

Pistas Educativas 127 (CITEC 2017), diciembre 2017, México, Tecnológico Nacional de México en Celaya

ANÁLISIS DE INFORMACIÓN DE ASPIRANTES A INGRESAR A INSTITUCIONES DE EDUCACIÓN SUPERIOR UTILIZANDO PENTAHO

Juan Ignacio Cerca Vázquez

Tecnológico Nacional de México en Celaya

nacho@itcelaya.edu.mx

Rubén Torres Frías

Tecnológico Nacional de México en Celaya

ruben.torres@itcelaya.edu.mx

Luis Alberto López González

Tecnológico Nacional de México en Celaya

luislao@itcelaya.edu.mx

José Jesús Sánchez Farías

Tecnológico Nacional de México en Celaya

jesus.sanchez@itcelaya.edu.mx

Resumen

Cuando concluye el proceso de admisión de una institución de educación superior (IES), en los registros institucionales como resultado del mismo, se cuenta con información recabada en base de datos mediante el registro y validación de aspirantes. Por la naturaleza del proceso, la recopilación y manejo de datos, dicha base de datos es consistente en datos de rendimiento académico, entre las que se mencionan: resultado de evaluaciones, becas, entre otros, sin perder de vista y tomando en consideración las políticas institucionales de manejo de información y datos personales, se identifica el procesamiento como un área de oportunidad para las entidades de desarrollo académico y seguimiento curricular

de dichas instituciones. Las áreas de impacto que se identifican dentro de este proceso es la toma de decisiones en base a datos estadísticos sobre deserción escolar, análisis de escuelas de procedencia, fortalezas y debilidades en áreas de conocimiento, programas de tutoría y seguimiento académico. El presente trabajo considera una propuesta de análisis utilizando las tecnologías de información y comunicación aplicada en la educación, en el ámbito de inteligencia de negocios utilizando Pentaho como herramienta de procesamiento y obtención de datos estadísticos, basados en procesos y actividades concretos, de donde proviene la propuesta de implementar una alternativa de apoyo a la toma de decisiones enfocada a los departamentos o áreas de las IES.

Palabra(s) Clave: Análisis de datos, Pentaho, Tecnologías de información, Toma de decisiones.

Abstract

At the end of the admission process of a superior education institution (SEI), there is a database with information collected through the registration and validation of applicants. Due to the nature of the process, data collection and management is consistent with data on academic performance, results of evaluations, scholarships, among others, which at all times and taking into account the institutional policies for the management of information and personal data, is identified as an opportunity area, for academic development entities and curricular monitoring of these institutions. The areas of impact identified within this process are decision-making based on statistical data on dropout, analysis of schools of origin, strengths and weaknesses in areas of knowledge, mentoring programs and academic follow-up. the present work considers a proposal of analysis using information and communication technologies applied in education, in the field of business intelligence using Pentaho as a tool for processing and obtaining statistical data, based on specific processes and activities, from which the proposal to implement an alternative of support to the decision making focused on the departments or areas of SEI.

Keywords: *Data analysis, Decision taking, Information technology, Pentaho.*

1. Introducción

Las Instituciones de Educación Superior (IES) en México se refieren a todas aquellas entidades definidas por la Secretaría de Educación Pública (SEP), como organismos responsables de realizar las funciones de docencia, generación y aplicación de conocimiento, extensión y difusión cultural, gestión de proyectos de investigación y la impartición de un amplio espectro de programas educativos a nivel licenciatura y posgrado, también definidos como estudios profesionales. Dentro de la SEP, se encuentra la Subsecretaría de Educación Superior, entidad de la que se constituye el "Tecnológico Nacional de México (TecNM). El 23 de julio de 2014 se publicó en el Diario Oficial de la Federación, el Decreto Presidencial por el que se crea la institución de educación superior tecnológica más grande de nuestro país, el Tecnológico Nacional de México. De acuerdo con el Decreto citado, "el TecNM se funda como un órgano desconcentrado de la Secretaría de Educación Pública, que sustituye a la unidad administrativa que se hacía cargo de coordinar este importante subsistema de educación superior" [SEP, 2017].

El Tecnológico Nacional de México se conforma por 266 instituciones y una Dirección General, distribuidas en los 31 estados de la República Mexicana y la Ciudad de México (CDMX). Los Institutos Tecnológicos del Tecnológico Nacional de México a lo largo de la historia, construida durante más de 65 años como IES a nivel nacional especializada en la formación de ingenieros. Para dimensionar la cobertura que posee el instituto, en el ciclo escolar 2014-2015 contó con una matrícula de 521,105 estudiantes inscritos en alguno de sus programas educativos, como se ve en tabla 1.

De acuerdo a la perspectiva nacional de estudiantes de bachillerato "La matrícula escolarizada de educación media superior ascendió a 4.8 millones de estudiantes en el ciclo escolar 2014-2015" [Presidencia de la República, 2017]. Tercer Informe de Gobierno. México D.F. el TECNM, tiene la capacidad de matricular a poco más del 9% de los estudiantes que potencialmente aspiren a estudiar una carrera.

Tomando como referencia dentro del TECNM al Instituto Tecnológico de Celaya, que será la entidad en la que se centre este trabajo, ya que de acuerdo a la información con que se cuenta, se proyecta implementar una tecnología de

información y comunicación, enfocada al área de conocimiento de inteligencia de negocios, la cual, es el área de las ciencias computacionales, que proporciona el fundamento teórico, del que parte la aplicación de modelos de explotación y unificación de datos, el apoyo se dará en la herramienta Pentaho, partiendo de un esquema de base de datos relacional, donde institucionalmente se encuentran ubicados físicamente los datos, para elaborar consultas de información que permitan el auxilio a la toma de decisiones en un posterior trabajo estratégico de las áreas responsables del proceso.

Tabla 1 Cobertura del TecNM para el ciclo escolar 2014-2015.

MATRÍCULA CICLO ESCOLAR 2014 - 2015			
Nivel de estudio	Institutos Tecnológicos Federales	Institutos Tecnológicos Descentralizados	Total
Técnico Superior Universitario	102	179	281
Licenciatura	305,318	211,191	516,509
Posgrado	3,690	625	4,315
Total	309,110	211,995	521,105
INSTITUCIONES QUE CONFORMAN EL TECNOLÓGICO NACIONAL DE MÉXICO			
Institutos Tecnológicos Federales y Centros	Institutos Tecnológicos Descentralizados		Total
Institutos Tecnológicos	Centros		
126	6	134	266

La inteligencia de negocios puede ser definida como una filosofía que incluye estrategias, procesos, aplicaciones, datos, productos, tecnologías y arquitecturas tecnológicas utilizadas para soportar la recolección, análisis y presentación de información del negocio [Dedic & Stainer, 2016].

El objetivo de implementar una tecnología de información y comunicación permitirá obtener reportes en tiempo real que enfocarán el trabajo colegiado de las entidades responsables del proceso en el análisis de resultados más que en la recolección, síntesis y presentación de datos.

El objetivo es el obtener ventaja del análisis de datos o también se denomina análisis descriptivo, ya que describe un estado pasado o actual, y delimita en una manera para que las áreas examinen los datos para entender las tendencias y obtener ideas. Las organizaciones hoy en día pueden elegir entre una sólida lista

de proveedores que ofrecen herramientas de BI. Según identifica a cerca de dos docenas de proveedores de BI y analíticos en su informe del cuadrante mágico 2017, con Microsoft, Glik y Tableau como líderes. Otros proveedores de herramientas de BI incluyen información, Builders, Sisense y Zoomdata. La mayor parte de estas herramientas poseen un costo, por lo que al identificar algunas alternativas se encontró a Pentaho como una opción que oferta un trial de 30 días suficiente para elaborar un proceso de evaluación. Contener las ventajas de la BI, se utilizará como software de prueba para este ejercicio.

Pentaho entonces es una herramienta de BI desarrollada por Hitachi Vantara y que proporciona una suite o conjunto de programas orientados hacia el procesamiento, minería de datos, graficación y reporte, con tipo de código *open source*.

2. Metodología

Para el uso de una tecnología de información se aplicará el conocimiento en el área de inteligencia de negocios o Business Intelligence (BI), que es una herramienta de gestión, y transformación de datos en inteligencia accionable que fungirá como apoyo a la toma de decisiones estratégicas y tácticas de una organización. Se tomará entonces nuestro servidor Pentaho, que ofrece facilidades en la compatibilidad de base de datos con las institucionales como se muestra en la figura 1. La descarga se realiza del sitio oficial de Pentaho que ofrece un trial de 30 días sin registrarse, posteriormente se requerirá el pago de una licencia para poder acceder al servidor, cabe señalar que, para entornos en producción, sí se requiere un licenciamiento.

- 8 Gb memoria RAM
- Procesador 64 bits
- 20 Gb libres posterior a la instalación
- Sistema operativo Linux, Windows o Mac de 64 bits

Adicionalmente es importante verificar el estatus de los puertos en servidores que ya posean la instalación de otro servicio http, ya que ocupará un puerto para la

interfaz web, la fase de configuración del servidor (figura 1) nos permitirá enlazar las vistas y consultas previamente obtenidas de la base de datos para la elaboración de los reportes.

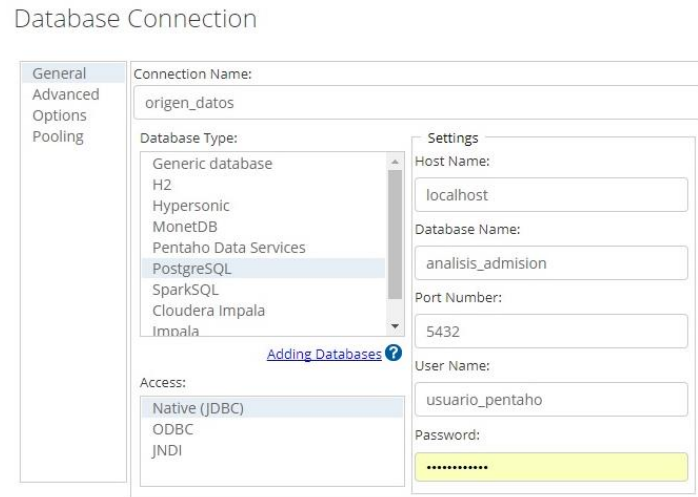


Figura 1 Configuración inicial de Pentaho.

Una vez concluida la instalación y configuración del servicio, se identifican las áreas de oportunidad que pueden ser impactadas por los resultados del proyecto, y con el propósito de integrar un resultado tangible y visual, el formato de los reportes, tanto como su contenido, datos y presentación, de esta manera, se considera estar en condiciones para definir con las entidades cliente, en este caso las áreas académicas cuál será, la clasificación y confidencialidad de los informes. De acuerdo a los parámetros definidos por el área académica se propone un primer resultado sujeto al análisis y validación de resultados, para ésta etapa que será posterior al resultado de esta investigación se desarrollará un trabajo colegiado de especialistas en ámbitos académicos, y que el resultado, al ser enfocado a la toma de decisiones, coadyuvará como herramienta de apoyo, para implementar acciones o políticas en base a estrategias, con una etapa de retroalimentación posterior donde pudiera aplicarse mejoras considerando nuevas áreas de oportunidad identificadas en el ejercicio.

Definiendo entonces que se busca obtener información que pueda ser clave en el apoyo a la toma de decisiones de las entidades responsables del proceso, tutores,

coordinadores de carrera y psicólogos, se iniciará definiendo que información es la que se posee para el procedimiento:

- El registro de datos de aspirantes.
- El registro de la escuela de procedencia, con ubicación geográfica y promedio de egreso.
- La encuesta de contexto que recolecta Ceneval.
- El cuestionario de los 16 factores de personalidad.
- Los exámenes de exploración de conocimientos.

En la conclusión del procedimiento de admisión y que se determina por parte de las autoridades institucionales, este registro de datos se encuentra definido en base de datos ubicada en los servidores institucionales, es importante señalar que esta información que es sensible por el manejo de datos personales, deberá ser resguardada y utilizada únicamente para los fines que se indican en el proceso, por tal motivo, en el sistema integral de información el estudiante es quien faculta al tutor, docente o coordinador, si comparte o no esta información, por tal motivo el universo de datos con el que se cuenta será el de aquellos estudiantes que proporcionen la autorización de tener acceso a sus datos. Para el proceso de admisión 2017 se tuvieron aproximadamente 2260 aspirantes, quienes presentaron los exámenes y entrevistas requeridas por el proceso correspondiente, la información resultante fue procesada por los sistemas institucionales y publicada oficialmente a los candidatos, de este número de aspirantes tomaremos un muestreo valido. Se obtendrá una consulta de base de datos para obtener los datos sobre las escuelas de procedencia, examen de exploración de ciencias básicas, examen de 16 factores de personalidad y encuesta de contexto de Ceneval.

El departamento de desarrollo académico ha requerido constantemente el procesamiento de la información de entrevistas y evaluaciones diagnosticas con el fin de proporcionar el seguimiento tutorial correspondiente, este primer ejercicio será enfocado en satisfacer esta demanda y entregar los resultados correspondientes al área de atención psicopedagógica departamental, ya

quedando a resguardo de este organismo el resultado detallado del presente trabajo. Como muestra de información obtenida en esta etapa metodológica la consulta de datos tiene un formato similar a lo que se presenta a continuación:

Escuelas de procedencia

Se aplica una proyección utilizando una consulta SQL a la base de datos aplicando inner join a las entidades de preficha, procedencia, municipios, esta proyección nos permite identificar qué porcentaje de hombres, mujeres, con que promedio, frecuencia y ubicación son candidatos o aspirantes a ingresar a la institución. Aquí se detectan las áreas de oportunidad siguientes:

- Procedencia de hombres y mujeres de los municipios y sus promedios de egreso.
- Instituciones con los más altos promedios (reclutamiento y seguimiento en bachillerato).
- Instituciones con programas de apoyo para becas o convenio de pase directo.
- Instituciones con los índices y promedios más bajos para estrategias de fortalecimiento (alumnos candidatos a ser identificados en el primer semestre).

Encuesta de 16 factores de personalidad

Se aplicará una proyección similar enlazando la preficha con la encuesta, en este caso existen estudiantes que por alguna razón no presentaron dicho cuestionario, y también podremos identificarlos, para elaborar esta proyección aplicaremos un *left join* entre el registro de prefichas, y el cuestionario de actitudes, tabla 2: el cuestionario de los 16 factores de personalidad de Catell se aplica como instrumento de apoyo a la evaluación de los candidatos, durante la entrevista de ingreso a la institución, con esta proyección se detectan áreas de oportunidad como:

- Elaboración de perfiles asociados a la transición de aspirante a estudiante del instituto.

- Patrones de conducta de aspirantes asociados a su desempeño escolar.

Tabla 2 Proyección que recupera la procedencia de los aspirantes.

ESTADO	LOCALIDAD	PREFICHA	FECHA NACIMIENTO	POBLACION	PROMEDIO BACH	SEXO
GUANAJUATO	SANTA CRUZ DE JUVENTINO ROSAS	P1603X	26/01/1998	SANTA CRUZ DE JUVENTINO ROSAS	10	F
GUANAJUATO	CELAYA	P1603X	18/10/1997	CELAYA	8	M
GUANAJUATO	CELAYA	P1603X	09/02/1998	CELAYA	8	F
GUANAJUATO	CELAYA	P1603X	07/05/1998	CELAYA	9	M

Proyección asociada al perfil

Se emitirá utilizando una vista una proyección asociada al perfil que devuelve Ceneval, junto con los resultados detallados de cada estudiante, en éste caso se opta por la implementación de una vista que permite la asociación de tuplas elaboradas a partir de múltiples entidades o tablas de la base de datos, y con estas tuplas implementar en las vistas de presentación de los datos, filtros que pueden ser por especialidad, por estatus entre otros, y que estos sean configurados directamente en el objeto de consulta, al realizar la vista se vería aproximadamente como se muestra en la tabla 3.

Tabla 3 Proyección para recuperar la prueba de 16 factores de personalidad.

FICHA	LUGAR	ESPECIALIDAD	PROMEDIO	ACEPTADO	RESPUESTAS FACTORES
F160224	48		70	SI	AAAACAAACAAAACC...
F160223	86		30.5	NO	AACAABCACAAAACCC...
F160222	32		68.5	SI	AAABABCCCAAAACCC...
F160221	223		23.5	NO	AAABCBABBBAAAAAB...

Cabe señalar que se detectan las áreas de oportunidad referentes a la relación entre las posibilidades de una discapacidad física, económica, familiar influyen en el desempeño escolar de los aspirantes y su evolución a estudiantes:

- Relación entre situación socioeconómica y niveles de deserción.
- Impacto de becas y apoyos previos en desempeño económico.
- Instituciones educativas y el nivel socioeconómico de aspirantes.

3. Resultados

La implementación de la herramienta permite la administración de las múltiples aplicaciones que contiene, en este trabajo se aplicó la herramienta de reporte, que viene incluida en la suite de prueba. la herramienta permite el diseño gráfico de la unidad de reporte una vez que se ha configurado el origen de datos y que se ha recuperado la consulta, es pertinente indicar que puede ser un origen de base de datos, o archivos separados por comas, para este ejercicio se trabajó con las proyecciones citadas en la metodología y con vistas, como se muestra en la figura 2, Algunos de los resultados, que no se encuentran clasificados fueron los siguientes:

- Reporte por procedencia, género, promedio, indicando que promedio poseen los estudiantes de determinados lugares de procedencia por género como se presentó en la tabla 3.
- Ciudad de Procedencia con mayor cantidad de aspirantes: Celaya (50%), Cortázar (17%), Apaseo el Alto (15%), Apaseo el Grande (14%).
- Nivel socioeconómico predominante en el instituto: Medio-Bajo (50%).
- Porcentaje de estudiantes que preferían estudiar otra carrera que a la que fueron admitidos (25%).
- Se genera un informe gráfico para los factores de personalidad.

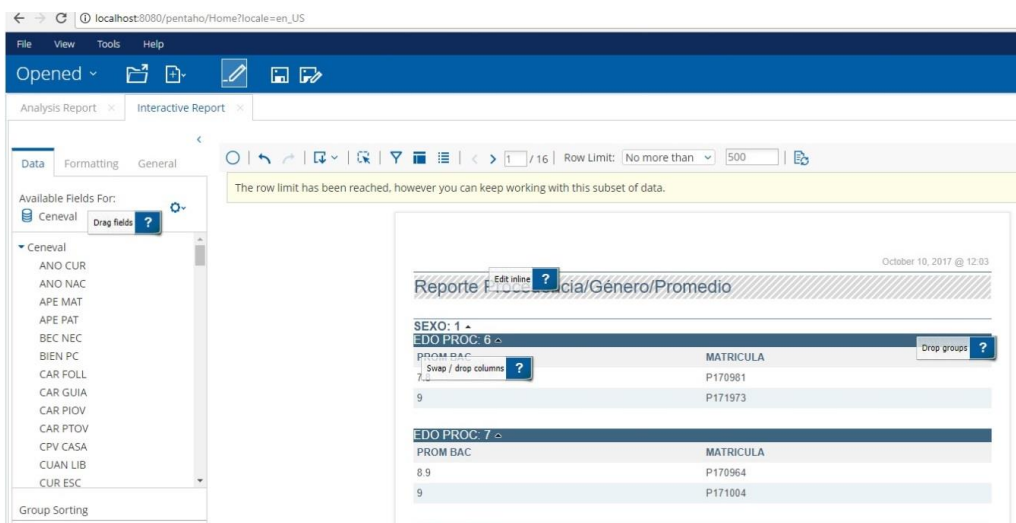


Figura 2 Configurando el reporte dado el origen de datos a partir de una vista.

Dentro de la información clasificada o reservada se obtuvieron reportes sensibles como:

- Afectación de discapacidad en base a desempeño escolar.
- Impacto del entorno familiar en el desempeño escolar.
- Instituciones públicas y privadas de procedencia con el mayor índice de nuevo ingreso, deserción y promedios.
- Clasificación de los resultados de contexto y actitudes por especialidad, este con el fin de elaborar el perfil del estudiante con respecto a la carrera que cursa, tabla 4.

Tabla 4 Fragmento de la vista con datos de la encuesta de contexto.

FICHA	ESCUELA_PROCEDENCIA	DSCSPECIALIDAD	ENTI_N AC	IMP_C AM	IMP_E CU	IMP_V ER	IMP_C ON	IMP_P AN	IMP_P CA
F172X XX	BACHILLERATO MANUEL CONCHA	INGENIERÍA EN GESTIÓN	GTO	NO	NO	NO	SI	NO	SI

Por citar algunos ejemplos del compendio de resultados que se tienen, el área académica de primera instancia considera asistirse del informe de la relación existente entre el promedio de bachillerato con el promedio obtenido en los exámenes de admisión, clasificándose por escuela, localidad, tipo de escuela, y carreras, esta estrategia identifica puntos importantes como la escuela de procedencia con el mayor número de ingresos, promedios (figura 3) y correspondencia entre su promedio de egreso y su promedio en el ITCelaya, los subsistemas escolares (estatal, federal, privado) agrupados por resultados con el fin de obtener la correspondencia entre carreras cursadas/solicitadas y el desempeño escolar en estas, identificando la compatibilidad entre estudios previos y posteriores en base a la especialidad, al final éste informe identifica plenamente las instituciones que perdurarán su convenio de pase directo y si éste es viable de manera parcial o total en especialidades cursadas-solicitadas.

En otro aspecto se considera un informe basado en la información del financiamiento o apoyo que los aspirantes reciben o recibieron en sus escuelas de procedencia, que al igual que el anterior ejemplo permite identificar y cuantificar la

cantidad de alumnos que reciben algún apoyo, por alguna circunstancia académica, capacidad diferente, estatal, federal, crear un punto inicial de identificación y seguimiento en otorgamiento de becas en el ITCelaya, el objetivo entonces es identificar a los alumnos que han tenido apoyo económico, su estatus y tomar decisiones sobre éste ámbito durante su estadía en el ITCelaya.

Reporte Promedio/Procedencia/Género

SEXO: 1

EDO PROC: 6

PROM BAC	MATRICULA
7.8	P170981
9	P171973

EDO PROC: 7

PROM BAC	MATRICULA
8.9	P170964
9	P171004

EDO PROC: 9

PROM BAC	MATRICULA
7.5	P172636
8	P172109
8	P172827
8.3	P171572
8.6	P171038

Figura 3 Ejemplo de un reporte exportado a PDF.

4. Discusión

Los resultados obtenidos en la investigación determinan una colección de datos que deberán ser sujetos a análisis y validación de las áreas correspondientes, el presente trabajo ha recibido como insumo la información real de los procesos de admisión vigentes en el ITCelaya, con éste resultado se han logrado implementar reportes para los tutores, como la encuesta de contexto y el reporte gráfico de 16 factores de personalidad, la ayuda brindada por la herramienta de reportes ha permitido dilucidar cuáles campos y datos son los más importantes para los tutores y entrevistadores para ser mostrados en las vistas y reportes correspondientes.

Dados los volúmenes de información la herramienta permite procesar en tiempo real y a través de un sistema de información los datos, enfocando los trabajos de las áreas correspondientes a la evaluación propia de los resultados, el impacto de

esta información es tangible en la elaboración de un perfil del estudiante dentro de una plataforma, que servirá como apoyo para el acompañamiento tutorial y la implementación de estrategias propias de las actividades pedagógicas.

Es importante señalar que para futuros procesos se seguirá elaborando un análisis de las tendencias y así poder identificar nuevas áreas de oportunidad, fortalezas y debilidades del proceso actual en base a objetivos concretos institucionales.

5. Bibliografía y Referencias

- [1] Dedic, N. & Stanier, C. (2016). *Towards Differentiating Business Intelligence, Big Data, Data Analytics and Knowledge Discovery*. United Kingdom: Staffordshire University.
- [2] Douglas Laney. (2017). *Infonomics: How to Monetize, Manage, and Measure Information as an Asset for Competitive Advantage*. USA: Routledge.
- [3] SEP. (2017). TecNM: <http://www.ses.sep.gob.mx/instituciones.html>.
- [4] Pentaho. (2017). Pentaho: <http://www.pentaho.com/> [10 Oct. 2017].
- [5] Presidencia de la República. (2015). *Tercer Informe de Gobierno*, Lic. Enrique Peña Nieto. México D.F.: Gobierno de los Estados Unidos Mexicanos.