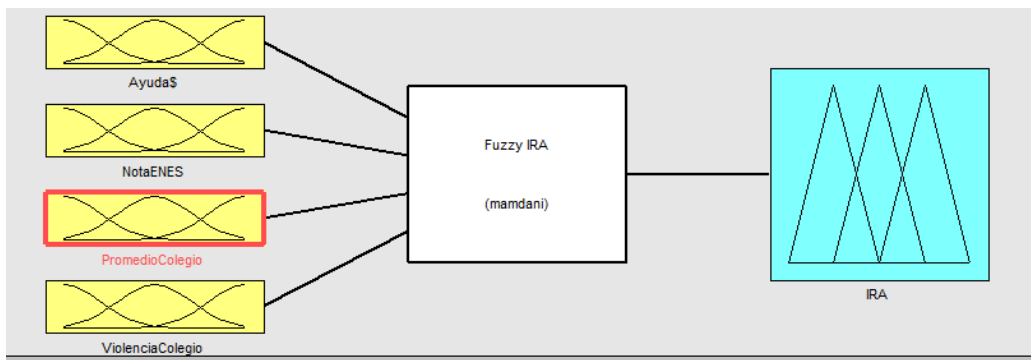


Figura 4. Esquema Difuso

2.3 Funciones Miembro

Se describen a continuación, cada una de las funciones miembro que componen el modelo difuso.

Función Ayuda\$

Esta función reparte mediante un tipo logístico la ayuda económica con la cual se ve favorecido cada estudiante en la EPN, su carácter creciente indica que a medida que un estudiante posee más ingresos económicos es menos probable que abandone sus estudios, llegando a un máximo de 350USD, que es un equivalente a un salario básico en el Ecuador.

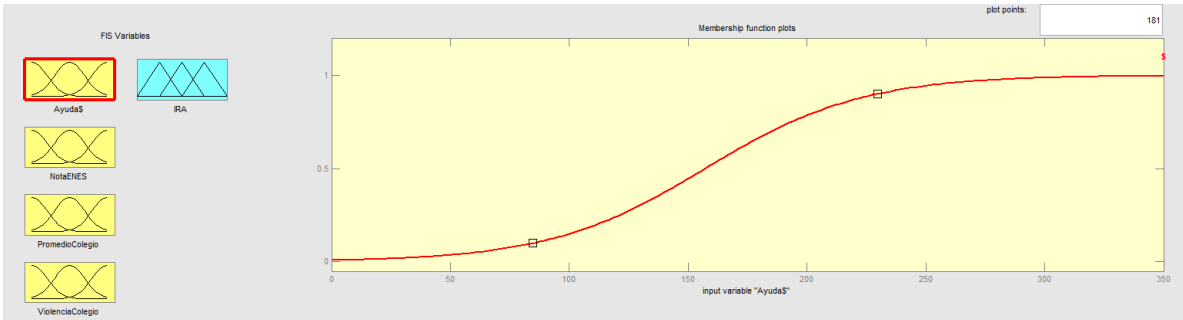


Figura 5. Función Ayuda\$

Función NotaENES

Esta función de pertenencia indica en que parámetros se encuentra la calificación de ingreso a la Universidad, es decir la calificación del Examen Nacional para la Educación Superior (ENES), se han creado tres subfunciones miembro clasificadas en Baja, Media y Alta, acorde a los intervalos indicados en la [Tabla 1](#), se ve entonces que una nota correspondiente al 85% es menos alta que media, pero es las dos en realidad, así mismo una nota cercana al 90% es mucho menos Media y es más Alta, una nota que supera el 95% es totalmente Alta y no da criterio de ambigüedad.

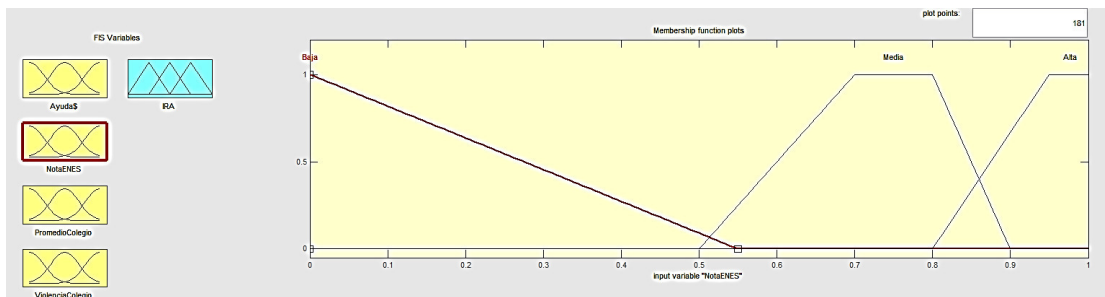


Figura 6. Función NotaENES

Función PromedioColegio

Esta función indica la difusificación de la nota de toda el recorrido académico durante la educación secundaria, clasificada según la [Tabla 2](#), se ha considerado este como una de las variables más relevantes, por lo que se realizó una segmentación más fina y el nivel de intersección entre las subfunciones miembro es mucho mayor, esto pretende no dejar escapar ninguna variabilidad y afinar dicha función miembro.

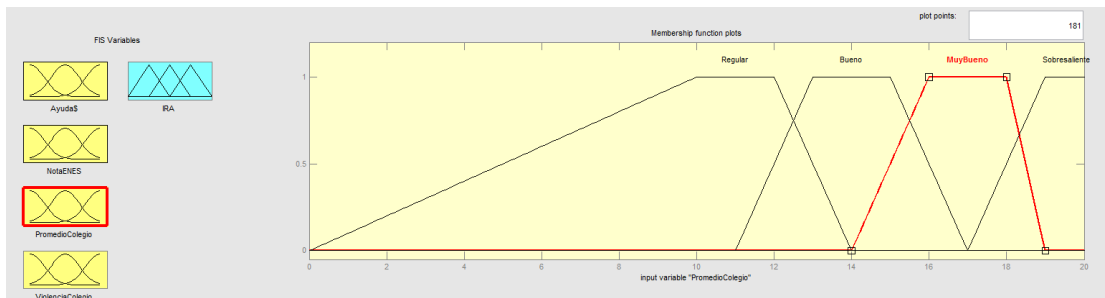


Figura 7. Función PromedioColegio

Función ViolenciaColegio

Esta función evalúa el criterio subjetivo del estudiante de la EPN, lo que intenta hacer es recoger la información de una variable netamente social como lo es la violencia en las aulas, concretamente la violencia entre estudiantes y/o con el profesor. De igual manera al ser una variable general, se ha difusificado poniendo atención en el refinamiento de los correspondientes dominios, los cuales presentan intersecciones que indican una mayor participación entre categorías y subfunciones miembro.

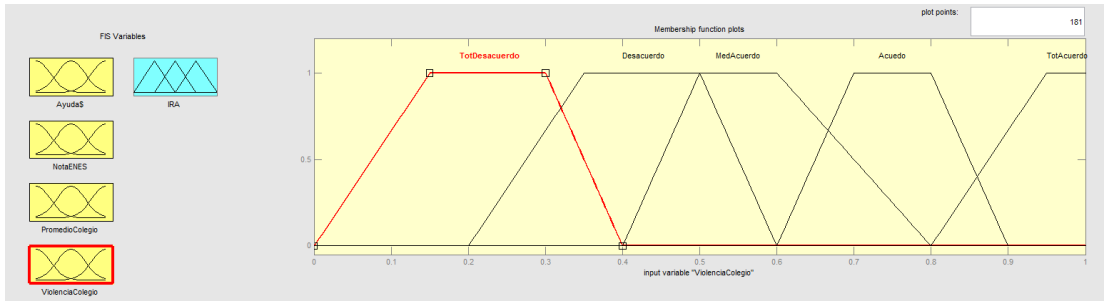


Figura 8. Función ViolenciaColegio

Función IRA

Finalmente tenemos la función de pertenencia del índice IRA, se ha optado por funciones semigaussianas por intervalos, debido a su suavidad, sin picos y con fuerte enlaceamiento en los dominios, debido a la intención de recoger de la mejor forma la variabilidad de las funciones miembro anteriores.

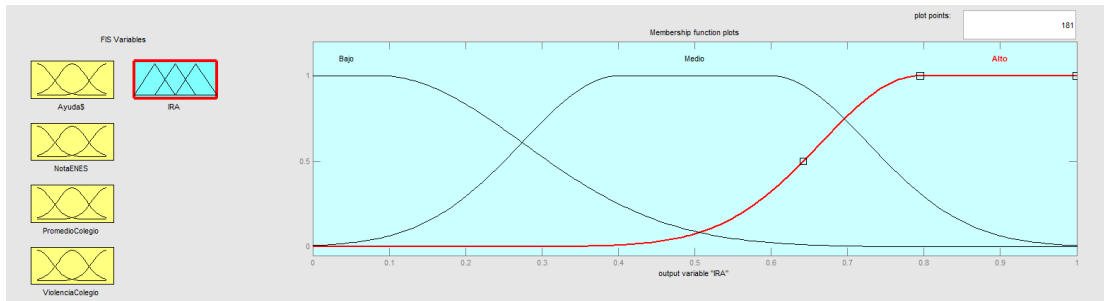


Figura 9. Función IRA

3. RESULTADOS

Los resultados enunciados a continuación son las superficies que resultan de la interacción del IRA vs las 4 variables difusas, las superficies en sí mismas constituyen el resultado del modelo y se describen a continuación:

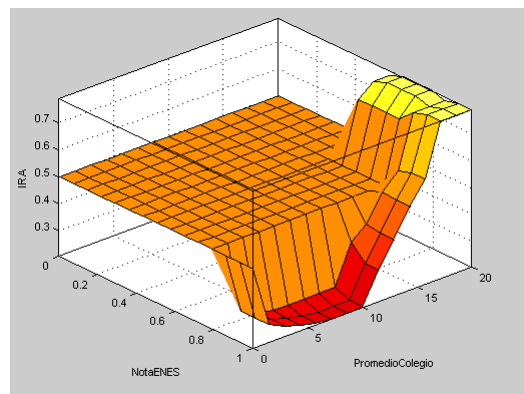


Figura 10. IRA vs NotaENES & PromedioColegio

En esta superficie se puede notar la influencia tanto de la nota Promedio del Colegio y la nota de acceso a la universidad, se observa una relación más fuerte entre la variable PromedioColegio que con la NotaENES, las reglas lógicas fueron consideradas de ésta manera porque la nota de colegio representa un histórico de 6 años, mientras que la nota de ingreso es en un solo momento de evaluación, acorde a los criterios indicados en (Escandell, O. y Marrero, G., 1999).

Se considera un factor más relevante dado que durante 6 años se han generado hábitos de estudio que son notoriamente más fuertes que un solo examen.

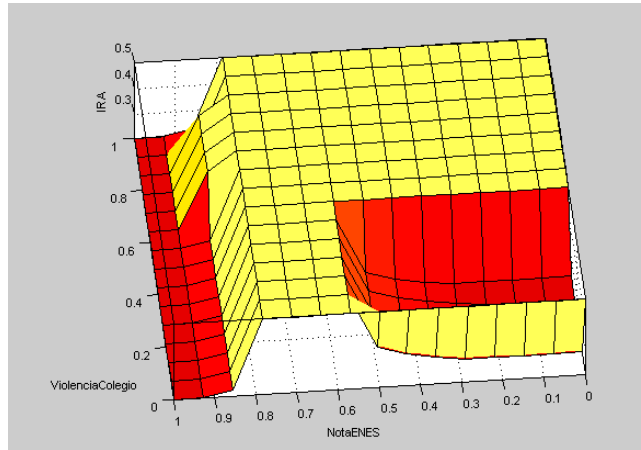


Figura 11. IRA vs NotaENES & ViolenciaColegio

En esta superficie se observa el carácter descendente de del IRA a medida que se obtienen mejores notas en el ENES, de igual manera se nota un aumento en el IRA a medida que los factores de violencia aumentan. Según se indica en (L. Cabrera, J.Bethencourt, P. Pérez, M. Afonso).

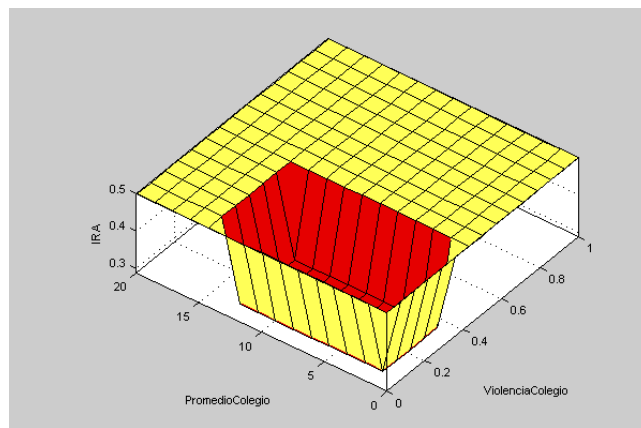


Figura 12. IRA vs PromedioColegio & ViolenciaColegio

En esta superficie se nota una cuenca en la cual el IRA disminuye cuando se obtiene un promedio alto, así existan índices de violencia, esto se debe a que el desempeño del colegio, considerado en la programación del modelo, es una variable de mucha más fuerza de correlación que la violencia experimentada en el colegio.

CONCLUSIONES

- Se construyó un modelo difuso, las reglas lógicas involucradas, y posterior a esto, la identificación de las variables más relevantes que están involucradas en la medición del IRA, este último aspecto es particularmente relevante puesto que nuestras bases de datos, con corte a Agosto 2016, cuentan con alrededor de 40 variables socio-económicas, demográficas y de segmento académico, la creación de las funciones miembro y el modelo como tal, permitieron identificar las más relevantes.
- Con la información obtenida en la EPN se desarrolló un modelo difuso que permite calcular el IRA. Con una fuente de datos lo suficientemente actualizada y depurada, y tras un desarrollo posterior, será posible el cálculo de dicho IRA, su comparación y posterior evolución.
- La complejidad de las funciones miembro utilizadas (lineales, constantes y semigaussianas) fue suficiente para captar el comportamiento de las variables que intervienen en el modelo.
- Se logró identificar en la base de datos utilizada, las variables académicas y socio-económicas sujetas a ser difusificadas.
- El modelo es lo suficientemente dinámico como para modificarlo acorde a las necesidades de las variables de entrada.
- Las reglas lógicas del modelo identifican a la variable *PromedioColegio* como la variable transversal al momento del cálculo del IRA.

REFERENCIAS

- [1] Ley Orgánica de Educación Superior (LOES) (2010). Registro Oficial Suplemento 29 de 12 de octubre del 2010.
- [2] Resolución 104-CEAACES-SO-12-2014 CEAACES (2014). Reglamento de Evaluación y Acreditación de Carreras. Consejo Nacional de Evaluación y Acreditación de la Educación Superior del Ecuador, <http://www.ceaaces.gob.ec/sitio/reglamentos/>
- [3] Resolución 132-CEAACES-SO-10-2015(2015). Reforma al Reglamento de Evaluación de Carreras, <http://www.ceaaces.gob.ec/sitio/>
- [4] Cabrera, L., Tomás, J., Álvarez, P. y Gonzalez, M. (2006). El problema del abandono de los estudios universitarios. RELIEVE, v. 12, n. 2, p. 171-203. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_1.htm
- [5] S. N. Sivanandam, S. Sumathi and S. N. Deepa. (2007). Introduction to Fuzzy Logic using MAMATLAB, Springer Verlag, 445, 157-277
- [6] Escandell, O. y Marrero, G. (1999). El abandono de los estudiantes en la universidad de Las Palmas de Gran Canaria: sus causas, propuestas de estrategias para su disminución. Las Palmas de Gran Canaria: La Caja de Canarias.
- [7] Cabrera L., Bethencourt J., Pérez P., Afonso M. (2006). El problema del abandono de los estudios universitario, RELIEVE, v. 12, n. 2. http://www.uv.es/RELIEVE/v12n2/RELIEVEv12n2_1.htm
- [8] Gomes, K. y Felicetti, V. (2015) Indicadores De Permanência: Contribuições Das Conferências Latino-Americanas Sobre O Abandono Na Educação Superior – Clabes. Quinta Conferencia Latinoamericana Sobre El Abandono En La Educación Superior. UNILASALLE. Brasil.
- [9] Del Corral J., González-Quejigo J., Villasalero, M. Análisis Del Abandono Universitario En La Uni-versidad De Castilla-La Mancha: Resultados Del Proyecto Alfa Guía. Quinta Conferencia Latinoamericana Sobre El Abandono En La Educación Superior. Universidad de Castilla. España