

Analítica del Aprendizaje, visualización de trayectoria académica

M. Peña¹, F. Bravo², L. Illescas-Peña²,

1 Universidad de Cuenca, Facultad de Ciencias Químicas, mario.pena@ucuenca.edu.ec

2 Universidad de Cuenca, Facultad de Filosofía, {fabian.bravo, lourdes.illescasp}@ucuenca.edu.ec

Abstract.

In the context of university academic management, several proposals have been developed for the study of analysis and visualization of learning trajectories. Bearing in mind that the educational trajectory is the trajectory of the student traveled at a given time from entry to the end of the stay, it can be considered that the use of technology could extract or highlight relevant information that is not seen directly with the tools traditional. The visualization of data in educational environments has become a challenge due to the large amounts of information available. The responsibilities of educational administrators require a clear visual proposal adapted to queries based on an academic context. Therefore, it was proposed to generate a tool that generates dashboard based on relevant variables of the students. To do this, the proposal began with a review of the literature that helped analyze the different ways of visualizing the data of academic trajectories. Subsequently, a dynamic visualization was formulated to explain the teachers and authorities through the learning analysis panel, based on the use of parallel coordinates that present multidimensional data over time. The sample was constituted by records of 1975 students of an Ecuadorian university, of the cohort that began in March 2013, distributed by faculties and careers. The technique used allowed us to discover trends and relationships between dimensions, improving the understanding of the trajectory patterns of students, the trends of school dropout, either increase or decrease performance, among other relationships. The consultations allowed to filter data by variables such as: faculties, careers, students and intervals of scores. Finally, the validation of the proposal was made based on the relevance of the dashboard, in response to the inquiries of the academic manager.

Resumen.

En el contexto de la gestión académica universitaria, se han desarrollado varias propuestas para el estudio del análisis y la visualización de las trayectorias de aprendizaje. Teniendo en cuenta que la trayectoria educativa es la trayectoria del estudiante desde la entrada hasta el final de los estudios universitarios, se puede considerar que el uso de la tecnología podría extraer o resaltar información rele-

vante que no se ve directamente con las herramientas tradicionales. La visualización de datos en entornos educativos se ha convertido en un desafío, debido a la gran cantidad de información disponible. Las responsabilidades de los administradores educativos requieren una clara propuesta visual adaptada a las consultas basadas en un contexto académico. Por lo tanto, se propuso generar una herramienta que genere un gráfico dinámico basado en variables relevantes de los estudiantes. Para ello, la propuesta comenzó con una revisión de la literatura que ayudó a analizar las diferentes formas de visualizar los datos de las trayectorias académicas. Posteriormente, se formuló una visualización dinámica para explicar a docentes y autoridades, basada en el uso de coordenadas paralelas que presentan datos multidimensionales a lo largo del tiempo. La muestra estuvo constituida por registros de 1975 estudiantes de una universidad ecuatoriana, de la cohorte que comenzó en marzo de 2013, distribuida por facultades y carreras. La técnica utilizada nos permitió descubrir tendencias y relaciones entre dimensiones, mejorando la comprensión de los patrones de trayectoria de los estudiantes, las tendencias de deserción escolar, el aumento o disminución en el rendimiento, entre otras relaciones. Las consultas permitieron filtrar datos por variables tales como: facultades, carreras, estudiantes e intervalos de puntajes. Finalmente, la validación de la propuesta se realizó en función de la relevancia de la visualización y en respuesta a las preguntas de un administrador académico.

Keywords. Learning Analytics, Academic Trajectory, Information Visualization, Learning Dashboards

1 Introducción

El estudio, la predicción del desarrollo académico, así como las investigaciones sobre los factores que describen el éxito académico y la permanencia de los estudiantes, son temas de alto impacto en la educación superior [1]. Es así, que el interés de las instituciones por identificar con claridad el desempeño estudiantil a través de tasas de deserción, retención, aprobación, tasas de egreso, etc., han generado investigaciones en el área, que surgen de una necesidad particular basadas en criterios académicos de acreditación, apoyo académico, planificación administrativa, etc.

En una época donde hay un pleno reconocimiento a la diversidad de circunstancias que definen las trayectorias de los jóvenes, estamos en condiciones de contribuir de mejor manera en su formación [2], por eso las instituciones realizan esfuerzos para conocer mejor a sus estudiantes “sus perfiles, necesidades, intereses, dificultades y proyectos de vida; ello requiere de investigación y del desarrollo de instrumentos y sistemas de seguimiento que faciliten el acceso a la información relevante” [3].

Por otro lado, el rendimiento académico es un importante predictor del comportamiento en otros niveles de educación y de otros resultados laborales importantes, como el desempeño laboral y el salario [4]. En este sentido, la teoría

sociocognitiva de la motivación plantea comprender el rendimiento académico con la premisa de que los comportamientos estudiantiles resultan ser basados en una función de los deseos de alcanzar objetivos particulares, centrándose principalmente en dos objetivos dominantes: aprendizaje y rendimiento [5]. Es así que, la definición de un rendimiento académico varía de acuerdo al enfoque o audiencia en función del significado que tenga para el sujeto de acuerdo a su situación particular, de ahí que puede considerarse como un concepto multidimensional, relativo y contextual [6]. El rendimiento inmediato utilizado con mayor frecuencia para evaluar el rendimiento académico de la enseñanza superior, considera a los resultados obtenidos por los estudiantes a lo largo de los estudios hasta obtener la titulación, y se mide en términos de éxito, es decir, superar las exigencias para aprobar una materia en función de un determinado periodo [6].

Paralelamente a estas investigaciones, una de las áreas dedicadas al estudio de los comportamientos académicos es la Analítica del Aprendizaje, utilizada mayoritariamente en educación superior para la toma de decisiones administrativas; y, que combina grandes conjuntos de datos, técnicas estadísticas y modelos predictivos. Se podría pensar que, es la práctica de extraer datos institucionales para producir "inteligencia procesable" [7]. En este contexto Cambell y Oblinger [7], definen la Analítica Educativa como un mecanismo que trabaja en 5 fases: captura, reporte, predicción, acción y refinamiento.

La fase de captura de la información en este estudio, requirió trabajar en base a una trayectoria académica. Para el cálculo de la trayectoria, Mares [8] propuso hacerlo en términos del promedio alcanzado en la universidad, la continuidad en los estudios y la manera de acreditar las materias. Estas actividades permitieron distinguir la orientación de las acciones hacia los alumnos o hacia la institución [9]. Basado en ello nuestra propuesta consideró analizar la primera categoría basada en competencias académicas generando una proyección que permitiera distinguir acciones orientadas directamente a los alumnos. Además, el uso de la tecnología en cada tarea diaria permite la recolección de grandes cantidades de datos observables en eventos que toman lugar en diferentes ambientes. La mayoría de las herramientas son capaces de almacenar los detalles de las acciones ejecutadas por los usuarios en archivos conocidos como *logs*. Estos archivos por supuesto, pueden ser analizados para inferir información que no es visible directamente por la mayoría de aplicaciones [10]. En educación, el registro de comportamientos académicos, son registrados por estos *logs* y pueden evidenciarse a través de los registros tradicionales de calificaciones y buscar visualizar el comportamiento académico de un colectivo en un periodo dado.

Por otro lado, la fase de reporte permitió monitorear el aprendizaje de los estudiantes considerando que es un componente esencial en la calidad de la educación superior y es uno de los factores determinantes que diferencian escuelas efectivas de profesores ineficaces [11]. Resulta entonces importante, definir la forma en la que puede representarse el comportamiento y las trayectorias generadas por los estudiantes durante el desarrollo de sus estudios. Si se considera además, que el uso de la tecnología permite no sólo extraer la información académica tales como calificaciones, datos de estudiantes, etc., es también importante la definición de

visualizaciones que se adapten a consultas formuladas por los gestores académicos responsables de la toma de decisiones.

Si bien es cierto, las visualizaciones de datos pueden ser utilizadas para apoyar un análisis, también pueden ser aplicadas en dominios educativos con experiencias de aprendizajes. Sin embargo, estas visualizaciones deben tomar en cuenta ciertos aspectos como: la forma de acceder a la información y proteger los datos, aplicación de filtros de protección, soporte multiusuarios y la disponibilidad [10]. Las herramientas de visualización utilizan datos generados en diferentes ambientes de aprendizaje, los llamados sistemas de gestión de cursos CMS (Vizualising student traking), generan mucha información en línea sobre las actividades de los participantes. Estos sistemas tienen demanda en las modalidades de estudios en línea o a distancia, y es propio de las herramientas el generar reportes para control de tutores o administradores.

Mediante una propuesta de visualización de información, se pretende “conocer y comprender al estudiante para así potenciar sus competencias y si es necesario buscar las estrategias para favorecer su integración” [12]. La información visual facilita a las autoridades académicas el observar las trayectorias y mediante esa presentación gráfica resaltar regularidades y anomalías que se presenten, para tomar decisiones académicas en función de esos datos. En este esfuerzo, la institución de educación superior puede apoyar a sus estudiantes a integrarse al ambiente académico, a través de estrategias de acompañamiento o intervención para fortalecer sus falencias y dificultades [13], que les permita a los estudiantes incorporarse plenamente al ambiente académico.

Las visualizaciones de aprendizaje pueden ser consideradas como una clase específica de aplicaciones informáticas personales que representan un área de estudio de la interacción hombre-máquina y que ayudan a las personas a entender mejor un comportamiento [14]. Estas herramientas ayudan a los usuarios a mejorar el autoconocimiento a través del uso de herramientas de visualización y análisis de sus historias personales. El autoconocimiento presenta varios beneficios, como fomentar la percepción de comportamientos e incrementar el autocontrol [15]. Al considerar la visualización como uno de los elementos de gran importancia dentro de la Analítica del Aprendizaje, se considera además que estas representaciones del estado actual e histórico del aprendizaje hacen flexible la toma de decisiones ya que muchas de estas visualizaciones son empleadas para dar soporte a los docentes y para obtener una visión de la actividad de un curso [16] o de una cohorte académica.

En esta propuesta, se genera una visualización de una trayectoria educativa, basada en el uso de coordenadas paralelas. El objetivo de esta técnica consiste en explorar y comprender conjuntos de datos complejos de alta dimensión, el cual es un problema importante y desafiante. Entre las diversas técnicas desarrolladas, las coordenadas paralelas (CP) se han adoptado ampliamente para la visualización de conjuntos de datos de alta dimensión y multivariados [17]. Esta técnica muestra cada tupla de dato multidimensional como una polilínea que interseca los ejes en la posición que corresponde al valor de los datos para la dimensión correspondiente.

Mapean el espacio k-dimensional en un gráfico de dos dimensiones usando k ejes equidistantes que son paralelos. Los ejes corresponden a las dimensiones y son linealmente escalados desde el mínimo hasta el máximo valor de cada dimensión [18]. Las CP se pueden considerar como diagramas de enlace de nodos, en los cuales los nodos están en los ejes paralelos y las aristas son aquellas polilíneas que vinculan los nodos en dos ejes vecinos [19] y pueden ser usadas para explorar el espacio de datos para casi cualquier tipo de tarea de visualización una vez que esta es aplicada a un dominio particular [17]. La ventaja al explorar datos visualmente radica en que el usuario se involucra directamente en el proceso de minería de datos. Con las CP hasta una docena de dimensiones de datos (o incluso más) se visualizan bien, especialmente cuando la visualización se combina con características de interacción adecuadas, tales como el reordenamiento en tiempo real de los ejes de coordenadas, el resaltado de polilíneas, el filtrado interactivo controlado por el usuario en dimensiones individuales, etc [20]. Las CP mejoran la visualización grandes volúmenes de datos multidimensionales, ofreciendo la posibilidad a detectar grupos [21], valores atípicos/ruido y tendencias [22], o correlaciones [23] de manera rápida y eficaz. Además, la estructura original de los datos no se pierde y es fácilmente observable [24]. Por la versatilidad y fácil interpretación que ofrecen las CP se usó esta técnica para presentar las trayectorias estudiantiles presentadas en este documento.

2 Metodología

En este estudio, como punto inicial se planteó realizar un seguimiento del progreso académico de los alumnos que pertenecían a una misma cohorte, en este caso, estudiantes que comenzaron sus estudios universitarios en el ciclo marzo 2013, en distintas facultades de la Universidad de Cuenca. Para analizar la evolución académica del alumnado que ingresó en esta universidad se realizó el seguimiento de ese rendimiento en 11 períodos o ciclos académicos, es decir, que el estudio analizó el registro de calificaciones hasta el ciclo marzo 2018.

La investigación es de tipo correlacional longitudinal, ya que se busca observar patrones y ofrecer predicciones sobre la trayectoria estudiantil a través del análisis de datos históricos en un periodo de tiempo definido. Los datos para el estudio fueron las calificaciones de 975 estudiantes que ingresaron a primer año de carrera en el periodo marzo-agosto de 2013. De esta cohorte se tiene todos los registros hasta el periodo marzo-agosto 2018 (11 periodos). La base de datos consta de 97303 registros sobre notas finales individuales (sobre 100 puntos) respecto a cada estudiante y en cada materia. Los campos relevantes que se utilizaron fueron: identificación del estudiante, facultad, carrera, nota final, estado de aprobación y periodo. Se calculó el promedio de calificaciones para cada estudiante en cada periodo de análisis.

Para la visualización de la información se generaron gráficos usando la librería D3 de JavaScript la cual permite utilizar documentos basados en datos, a través de la combinación de componentes de visualización. Particularmente se utilizaron coordenadas paralelas, logrando presentar información de la trayectoria de cada estudiante, y datos de las variables sociodemográficas. Esta visualización es dinámica,

y permite al usuario interactuar con los datos permitiendo comprender de forma visual las trayectorias académicas que describen los estudiantes para encontrar variables que puedan estar asociadas con sus rendimientos académicos.

En los ejes extremos se colocaron variables categóricas que contienen las 48 carreras y 12 facultades de estudio. Los demás ejes son variables continuas que corresponden al promedio de notas de todas las materias para cada periodo. Estos ejes son dimensiones que tienen un orden fijo pues son puntos de tiempo. Las polilíneas que unen los ejes representan la carrera, facultad y los promedios en cada periodo para todos los estudiantes. Esta polilínea representa la trayectoria para cada estudiante. Finalmente, para que la visualización sea efectiva, se agregaron interacciones que permiten filtrar y resaltar trayectorias por grupos o de forma individual. Las interacciones permiten realizar un minado de los datos directamente en el gráfico.

3 Resultados

La propuesta de visualización está basada en la presentación de información en coordenadas paralelas, que permitirían a los gestores educativos visualizar la trayectoria de un estudiante, o de todos los estudiantes de una carrera o una facultad inclusive.

La Fig[1] muestra en el eje izquierdo las 48 diferentes carreras de la Universidad de Cuenca; mientras en el eje derecho visualiza las 11 diferentes facultades de la Universidad de Cuenca; al centro en los ejes paralelos se aprecian los rendimientos académicos de los estudiantes en cada semestre, de manera que si se unen los puntos que representan los rendimientos de un estudiante en cada semestre, se dibuja una línea que muestra su trayectoria académica.

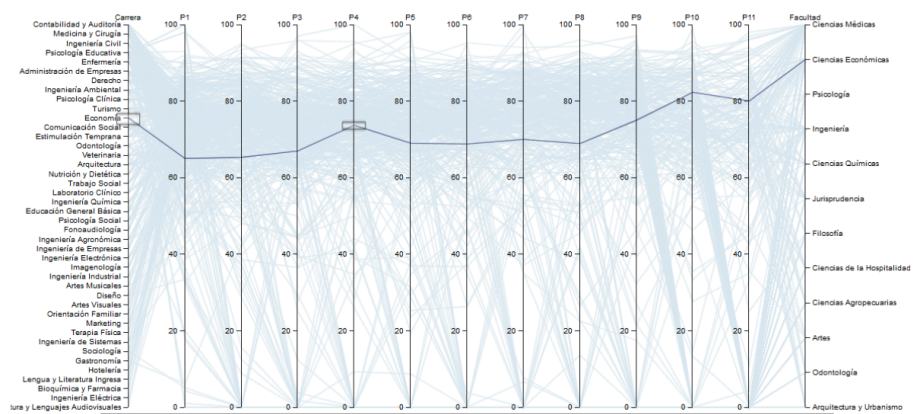


Fig. 1 Trayectoria de un estudiante de la carrera de Economía de la Facultad de Ciencias Económicas y Administrativas

Es también posible evidenciar las trayectorias de los estudiantes de una carrera, pudiéndose observar los comportamientos del grupo, así como las trayectorias individuales. En estas gráficas se pueden identificar algunas trayectorias especialmente irregulares de algunos estudiantes que desertan y en algunos casos retoman sus estudios. También es posible seleccionar la visualización de las trayectorias de los estudiantes de una facultad completa. En la Fig[2] un ejemplo aplicado a la carrera de ingeniería ambiental, en la que se evidencia además de los abandonos, el promedio de calificaciones obtenidas en cada periodo de cursado.

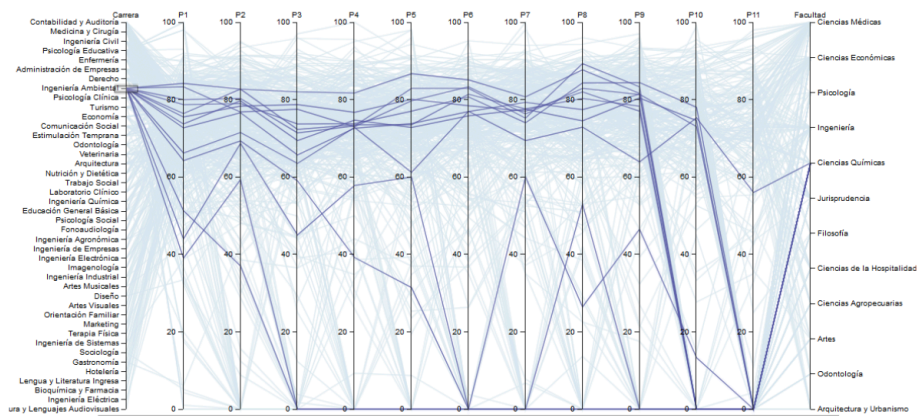


Fig. 2 Trayectorias de los estudiantes de la carrera de Ingeniería Ambiental de la Facultad de Ciencias Químicas

Es posible aplicar otro tipo de filtros, como en este caso se seleccionó a aquellos estudiantes que en el primer y segundo semestre obtuvieron promedios de hasta el 60/100 y se puede observar en la Fig[3], que muchos se retiran o sus trayectorias son bastante dilatadas y con muy bajos rendimientos. En este caso no se seleccionaron facultades o carreras específicas, es decir que el comportamiento tras la obtención de una calificación promedio regular, generalizaría un comportamiento similar a lo largo de los estudios con tendencia al abandono.

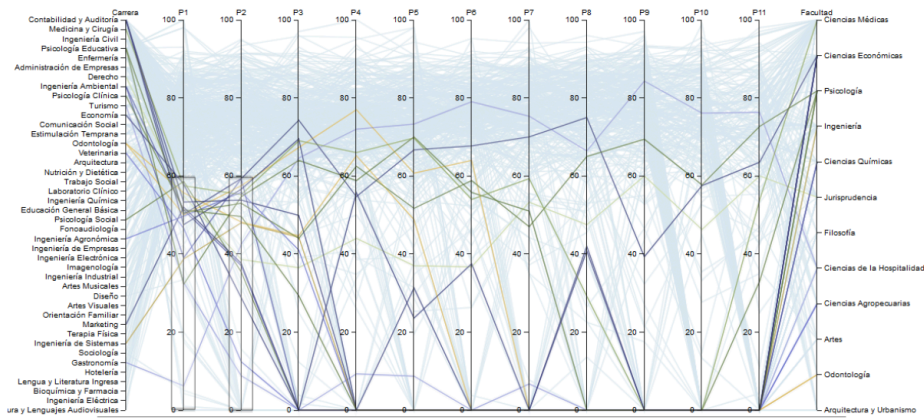


Fig. 3 Trayectorias de los estudiantes de la Universidad de Cuenca que en los dos primeros semestres obtuvieron hasta 60/100

Finalmente, este tipo de visualización permite la aplicación de filtros, en la que puede buscarse una tendencia, en la Fig[4] se generó una visualización con la aplicación de condiciones, filtrando trayectoria de aquellos estudiantes que incrementaron sus rendimientos académicos consecutivamente en los dos primeros semestres.

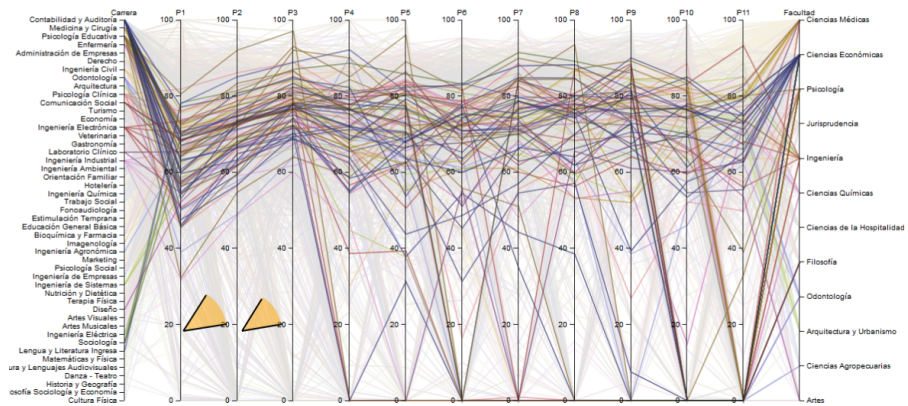


Fig. 4 Trayectorias de estudiantes de la Universidad de Cuenca que en los dos primeros semestres muestran consecutivamente incrementos de sus rendimientos académico

Las posibilidades de visualización de trayectorias académicas de individuos o de grupos que cumplen con ciertas condiciones que pueden ser seleccionadas en el mismo gráfico, permiten a los gestores educativos interpretar por si mismos la información y realizar un seguimiento exhaustivo de los estudiantes o carreras, así se pueden identificar algún tipo de problemas y decidir de forma inmediata alguna

acción preventiva o remedial con el fin de apoyar a estudiantes de bajos rendimientos, en riesgo de desertar; o también, de mejorar algunos indicadores de eficiencia académica con fines de acreditación institucional.

4 Conclusiones

La visualización de trayectorias estudiantiles no es una tarea trivial y requiere de nuevas herramientas para obtener resultados satisfactorios. Por lo tanto, en este artículo se describe un método para revelar las trayectorias estudiantiles visualizadas con coordenadas paralelas. Al calcular resúmenes estadísticos, se limita y posiblemente oculta información valiosa que puede estar dispersa en los datos. Mientras que al gestionar directamente con los datos crudos y con el poder intuitivo de la mente humana podemos extraer todo el conocimiento inmerso en los datos. La técnica de las coordenadas paralelas ha mostrado ser efectiva para presentar datos multidimensionales de variables cuantitativas e incluso cualitativas, que ayudan a potencializar los análisis sobre la información sin la pérdida de su estructura inherente. Además, la inclusión de interacciones tales como: resaltado y filtrado, son imprescindibles para poder minar directamente en el gráfico. Esto permite una retroalimentación inmediata y los diferentes aspectos y características de las trayectorias pueden ser reveladas. Las coordenadas paralelas son una herramienta poderosa y de fácil interpretación que hace posible poder observar los patrones, grupos, correlaciones y tendencias estudiantiles. Las posibilidades para utilizar las coordenadas paralelas sobre diferentes abstracciones de los datos son infinitas. El aporte que ofrece el visualizador a las autoridades es totalmente importante, ya que proporciona información relevante de manera clara para la toma de decisiones. Finalmente, se recomienda aplicar esta técnica sobre datos de otra institución de educación, tal vez con políticas diferentes de aprobación y cursado académico. La validación de la visualización podría permitir generalizar la herramienta para cualquier contexto educativo.

5 Bibliografía

- [1] L. M. Ruban and D. B. McCoach, “Gender Differences in Explaining Grades Using Structural Equation Modeling,” *Rev. High. Educ. J. Assoc. Study High. Educ.*, vol. 28, no. 4, pp. 475–502, 2005.
- [2] G. Vargas and V. Andrea, “Master’s Thesis Measurement of Scientific and Technological Production: Innovation Trends in Latin America and the Caribbean,” 2017.
- [3] M. Silva, “El primer año universitario. Un tramo crítico para el éxito académico,” *Perfiles Educ.*, vol. 33, no. SPE, pp. 102–114, 2011.
- [4] N. R. Kuncel, M. Crede, and L. L. Thomas, “The Validity of Self-Reported Grade Point Averages, Class Ranks, and Test Scores: A Meta-Analysis and Review of the Literature,” *Rev. Educ. Res.*, vol. 75, no. 1, pp. 63–82, 2005.

- [5] P. Fenollar, S. Román, and P. J. Cuestas, "University students' academic performance: An integrative conceptual framework and empirical analysis," *Br. J. Educ. Psychol.*, vol. 77, no. 4, pp. 873–891, 2007.
- [6] J. De Miguel, F., & Arias, "La evaluación del rendimiento inmediato en la enseñanza universitaria," *Rev. Educ.*, vol. 320, no. July, pp. 353–377, 1999.
- [7] J.P. Campbell, Peter B. DeBlois, "Academic Analytics: A New Tool for a New Era | EDUCAUSE," 42, 2007. [Online]. Available: <https://er.educause.edu/articles/2007/7/academic-analytics-a-new-tool-for-a-new-era>. [Accessed: 14-May-2018].
- [8] M. Chaín, R. y Ramírez, "TRAYECTORIAS ESCOLARES EN LA UNIVERSIDAD VERACRUZANA," *Revista de la Educación Superior*, 1997. [Online]. Available: <https://cdigital.uv.mx/bitstream/handle/123456789/5819/Publragu.htm?sequence=2&isAllowed=y>. [Accessed: 20-Jun-2018].
- [9] G. Mares, H. Rocha, O. Rivas, E. Rueda, R. Cabrera, and J. Tovar, "Identification of factors vinculated to desertion and academic trajectory of Psychology students at FES Iztacala," vol. 17, no. 55, pp. 189–207, 2012.
- [10] G. Dodd, "Glass: A Learning Analytics Visualization Tool," *Constr. Mater. Their Nat. Behav. Fourth Ed.*, no. May, p. 507, 2010.
- [11] R. Mazza and V. Dimitrova, "Visualising student tracking data to support instructors in web-based distance education," *Proc. 13th Int. World Wide Web Conf. Altern. track Pap. posters - WWW Alt. '04*, p. 154, 2004.
- [12] L. A. Villafra de Monroy, "LA INTEGRACIÓN DE LOS NUEVOS ESTUDIANTES DE LA UNIVERSIDAD AUTÓNOMA DE BUCARAMANGA A LA VIDA UNIVERSITARIA: UNA EXPERIENCIA INTEGRAL," *CLABES*, vol. 13, no. 978-956-303-372-4, 2013.
- [13] J. Matus, "Integración a la vida universitaria," *CLABES*, vol. 91, pp. 399–404, 2017.
- [14] I. Li, A. Dey, and J. Forlizzi, "A Stage-Based Model of Personal Informatics Systems," *J. Appl. Pharm. Sci.*, vol. 4, no. 1, pp. 29–32, 2014.
- [15] R. Loewenstein, G., Read, D., & Baumeister, *Time and Decision: Economic and Psychological Perspectives of Intertemporal Choice*. Retrieved from. Russell Sage Foundation., 2003.
- [16] K. Verbert, E. Duval, J. Klerkx, S. Govaerts, and J. L. Santos, "Learning Analytics Dashboard Applications," *Am. Behav. Sci.*, vol. 57, no. 10, pp. 1500–1509, 2013.
- [17] M. Tory, S. Potts, and T. Möller, "A parallel coordinates style interface for exploratory volume visualization," *IEEE Trans. Vis. Comput. Graph.*, vol. 11, no. 1, pp. 71–80, 2005.
- [18] D. Keim, "Information Visualization and Visual Data Mining," *IEEE Trans. Vis. Comput. Graph.*, vol. 8, no. 1, pp. 1–8, 2013.
- [19] H. Zhou, X. Yuan, H. Qu, W. Cui, and B. Chen, "Visual clustering in parallel coordinates," *Comput. Graph. Forum*, vol. 27, no. 3, pp. 1047–1054, 2008.
- [20] H. Hauser, F. Ledermann, and H. Doleisch, "Angular brushing of extended parallel coordinates," *Proc. - IEEE Symp. Inf. Vis. INFO VIS*, vol. 2002–

- Janua, pp. 127–130, 2002.
- [21] J. Heinrich and D. Weiskopf, “State of the Art of Parallel Coordinates,” *STAR Proc. Eurographics*, pp. 95–116, 2013.
 - [22] M. Novotný and H. Hauser, “Outlier-preserving focus+context visualization in parallel coordinates,” *IEEE Trans. Vis. Comput. Graph.*, vol. 12, no. 5, pp. 893–900, 2006.
 - [23] A. O. Artero, M. C. Ferreira De Oliveira, and H. Levkowitz, “Uncovering clusters in crowded parallel coordinates visualizations,” *Proc. - IEEE Symp. Inf. Vis. INFO VIS*, pp. 81–88, 2004.
 - [24] J. Johansson, P. Ljung, M. Jern, and M. Cooper, “Revealing structure within clustered parallel coordinates displays,” *Proc. - IEEE Symp. Inf. Vis. INFO VIS*, pp. 125–132, 2005.