

Aplicación del aprendizaje automático con árboles de decisión al estudio de las variables del modelo de indicadores de gestión de las universidades públicas

Automatic with learning application tress to study decision variables model
management indicators public universities

Jhon Jairo Santa Chávez, Juan De Jesús Veloza Mora, Reinel Arias Montoya
^{1,3} *Facultad de Ingenierías, Universidad del Libre, Pereira, Colombia*
jjsanta@unilibrepereira.edu.co
juanveloza@unilibrepereira.edu.co
ram@etp.com.co

Resumen— El estudio de las variables del modelo de indicadores de gestión del Sistema de Universidades Públicas (SUE) desde el punto de vista institucional, se convierte en un estudio de carácter prioritario, ya que actualmente el Ministerio de Educación asigna los recursos a distribuirse entre las universidades públicas con base en estos indicadores. Las universidades colombianas no cuentan con un modelo que permita realizar estudios previos de dichos indicadores, reduciendo la posibilidad de mejorar los aspectos necesarios para tener acceso a mayores recursos que permitan su funcionamiento óptimo. En este estudio se analizan las variables que inciden dentro del modelo de indicadores de gestión y se propone una metodología general basada en árboles de decisión, específicamente el método de algoritmo que ayudará a elaborar un modelo matemático para determinar los ajustes necesarios que debe hacer la Universidad Tecnológica de Pereira, con el fin de mejorar sus índices de gestión.

Palabras clave— Modelo matemático, sistema de gestión de universidades públicas, arboles de decisión, minería de datos, algoritmo inductivo, patrones, modelos

Abstract— The study of model variables of performance indicators Public University System (SUE) from the institutional point of view, it becomes a priority study, and that the Ministry of Education allocates resources to be distributed among the universities public based on these indicators. Colombian universities do not have a model to previous studies of these indicators, reducing the possibility of improving the aspects necessary to have access to greater resources to its optimum performance.

In this study we analyze the variables that influence within the model of management indicators and proposes a general methodology based on decision trees, specifically the algorithm method which will help develop a mathematical model to determine the necessary adjustments should the Universidad Tecnológica de Pereira, in order to improve their management indexes.

Key Word — Mathematical model, management system of public universities, decision trees, data mining, inductive algorithm, patterns, models

I. INTRODUCCIÓN

De acuerdo con la Ley 30 de 1992, las fuentes de recursos de las universidades estatales son tres: los provenientes de la Nación, los aportados por entidades territoriales y los propios de cada institución, los cuales se obtienen a través de venta de servicios, matrículas, derechos académicos, programas de extensión, consultorías o convenios de investigación, entre otros.. En cuanto a los recursos provenientes de la Nación, estos se aseguran y regulan mediante dos normatividades: los artículos 86 y 87 de la Ley 30. El artículo 86 garantiza a las universidades del Estado recursos fijos que provienen de la nación y de las entidades territoriales, los cuales deben ser destinados a funcionamiento e inversión; el artículo 87 por su parte fija el incremento anual de los recursos que aporta la nación por el artículo 86 de acuerdo con el crecimiento de la economía y de conformidad con los objetivos previstos para el sistema de universidades estatales. Los aportes de la nación provenientes del artículo 86, que constituyen en muchos casos

en la principal fuente de financiamiento de las universidades, son distribuidos mediante una ecuación que garantiza su incremento en pesos constantes, tomando como base los presupuestos de 1993. A pesar de lo anterior, dicha ecuación no tiene en cuenta criterios de gestión de las universidades en términos de la eficiencia en el uso de los recursos o del cumplimiento de las metas institucionales y de la política sectorial. El gobierno nacional en el anterior Plan de Desarrollo (Ley 812 de 2003) buscó una modificación al esquema de distribución de estos recursos a partir de la aplicación de un modelo de indicadores de gestión para la asignación de los mismos, el cual fue desarrollado por el Sistema de Universidades Estatales (SUE) y permite fijar criterios generales para identificar el nivel de eficiencia de las instituciones de educación superior en el manejo y productividad de sus recursos. A pesar de lo anterior, el artículo 84 de la Ley 812 de 2003 fue declarado inexecutable y por tanto la distribución de los recursos de funcionamiento e inversión de las universidades estatales referidos al artículo 86 de la Ley 30 sigue teniendo su tendencia inercial. No obstante, la distribución de los recursos correspondientes al artículo 87 de la Ley 30 de 1992, continúa aplicando el modelo de indicadores de gestión diseñado por el sistema de universidades estatales, el cual se ha ido perfeccionando con la participación activa de todas las instituciones que lo componen. Se trata de un instrumento que combina la estadística y la matemática en un modelo técnico que provee unos indicadores claros, los cuales permiten a la sociedad enterarse de cómo las universidades estatales producen resultados concretos.

El problema fundamental es que las universidades desconocen la forma en como el Ministerio de Educación pondera estos índices de gestión para realizar la categorización y asignar en base a esta los recursos que provienen del estado.

El presente trabajo pretende construir un modelo que comprende los árboles de decisión y el análisis de datos, con la metodología de selección automática de atributos importantes, para realizar luego la toma de decisiones en función al modelo de indicadores que permita mejorar el funcionamiento de Universidad Tecnológica de Pereira.

II. CONTENIDO

A. Metodología

1. La información

En el marco del cumplimiento de sus objetivos misionales, la gestión de las universidades estatales está direccionada por dos grandes elementos: por el marco nacional de la política sectorial, contemplado en el Plan Nacional de Desarrollo y por las competencias legales de las Instituciones de Educación Superior definidas en la Ley Orgánica de Educación Superior (Ley 30 de 1992).

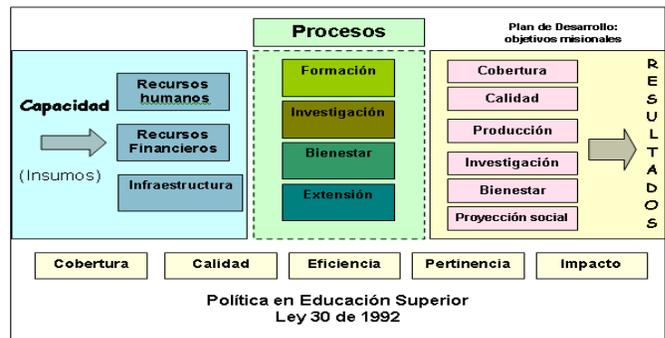


Figura 1. Base conceptual del modelo de indicadores de gestión.

Considerado el insumo primario de información a ser sometida al modelo planteado.

2. Las herramientas de minería de datos.

i. El proceso KDD

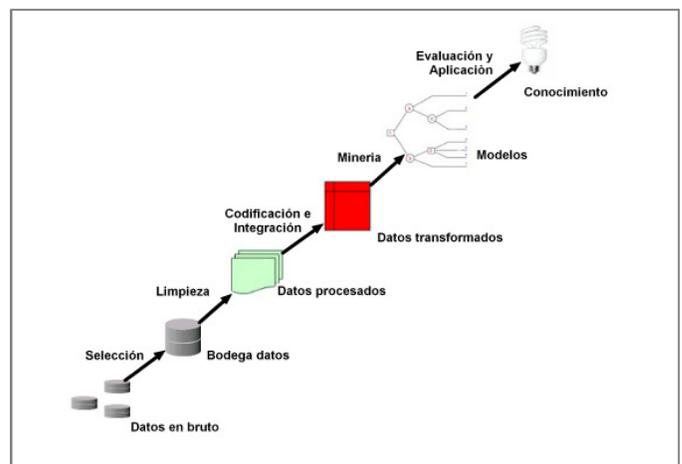


Figura 2. El proceso KDD .

El descubrimiento de conocimientos en las bases de datos, por sus siglas en inglés (knowledge discovery in databases o

KDD), combina métodos de la teoría de bases de datos, de la estadística, del reconocimiento de patrones, inteligencia artificial, computación de alto desempeño y otros, describiendo el proceso de encontrar estructuras interesantes y útiles en los datos. Por estructura se entiende patrones o modelos. Un patrón se define clásicamente como una descripción incluyente de un conjunto de datos. El proceso KDD involucra varios pasos tales como la selección, limpieza y transformación de los datos. Existe un paso crítico en este proceso que es la minería, es decir, la búsqueda de patrones en esos datos. Para un conjunto arbitrario de datos, existen más patrones posibles que datos en sí, el objetivo de la minería de datos es saber que patrones buscar y enumerar. Dado que la mayoría de patrones serian completamente superfluos y quizás hasta serian producto de la aleatoriedad o a veces describirían sentencias completamente obvias, es por este motivo que se hace atractiva la minería de datos al dedicarse a la búsqueda de patrones interesantes y útiles en los mismos. El proceso a recorrer desde los datos en bruto hasta encontrar "conocimiento" en los mismos, es un proceso largo y arduo. Los datos en bruto generados por cualquier proceso por ejemplo uno de manufactura, compras de los clientes o una central telefónica son difíciles de obtener. El primer paso es recolectar y organizar estos datos en una bodega que suministre una visión lógica unificada de diversos aspectos del negocio o la organización. Al construir una bodega de datos, se deben afrontar problemas básicos como la limpieza y consolidación de los mismos tales como creación de nuevos datos agregados o discretización y codificación de los mismos, para prepararlos como insumo de la herramienta de minería que pretende la obtención del conocimiento.

ii. Los arboles de decisión.

De las herramientas disponibles en minería de datos tenemos las de descubrimiento del conocimiento, donde se encuentran los Algoritmos de árboles de decisión como una de las principales herramientas predictivas que ayudan a determinar los atributos de mayor incidencia en una decisión positiva o negativa entregándolos en un árbol y mostrando las reglas que guían al destinatario sobre cómo obtener esta decisión. Es un algoritmo TDIDT cuya heurística principal es buscar el mejor atributo para ubicarlo en la raíz del árbol y para esto utiliza un estadístico llamado mayor ganancia de información el cual esta expresado como la diferencia de las siguientes entropías según el teorema de SHANON..

La entropía de conjunto dada por la siguiente expresión:

$$H(S) = -p_p \log_2 p_p - p_n \log_2 p_n$$

La entropía ponderada definida como el valor esperado de la entropía del conjunto cuando se particiona de acuerdo a un atributo A en particular. Esta entropía se calcula de acuerdo a la siguiente expresión:

$$H(S, A) \equiv \sum_{v \in \text{Valores}(A)} \frac{|S_v|}{|S|} H(S_v)$$

Con las dos expresiones anteriores se puede ahora obtener el estadístico de ganancia de información para cada atributo A de la información mediante la siguiente expresión:

$$G(S, A) \equiv H(S) - H(S, A)$$

Es utilizando el mayor valor de esta ganancia que se selecciona el atributo que debe ir ubicado en la raíz del árbol.

Como segundo elemento de la Heurística básica tenemos la redefinición del conjunto base para cada rama del árbol que corresponde a cada valor del atributo seleccionado en el paso anterior.

Este proceso debe repetirse recursivamente para cada nuevo nodo generado en el árbol. Cuando un nodo no tiene descendientes distintos entonces se convierte en una rama que contiene la decisión a tomar.

A continuación se muestra, haciendo uso del ejemplo típico en la teoría de este algoritmo sobre la decisión de ir o no a jugar tenis condicionado al valor de otras variables:

Información Base:

Estado	Temp	Humedad	Viento	Juego tenis
Soleado	Alta	Alta	Leve	No
Soleado	Alta	Alta	Fuerte	No
Nublado	Alta	Alta	Leve	Si
Lluvia	Suave	Alta	Leve	Si
Lluvia	Baja	Normal	Leve	Si
Lluvia	Baja	Normal	Fuerte	No
Nublado	Baja	Normal	Fuerte	Si
Soleado	Suave	Alta	Leve	No
Soleado	Baja	Normal	Leve	Si
Lluvia	Suave	Normal	Leve	Si
Soleado	Suave	Normal	Fuerte	Si
Nublado	Suave	Alta	Fuerte	Si
Nublado	Alta	Normal	Leve	Si
Lluvia	Suave	Alta	Fuerte	No

El árbol resultante después de aplicar esta Heurística.

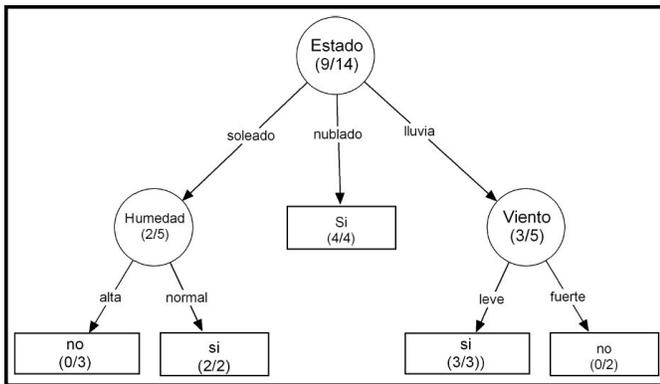


Figura 3. Árbol de decisión para el juego de tenis.

Y las reglas producidas por el algoritmo:

La persona juega tenis si:

Estado=Soleado y Humedad=normal o

Estado=Nublado o

Estado=lluvia y Viento=Leve

En la etapa de evaluación y aplicación se observa la facilidad de entendimiento, por parte del usuario destinatario, tanto del árbol como de las reglas de decisión lo cual es tomado como referente para seleccionar el patrón hallado.

iii. El modelo planteado

El modelo planteado incluye los pasos explicados en el proceso KDD desarrollados desde la selección hasta la codificación pasando por la limpieza y entregándolos a la herramienta de minería de árboles de decisión utilizando scripts programados en PHP y soportados por el SQL, los cuales en su conjunto conforman el modelo que permite la obtención del conocimiento mediante la exploración en la base de datos haciendo uso de herramientas de la minería de datos.

III. RESULTADOS.

Resultados para un problema de clasificación de presencia o ausencia de enfermedades del corazón.

Inicialmente se muestra la aplicabilidad del algoritmo de clasificación en el campo de la instrumentación física. La Universidad de California en Irvine publica desde 1987 una amplia colección de problemas en estas y otras áreas, la mayoría de ellos con datos reales aportados por diversos investigadores e instituciones y que son ampliamente referenciados en publicaciones científicas, llamada el *Machine Learning Repository* y que actualmente alberga cerca de 200 conjuntos de datos para la investigación y prueba de métodos y algoritmos de aprendizaje automático. En el análisis de la eficiencia del algoritmo utilizado en este trabajo, se hace uso de uno de los conjuntos de datos publicados en este repositorio aplicable a la instrumentación física. En el ejercicio se toma una base de datos 270 instancias o registros que

contienen información variada entre instrumentos de medida y datos ingresados los cuales sirven como insumo de prueba del algoritmo de clasificación utilizado para observar su funcionalidad. Esta base de datos contiene 13 atributos (que se han extraído de un conjunto mayor de 75.

Estos datos fueron extraídos de:

<http://archive.ics.uci.edu/ml/datasets/Statlog+%28Heart%29>

Se realizaron diferentes ensayos para observar la funcionalidad del algoritmo de clasificación diseñado, probando con parte de los datos como aprendizaje y seguidamente observando el nivel de aciertos cuando lo usamos para predecir el resultado con el restante grupo de datos.

A continuación se muestra un esquema de presentación de resultados de un ensayo:

Total casos: 270
Casos acertados: 247
Porcentaje error: 8.52%
t_ejec=10s

Ejecucion con nivel_max=5

Clases

clase[0]=2

clase[1]=1

Positivo:2 Arbol resultante basado en criterio de proporción

```

[] (-/-) thal_c
    [<= 4.5] (33/152) dipe_c
        [<= 2.5] (26/142) fcma_c
            [<= 112.0] (6/7) 2
            [> 112.0] (20/135) 1
        [> 2.5] (7/10) edad_c
            [> 45.5] (7/8) 2
            [<= 45.5] (0/2) 1
    [> 4.5] (87/118) paer_c
        [> 109.0] (86/112) tddt
            [3] (12/21) 2
            [2] (5/7) 2
            [4] (67/76) 2
            [1] (2/8) 1
        [<= 109.0] (1/6) 1
  
```

Reglas de decisión generadas

Regla 1:

```

Si thal_c es <= 4.5
y dipe_c es <= 2.5
y fcma_c es <= 112.0
entonces fail -> 2
  
```

Regla 2:

Si $thal_c$ es ≤ 4.5
 y $dipe_c$ es > 2.5
 y $edad_c$ es > 45.5
 entonces fail -> 2

Regla 3:

Si $thal_c$ es > 4.5
 y $paer_c$ es > 109.0
 y $tddt$ es 3
 entonces fail -> 2

Regla 4:

Si $thal_c$ es > 4.5
 y $paer_c$ es > 109.0
 y $tddt$ es 2
 entonces fail -> 2

Regla 5:

Si $thal_c$ es > 4.5
 y $paer_c$ es > 109.0
 y $tddt$ es 4
 entonces fail -> 2

Resultados para el problema de clasificación usado con los indicadores de gestión de las universidades públicas para determinar ubicación en un ranking predefinido.

En este caso se utilizó el algoritmo para predecir la ubicación dentro de un ranking determinado en cada ensayo dependiendo del valor de los indicadores de gestión entrados como insumo. Se realizaron diferentes ensayos haciendo uso de indicadores de gestión de 32 universidades públicas, en los últimos 10 años, como datos de entrenamiento para el algoritmo, y con estos mismos datos se probó la asertividad del modelo prediciendo ante diferentes valores de indicadores de prueba para la Universidad Tecnológica de Pereira.

En este enlace se explica todo sobre los datos usados por el algoritmo:

<http://www.mineducacion.gov.co/sistemasdeinformacion/1735/article-212353.html>

La siguiente es la codificación adoptada para el proceso:

PTCE: profesores tiempo completo equivalente
 GEPA: gasto en personal administrativo
 RFIN: recursos financieros
 RFIS: recursos físicos
 PPRE: programas de pregrado
 PPOS: programas de postgrado
 MPRE: matrícula pregrado
 MPOS: matrícula postgrado
 MPPA: matrícula pregrado en programas acreditados
 MPCu: matrícula primer curso
 GPRE: graduados pregrado
 GPOS: graduados posgrado
 GRMD: graduados en maestría y doctorado
 SPRI: Saber pro Inglés
 GRML: graduados en el mercado laboral.
 MVEC: movilidad de estudiantes colombianos.

MVEE: movilidad de estudiantes extranjeros.
 SPQS: Saber pro Quintil Superior
 REVI: revistas indexadas
 ARTI: artículos en revistas indexadas
 PATT: patentes
 MVDC: movilidad docentes
 PRCU: producción cultural
 ESEX: estudiantes en extensión
 ENVE: entidades con vínculo de extensión
 PRLI: productos licenciados
 EDCN: educación continua
 APEPRE: apoyo socioeconómicos a estudiantes de pregrado
 APEPOS: apoyo socioeconómicos a estudiantes de posgrado
 NERE: número de estudiantes retenidos.
 CLSI: clasificación.

A continuación se muestra un esquema de presentación de resultados de un ensayo:

Prueba contra los datos de entrenamiento

Total casos: 279
 Casos acertados: 248
 Porcentaje error: 11.11%
 Problema con nivel 3 y $P=5$

Clases

- clase[0]=S
 - clase[1]=P
- Positivo:P

Árbol resultante basado en criterio de proporción

```
[ ] (-/-) MVDC_C
  [ <= 83.5 ] (32/245) GPRE_C
    [ > 8.0 ] (28/240) S
    [ <= 8.0 ] (4/5) APEPRE_C
      [ <= 32.0 ] (0/1) S
      [ > 32.0 ] (4/4) P
  [ > 83.5 ] (18/34) MPRE_C
    [ <= 40173.5 ] (18/28) APEPRE_C
      [ <= 17543.5 ] (18/26) P
      [ > 17543.5 ] (0/2) S
    [ > 40173.5 ] (0/6) S
```

Reglas de decisión generadas

Regla 1:

Si $MVDC_C$ es ≤ 83.5
 y $GPRE_C$ es ≤ 8.0
 y $APEPRE_C$ es > 32.0
 entonces ranking -> P

Regla 2:

Si $MVDC_C$ es > 83.5
 y $MPRE_C$ es ≤ 40173.5
 y $APEPRE_C$ es ≤ 17543.5
 entonces ranking -> P

Función clasificadora generada

```

funcionclasifica($PTCE,$GEPa,$RFIN,$RFIS,$PPRE,$PPO
S,$MPRE,$MPOS,$MPPA,$MPCU,$GPPE,$GPOS,$GRMD,
$SPRI,$GRML,$MVEC,$MVEE,$REVI,$ARTI,$PATT,$MV
DC,$PRCU,$SESEX,$ENVE,$PRLI,$EDCN,$APEPRE,$APE
POS,$NERE) {
$ranking="S";
if ( $MVDC <= 83.5
&& $GPPE <= 8.0
&& $APEPRE > 32.0
) $ranking=P;
if ( $MVDC > 83.5
&& $MPRE <= 40173.5
&& $APEPRE <= 17543.5
) $ranking=P;
return($ranking);
}

```

El modelo presenta una tasa de error que no supera el 20%.

IV. CONCLUSIONES

El modelo creado puede ser empleado para generar diagnósticos que permitan fortalecer y mejorar aspectos de la gestión de la universidad de tal manera que la Universidad se podrá gerenciar de manera más eficiente, integrando los resultados del modelo en sus planes de desarrollo.

Los anteriores indicadores encontrados como prioritarios con árboles de decisión coinciden parcialmente con los utilizados por Boletín Científico Sapiens Research en su Ranking: Programas de maestrías y doctorados, Grupos de investigación validados por Colciencias y Revistas indexadas en Publindex.

Se muestra la efectividad de utilizar el método de clasificación mediante la aplicación de un algoritmo de árboles de decisión para el manejo de información de alto volumen y análisis multivariado para el apoyo en la toma de decisiones en las universidades públicas al aplicar como datos de entrada, para el aprendizaje, los indicadores de gestión incluida su clasificación y utilizar el algoritmo como un pronosticador ante la variación de ciertos parámetros y valores en la entrada.

Los árboles de decisión obtenidos son de muy fácil entendimiento y de utilización, ya que limita la labor de priorización de los indicadores de gestión de 30 indicadores a unos pocos (cuatro o cinco) indicadores, y cuando se tengan los indicadores de gestión del año 2012 se podrá usar el árbol hallado con los indicadores desde el 2006 hasta el 2011.

RECOMENDACIONES

La Universidad Tecnológica de Pereira debe trabajar en el fortalecimiento de los indicadores considerados de mayor peso en la incidencia de una mejor posición dentro de los

recursos que se transferirán a las universidades en la vigencia 2014.

Cuando el Ministerio de Educación publique los indicadores y la correspondiente distribución de recursos de la presente vigencia se debe ejecutar nuevamente el algoritmo aquí planteado para buscar disminuir el error en las predicciones para las posteriores distribuciones de recursos.

En este trabajo mostramos la forma de utilizar el algoritmo pero corresponde al área de planeación dar la verdadera aplicabilidad que tiene este algoritmo en el diseño de simulaciones apoyadas con la herramienta para la toma de decisiones más acertadas al respecto.

REFERENCIAS

Referencias de publicaciones periódicas:

- [1]. ARENAS VALENCIA, Wilson. Una mirada crítica a los indicadores de desempeño de las universidades públicas desde la perspectiva del análisis envolvente de datos -DEA-. Tesis de Grado Universidad Tecnológica de Pereira. 2005.
- [2]. ARIAS MONTTOYA, Reinel. Detección temprana de fallas en la red de internet banda ancha aplicando minería de datos. Proyecto de Grado, Universidad Tecnológica de Pereira, 2010.
- [3]. ARIAS RODRÍGUEZ, A. El control de la gestión financiera de las universidades: nuevas tendencias. 1997.
- [4]. ASCUN. Asociación colombiana de universidades. Políticas y estrategias para la educación superior de Colombia 2006 – 2010: De la Exclusión a la Equidad II. Hacia la construcción de un sistema de educación superior más equitativo y competitivo, al servicio del país. Bogotá D.C. Enero de 2007.
- [5]. BURGUES, C. A tutorial on Classification Techniques for Pattern Recognition, Knowledge Discovery and Data Mining”, 2(2), 2003.
- [6]. CALANCHA ZUÑIGA, Niefar Abgar. Breve aproximación a la técnica de árbol de decisiones.
- [7]. GODSHALK EB, Timothy DH. Factor and principal component analyses as alternatives to index selection. Teoretical Aplied Gen 1998.
- [8]. JARAMILLO, Alberto. El financiamiento de la educación superior en Colombia: retos y tensiones, 2010.
- [9]. SANTA CHÁVEZ, John Jairo; ARDILA URUEÑA, William; PÉREZ HERNÁNDEZ, Lucas. Estudio de las variables del modelo de indicadores de gestión de las universidades públicas colombianas. 2012.

- [10]. SILVA, Jaime; GARCÍA, Manuel Felipe; MARTÍNEZ, Enrique. Situación presupuestal y de eficiencia de las universidades públicas colombianas. 2002.
- [11]. SOTO MEJÍA, José A., ARENAS VALENCIA, Wilson y TREJOS C., Álvaro A. La producción académica como uno de los indicadores del desempeño de las universidades públicas colombianas desde la perspectiva del Análisis Envoltante de Datos. En: Scientia Et Technica, Vol. XI, N° 28, Universidad Tecnológica de Pereira, octubre de 2005, p. 114.
- [12]. TORO, M, Eliana, PÉREZ, H, Lucas. Reducción de la dimensionalidad con componentes principales y técnica de búsqueda de la proyección aplicada a la clasificación de nuevos datos. Universidad Distrital Francisco José de Caldas. Revista Tecnura, Año XIII, N° 35, Agosto de 2007.
- [13]. TSUKAMOTO, Y, NAMATAME, A. Evolving neural network models. Evolutionary Computation, 1996. Proceedings of IEEE International Conference on 20-22 May 1996.
- [14]. UNIVERSIDAD NACIONAL DE COLOMBIA. Claves para el debate público. Sistema Universitario Estatal (SUE). El dilema entre los intereses colectivos de las Universidades y los del Gobierno de Turno. Bogotá, Colombia, octubre de 2010, número 41.
- [15]. UNIVERSIDAD TECNOLÓGICA DE PEREIRA. Taller de indicadores de eficiencia y su uso como tablero de mando. Pereira, 2004.
- [16]. UNIVERSIDAD TECNOLÓGICA DE PEREIRA. Información sobre las variables de los indicadores de las universidades públicas. Estudio de las variables del modelo de indicadores de la gestión de las universidades públicas colombianas Scientia et Technica Año XV, No. 45, Diciembre de 2010.
- [17]. VICEMINISTERIO DE EDUCACIÓN SUPERIOR. Metodología para la distribución de recursos. Ley 30 de 1992, Artículo 87. Año 2008
- [20]. JIMÉNEZ MOSCOVIT, Leonardo. Un modelo conceptual para el desarrollo de árboles de decisión con programación genética.
- [21]. JIMENEZ, L. and LANDGREBE, D. A. Projection Pursuit in High Dimensional Data Reduction: Initial Conditions, Feature Selection and the Assumption of Normality, presented at the IEEE International Conference on Systems, Man and Cybernetics, Vancouver Canada, October 1995.
- [22]. LEY 30 DE 1992. Por la cual se organiza el servicio público de la Educación Superior.
- [23]. MINISTERIO DE EDUCACIÓN NACIONAL. Modelo de Indicadores de Gestión, Propuesta de distribución recursos Artículo 87 Ley 30 de 1992, septiembre 4 de 2012.
- [24]. MONTGOMERY DC, Peck EA. Introduction to linear regression analysis". 2a Edition, J. Wiley and Sons, New York 1986.
- [25]. PÉREZ, L, MORA, J, PÉREZ, S, Reducción de las Características asociadas al problema de localización de fallas en sistemas de distribución. Scientia et technica, Vol XIII, N° 35, Agosto de 2007.
- [26]. PÉREZ, C. y SANTÍN, D. Minería de Datos: técnicas y herramientas. Madrid: Ediciones Paraninfo, S.A. 2008.
- [27]. PROCEEDINGS OF 2005 INTERNATIONAL CONFERENCE. On Volume3, 18-21 Aug. 2005.
- [28]. E. Clarke, *Circuit Analysis of AC Power Systems*, vol. I. New York: Wiley, 1950, p. 81.
- [29]. G. O. Young, "Synthetic structure of industrial plastics," in *Plastics*, 2nd ed., vol. 3, J. Peters, Ed. New York: McGraw-Hill, 1964, pp. 15-64.
- [30]. J. Jones. (1991, May 10). *Networks*. (2nd ed.) [Online]. Available: <http://www.atm.com>

Referencias de libros:

- [18]. JOHNSON RA, WICHERN DW. Applied multivariate statistical analysis. U. Wisconsin, Madison. PrenticeInc., 1982
- [19]. JIAN-PEI, Zhang, ZHONG-WEI LI, Jing Yang. A parallel SVM training algorithm on large – scale classification problems. Machine Learning and Cybernetics. 2005.