

3.5. CONFERENCIA PARALELA 5

Identificación de las características que inciden en el rendimiento académico de los estudiantes de pregrado de la universidad de Nariño usando minería de datos.

Sandra Viviana Escobar Madroño, Jaime Harvey Enríquez Tulcán,
sandraescobar2008@hotmail.com, j.enriquez@utp.edu.co, Universidad de Nariño.

Resumen: El rendimiento académico en colegios y universidades es un tema de investigación de interés continuo; se aprovecha cualquier técnica, metodología o herramienta nueva para estudiarlo. Aquí, se abordó el problema de identificar características de rendimiento académico en estudiantes de pregrado entre los años 2010 y 2014 de la universidad de Nariño usando minería de datos. Interesa predecir el promedio académico acumulado y determinar qué variables influyen más en la predicción a través de técnicas cuya variable objetivo sea numérica, se selecciona la mejor de ellas, para este caso máquina de vectores de soporte (SMV-R); y se continúa el estudio usando la metodología CRISP-DM. Adicionalmente, se ejecuta un análisis de clustering para tratar de encontrar patrones en los datos y se obtienen conclusiones.

Palabras claves. Rendimiento, académico, minería, datos

1. Presentación.

La Minería de Datos es un campo interdisciplinario, que en términos generales; organiza, procesa, analiza y genera reportes de grandes volúmenes de datos con el objetivo de descubrir información útil en la forma de relaciones, tendencias y patrones significativos y nuevos que pueden ayudar en la toma de decisiones y mejoramiento de procesos para organizaciones o empresas de diferente tipo.

(Pérez López and Santín Gonzalez) mencionan respecto de la minería de datos “Las técnicas de minería de datos persiguen el descubrimiento automático del conocimiento contenido en la información almacenada de modo ordenado en grandes bases de datos. Estas técnicas tienen como objetivo descubrir patrones, perfiles, y tendencias a través del análisis de los datos utilizando tecnologías de reconocimiento de patrones, redes neuronales, lógica difusa, algoritmos genéticos y otras técnicas avanzadas de análisis de datos”, a esta lista de técnicas se puede agregar: clustering, predicción y otras técnicas de análisis multivariado.

En los últimos años ha habido un interés creciente en utilizar la Minería de Datos en la educación, en colegios y especialmente en universidades; para analizar el rendimiento académico estudiantil y en función de los resultados tomar decisiones que beneficien a los estudiantes y a las instituciones.

En la actualidad existe una elevada preocupación en torno al bajo nivel de aprovechamiento estudiantil reflejado en altos índices de mortalidad académica, deserción, entre otros y los factores que pueden influir en este, por lo que varios centros educativos muestran especial interés en investigar este tema con el fin de establecer políticas institucionales que permitan actuar de manera preventiva frente a situaciones que puedan afectar el rendimiento académico de un estudiante y no de una manera recuperadora como sucede actualmente.

El rendimiento académico es un factor que se ve influenciado tanto por las condiciones internas de las instituciones educativas, como por la labor docente llevada a cabo por cada uno de sus maestros, sin embargo no se puede desconocer cómo las características propias de los estudiantes influyen de manera positiva o negativa en el rendimiento académico. Por lo tanto existen muchos factores que intervienen o condicionan este rendimiento, de ahí que este trabajo centre su interés en determinar las características relacionadas con las condiciones de vida, situación socioeconómico, características demográficas, antecedentes escolares y el rendimiento académico de los estudiantes de pregrado de la Universidad de Nariño, definiendo; en esta investigación, el rendimiento académico como el promedio acumulado resultado de las valoraciones cuantitativas emitidas por semestre académico.

2. Desarrollo de la temática.

1. Introducción Minería de datos y Rendimiento Académico
2. Definición del problema
3. Antecedentes
4. Objetivos
5. Diseño metodológico
6. Datasets
7. Correlación
8. Importancia de las variables
9. Máquinas de vectores de soporte
10. K-Means
11. Entendimiento y organización de los datos
12. Modelamiento
13. Evaluación
14. Resultados
15. Conclusiones
16. Recomendaciones

3. Conclusiones.

1. De los modelos de minería de datos para predicción, el mejor de todos fue las máquinas de vectores de soporte, esta misma situación se presenta en los antecedentes revisados que utilizaron esta técnica y que de verdad la recomiendan para problemas de rendimiento académico.
2. Como están provistos los datos, el modelo de segmentación no permite identificar claramente un clúster para el rendimiento académico bajo.
3. Luego de haber analizado los modelos de predicción, se puede ver que el rendimiento es más bajo en los 2 primeros semestres. Esta situación se evidencia en el análisis de las variables para el problema de predicción en ambos repositorios de las carreras técnicas y las humanísticas.
4. Las variables de los puntajes del examen ICFES si influyen en el rendimiento académico.
5. El sexo, la situación socioeconómica y la demográfica del estudiante afectan su rendimiento académico.
6. De los resultados se puede ver que si un estudiante tiene un promedio ponderado alto para el ingreso a la universidad, no garantiza que su rendimiento académico sea el mejor.

7. El rendimiento medio es compartido por hombres y mujeres, el rendimiento muy bajo es una característica mayoritariamente masculina y las mujeres tienen mejor rendimiento académico que los hombres, se puede evidenciar en los modelos de predicción y segmentación.

4. Referencias bibliográficas.

- Alcover, R., et al. *Análisis Del Rendimiento Académico En Los Estudios de Informática de La Universidad Politécnica de Valencia Aplicando Técnicas de Minería de Datos*.
- Álvarez, María Teresa, and Hernán García. *Factores Que Predicen El Rendimiento Universitario*. Universidad de Nariño, 1996.
- Bramer, Max A. *Principles of Data Mining*. 2007, doi:10.1007/978-1-84628-766-4.
- Carvajal Olaya, Patricia, et al. *MODELOS DE PREDICCIÓN DEL RENDIMIENTO ACADÉMICO EN MATEMÁTICAS I EN LA UNIVERSIDAD TECNOLÓGICA DE PEREIRA*. 2009, <http://revistas.utp.edu.co/index.php/revistaciencia/article/download/2323/1237>.
- Chapman, Pete, et al. "Crisp-Dm 1.0." *CRISP-DM Consortium*, 2000, p. 76, doi:10.1109/ICETET.2008.239.
- Deng, Naiyang, et al. *Support Vector Machines Optimization Based Theory, Algorithms, and Extensions*. 2013.
- Harwati, et al. "Mapping Student's Performance Based on Data Mining Approach (A Case Study)." *Agriculture and Agricultural Science Procedia*, vol. 3, 2015, pp. 173–177, doi:10.1016/j.aaspro.2015.01.034.
- Hastie, Trevor, et al. "The Elements of Statistical Learning." *Elements*, vol. 1, 2009, pp. 337–387, doi:10.1007/b94608.
- IBM, Corp. *Nodos de Modelado de IBM SPSS Modeler 18.0*. 2016.
- Kantardzic, Mehmed. *Data Mining Concepts, Models, Methods, and Algorithms*. Vol. 12, IEEE Press, 2011, doi:10.1007/978-3-642-19721-5.
- Ktona, Ana, et al. "Extracting Relationships between Students' Academic Performance and Their Area of Interest Using Data Mining Techniques." *2014 Sixth International Conference on Computational Intelligence, Communication Systems and Networks*, 2014, pp. 6–11, doi:10.1109/CICSyN.2014.18.
- Lanzarini, Laura, et al. *Academic Performance of University Students and Its Relation with Employment*. 2015.
- Merchán, S. M., and J. A. Duarte. "Analysis of Data Mining Techniques for Constructing a Predictive Model

- for Academic Performance." *IEEE Latin America Transactions*, vol. 14, no. 6, 2016, pp. 2783–2788, doi:10.1109/TLA.2016.7555255.
- Nariño, Universidad de. *ESTATUTO ESTUDIANTIL DE PREGRADO DE LA UNIVERSIDAD DE NARIÑO*. 1998, p. 60, <http://www2.udenar.edu.co/recursos/wp-content/uploads/2017/08/document-est.pdf>.
- Orallo, Jose Hernandez, et al. "Introducción a La Minería De Datos." *Introduccion a La Minería de Datos*, 2004, <http://users.dsic.upv.es/~flip/LibroMD/>.
- Pérez López, César. *MODELOS PREDICTIVOS, REDES NEURONALES Y TECNICAS DE SEGMENTACION Con IBM SPSS MODELER*. 2016.
- Pérez López, César, and Daniel Santín Gonzalez. *Minería de Datos. Técnicas Y Herramientas*. Paraninfo, 2007.
- Perez Marqués, Maria. *Minería de Datos a Través de Ejemplos*. 2015.
- Poh, Norman, and Ian Smythe. "To What Extend Can We Predict Students' Performance? A Case Study in Colleges in South Africa." *IEEE SSCI 2014 - 2014 IEEE Symposium Series on Computational Intelligence - CIDM 2014: 2014 IEEE Symposium on Computational Intelligence and Data Mining, Proceedings*, 2015, pp. 416–421, doi:10.1109/CIDM.2014.7008698.
- Porcel, Eduardo Adolfo; Dapozo, Gladys Noemí;, and Maria Victoria Lopéz. *Predicción Del Rendimiento Académico de Alumnos de Primer Año de La FACENA (UNNE) En Función de Su Caracterización Socioeducativa*. Aug. 2010, <https://redie.uabc.mx/redie/article/view/264/730>.
- Saltelli, Andrea, et al. *Sensitivity Analysis*. Wiley, 2009.
- Slim, Ahmad, et al. "Predicting Student Success Based on Prior Performance." *IEEE SSCI 2014 - 2014 IEEE Symposium Series on Computational Intelligence - CIDM 2014: 2014 IEEE Symposium on Computational Intelligence and Data Mining, Proceedings*, 2015, pp. 410–415, doi:10.1109/CIDM.2014.7008697.
- Sorour, Shaymaa E., et al. "Comments Data Mining for Evaluating Student's Performance." *Proceedings - 2014 IIAI 3rd International Conference on Advanced Applied Informatics, IIAI-AAI 2014*, 2014, pp. 25–30, doi:10.1109/IIAI-AAI.2014.17.
- SPSS Modeler. *Algorithms Guide*. 2016.
- Tobergte, David R., and Shirley Curtis. "Machine Learning with R." *Journal of Chemical Information and Modeling*, vol. 53, no. 9, 2013, doi:10.1017/CBO9781107415324.004.
- Vialardi, César, et al. "A Data Mining Approach to Guide Students through the Enrollment Process Based on

Academic Performance.” *User Modeling and User-Adapted Interaction*, vol. 21, no. 1–2, 2011, pp. 217–248, doi:10.1007/s11257-011-9098-4.

Wendler, Tilo, and Sören Gröttrup. *Data Mining with SPSS Modeler*. 2016, doi:10.1007/978-3-319-28709-6.