Universidad de San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería Ciencias y Sistemas

PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

Randy Fernando Juárez Najarro
Asesorado por el Ing. José Julio Pineda Chinchilla

Guatemala, septiembre de 2019

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

TRABAJO DE GRADUACIÓN

PRESENTADO A LA JUNTA DIRECTIVA DE LA
FACULTAD DE INGENIERÍA
POR

RANDY FERNANDO JUÁREZ NAJARRO
ASESORADO POR EL ING. JOSÉ JULIO PINEDA CHINCHILLA

AL CONFERÍRSELE EL TÍTULO DE

INGENIERO EN CIENCIAS Y SISTEMAS

GUATEMALA, SEPTIEMBRE DE 2019

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

FACULTAD DE INGENIERÍA

NÓMINA DE JUNTA DIRECTIVA

DECANA	Inga. Aurelia Anabela Cordova Estrada
VOCAL I	Ing. José Francisco Gómez Rivera
VOCAL II	Ing. Mario Renato Escobedo Martínez
VOCAL III	Ing. José Milton de León Bran
VOCAL IV	Br. Luis Diego Aguilar Ralón
VOCAL V	Br. Christian Daniel Estrada Santizo
SECRETARIO	Ing. Hugo Humberto Rivera Pérez

TRIBUNAL QUE PRACTICÓ EL EXAMEN GENERAL PRIVADO

DECANO	Ing. Pedro Antonio Aguilar Polanco
EXAMINADOR	Ing. César Augusto Fernández Cáceres
EXAMINADOR	Ing. Luis Fernando Espino Barrios
EXAMINADOR	Ing. Sergio Arnaldo Méndez Aguilar
SECRETARIA	Inga. Lesbia Magalí Herrera López

HONORABLE TRIBUNAL EXAMINADOR

En cumplimiento con los preceptos que establece la ley de la Universidad de San Carlos de Guatemala, presento a su consideración mi trabajo de graduación titulado:

PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA

Tema que me fuera asignado por la Dirección de la Escuela de Ingeniería Ciencias y Sistemas, con fecha marzo 2019.

Randy Fernando Juárez Najarro

Guatemala, 28 de junio de 2019

Ing. Carlos Alfredo Azurdia Morales
Ingeniería en Ciencias y Sistemas
Facultad de Ingeniería
Universidad de San Carlos de Guatemala

Ingeniero Carlos Azurdia:

Por este medio me permito informarle que he procedido a revisar el trabajo de tesis titulado "PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA", elaborado por el estudiante Randy Fernando Juárez Najarro quien se identifica con carné 201404211, el cual, a mi criterio, cumple con los objetivos propuestos para su desarrollo y por tanto lo doy por aprobado.

Sin otro particular, me suscribo a usted,

Atentamente,

Ing. José Julio Pineda Chinchilla
Colegiado No. 10,340

José Julio Pineda Chinchilla
Ingeniero en Ciencias y Sistemas
Colegiado No. 10,340

Universidad San Carlos de Guatemala
Facultad de Ingeniería
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 17 de julio de 2019

Ingeniero
Carlos Gustavo Alonzo
Director de la Escuela de Ingeniería
En Ciencias y Sistemas

Respetable Ingeniero Alonzo:

Por este medio hago de su conocimiento que he revisado el trabajo de graduación del estudiante **RANDY FERNANDO JUÁREZ NAJARRO** con carné **201404211 y CUI 3001 03778 0101** titulado **"PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERIA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL INDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA"** y a mi criterio el mismo cumple con los objetivos propuestos para su desarrollo, según el protocolo aprobado.

Al agradecer su atención a la presente, aprovecho la oportunidad para suscribirme,

Atentamente,

Ing. Carlos Alfredo Azurdia
Coordinador de Privados
y Revisión de Trabajos de Graduación

ESCUELA DE INGENIERÍA EN CIENCIAS Y SISTEMAS

UNIVERSIDAD DE SAN CARLOS
DE GUATEMALA

FACULTAD DE INGENIERÌA
ESCUELA DE INGENIERÍA EN
CIENCIAS Y SISTEMAS
TEL: 24767644

*El Director de la Escuela de Ingeniería en Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer el dictamen del asesor con el visto bueno del revisor y del Licenciado en Letras, del trabajo de graduación "**PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS DE GUATEMALA**", realizado por el estudiante, RANDY FERNANDO JUÁREZ NAJARRO aprueba el presente trabajo y solicita la autorización del mismo.*

"ID Y ENSEÑAD A TODOS"

Ing. Carlos Gustavo Alonzo
Director
Escuela de Ingeniería en Ciencias y Sistemas

Guatemala, 11 de septiembre de 2019

Universidad de San Carlos
De Guatemala

Facultad de Ingeniería
Decanato

Ref. DTG.365.2019

La Decana de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala, luego de conocer la aprobación por parte del Director de la Escuela de Ingeniería en Ciencias y Sistemas, al trabajo de graduación titulado: **PRESENTACIÓN DE UN PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS DE LA UNIVERSIDAD DE SAN CARLOS GUATEMALA,** presentado por el estudiante universitario: **Randy Fernando Juárez Navarro,** y después de haber culminado las revisiones previas bajo la responsabilidad de las instancias correspondientes, se autoriza la impresión del mismo.

IMPRÍMASE.

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DECANA
FACULTAD DE INGENIERÍA

Inga. Aurelia Anabela Cordova Estrada
Decana

Guatemala, Septiembre de 2019

/cc

ACTO QUE DEDICO A:

Dios	Por darme el don de la vida.
Mis padres	Audie Juárez y Miriam Najarro. Por darme hasta la última gota de amor.
Mis hermanos	Audie, Steaven y Kevyn, Juárez Najarro, por ser las estrellas a las cuales sigo.
Mis tíos	Por estar atentos y brindarme su apoyo
Mis primos	Por los momentos muy gratos de niñez.

AGRADECIMIENTOS A:

Universidad de San Carlos de Guatemala	Por alojarme durante mi carrera y ser fuente de muchas experiencias.
Facultad de Ingeniería	Por ser el pilar de mi carrera y futuro.
Mis amigos	Oscar, Bryan, Pablo y Cindy. Por motivarme y apoyarme siempre.
Mis amigos de ciencias y sistemas	Por el apoyo mutuo para finalizar la carrera.
Mis amigos de la facultad	Por estar desde el inicio acompañándome.
Ingeniero asesor	Ingeniero José Julio Pineda Chinchilla, por no escatimar esfuerzo en sus enseñanzas.

ÍNDICE GENERAL

ÍNDICE DE ILUSTRACIONES

FIGURAS

TABLAS

LISTA DE SÍMBOLOS

Símbolo	Significado
csv	*Comma separated values*
cd	Disco compacto **(***Compact Disc*)
xml	*Extensible markup language*
m	Formato de archivo de Matlab
bsi	Formato estándar para almacenar datos
xslx	Formato de Microsoft Excel 2007
dat	Formato tipo de archivo *data*
gnu	*GNU's not Unix*
json	*JavaScript object notation*
url	*Universal resource locator*

GLOSARIO

Arquitectura de software	Es el diseño de alto nivel que modela la estructura de un sistema.
Canvas	Espacio en blanco que se utiliza para diseñar o diagramar libremente.
Cifrado	Proceso mediante el cual se transforma un texto en signos y solo se puede comprender si se conoce la clave.
Clúster	Unión de computadores mediante fibras de alta velocidad que tiene como objetivo aumentar el nivel de procesamiento al sumar los recursos y actuar como un solo servidor.
Correlación	Medida de dependencia que existe entre variables.
Evidencia	Certeza clara y manifiesta de una cosa, de tal forma que no puede ser negada bajo ningún punto de vista.
Función *Hash*	Función computable que tiene como entrada un conjunto de elementos y los transforma en cadenas de texto de longitud fija.

GNU General Public License	Es una licencia gratuita para software, que permite libertad de distribución y gratuidad a los usuarios. Es apoyada por Free Software Fundation, Inc.
Google Sheets	Herramienta de Google que permite la ejecución de hojas de cálculo en un navegador web.
Hardware	Conjunto de elementos materiales que constituyen el soporte físico de un computador.
Internet	Red informática de comunicación internacional que permite el intercambio de todo tipo de información entre sus usuarios.
Muestreo	Estudio de un número parcial de datos de un colectivo para deducir las características de la totalidad.
Muestra	Porción de un conjunto seleccionado por un método que permite considerar la muestra significativa.
Navegador web	Herramienta informática que permite a un usuario el acceso a internet.
Patrón	Modelo que sirve de muestra para sacar otro objeto igual.
Pedagogía	Ciencia que se ocupa de la educación y enseñanza.

Plugins	Complemento a una aplicación informática, para agregarle una función nueva.
Ranking	Clasificación de mayor a menor, útil para establecer criterios de valoración.
Remuestreo	Métodos estadísticos que permiten estimar la precisión de muestras y validar modelos mediante la validación cruzada.
Servidor	Software en ejecución, encargado de atender peticiones de clientes y resolver devolviendo una respuesta.
Sistema operativo	Programa o conjunto de programas que realizan funciones básicas y permiten el desarrollo de otros programas.
Sobreajuste	Problema que surge de utilizar el grado de libertad que aparece cuando hay un conjunto grande de hipótesis posibles, para encontrar regularidades poco significativas en los datos.
Software	Componentes lógicos que se ejecutan sobre un sistema informático físico y permiten efectuar tareas programadas.

Valor atípico	Valor que se aleja de manera disruptiva del comportamiento del modelo y, por lo tanto, puede corromperlo.

RESUMEN

A lo largo de la historia, la Universidad San Carlos de Guatemala se ha caracterizado por representar un papel importante para la sociedad. Aún en la actualidad es referente en muchos procesos políticos, científicos y sociales, además de contener a la mayor cantidad de estudiantes del país.

La Facultad de Ingeniería posee una gran representación debido a la importancia de la tecnología en el mundo. Para ingresar al resto de carreras que ofrece la universidad, todo aspirante debe realizar un proceso, en el cual se recolecta mucha información sobre los estudiantes.

La carrera de Ciencias y Sistemas de la Facultad de Ingeniería es la que tiene más estudiantes de nuevo ingreso, pero son pocos quienes logran egresar de manera eficiente. Muestra la mayor tasa de deserción estudiantil de la facultad. Según investigaciones de universidades como el Politécnico de Nueva Zelanda, es posible utilizar datos socioeconómicos y psicológicos de los estudiantes, como los recolectados en el proceso de ingreso, para impulsar investigaciones que reduzcan la tasa de deserción estudiantil.

Sin embargo, dados los peligros a los que se exponen a los estudiantes en estas investigaciones, es imperativo el uso de instrumentos legales y herramientas tecnológicas, en aras de resguardar su privacidad.

JUSTIFICACIÓN

La Universidad de San Carlos de Guatemala es la institución educativa de nivel superior con mayor impacto en la sociedad guatemalteca, debido al contexto histórico, la representación a nivel nacional, la importancia que ostenta internacionalmente y la gran cantidad de estudiantes que ingresan anualmente.

Cada año, una gran cantidad de estudiantes pasan por el procedimiento obligatorio para ingresar a la universidad. Inician con la evaluación vocacional; luego, una encuesta socioeconómica obligatoria. Finalmente, para cada estudiante se genera información de su desempeño académico.

Con la información mencionada es posible efectuar estudios estadísticos y análisis de correlaciones entre datos. Por la cantidad de datos que se proyectan, es necesario el auxilio de la tecnología, las ciencias de la computación y, más específicamente, la minería de datos.

Con el auxilio de dichos estudios estadísticos y la investigación es posible reducir el índice de deserción estudiantil, así como mejorar las estrategias pedagógicas con base en los factores psicológicos para potenciar el desempeño de los estudiantes, ahorrar recursos y priorizarlos.

OBJETIVOS

General

Crear un plan de implementación de una herramienta de minería de datos para determinar los factores clave que afectan el índice de éxito de cierre de pensum de estudiantes de ingeniería en ciencias y sistemas.

Específicos

1. Demostrar la importancia de la Universidad San Carlos de Guatemala para la sociedad, según su contexto histórico y actual.

2. Describir el número de estudiantes que desertan de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería.

3. Identificar las dependencias clave en el proceso de ingreso para los aspirantes y las de mayor impacto a lo largo de la carrera universitaria, así como los datos que recolectan en el proceso.

4. Proponer un plan de implementación en donde se consideren los factores de privacidad de la información y el marco legal al que se está sujeto.

5. Demostrar la factibilidad del plan de implementación, a través de una demostración, con datos reales, en la Escuela de Ciencias y Sistemas.

ALCANCE Y LIMITACIONES

En la investigación se propondrá un plan que toma en cuenta específicamente a la escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala.

Se presentará un plan que detalla la arquitectura del software y diagramas básicos del mismo, así como la elección de la herramienta de minería de datos que más se ajusta a las necesidades de la universidad. Además, la guía de instalación, costos, recursos humanos y una muestra a manera de ejemplo del análisis de propuesto.

Actualmente, el acceso a la información de las dependencias que manejan la información de los estudiantes de la Universidad San Carlos de Guatemala es limitado, dado que no existe un procedimiento estándar para la recolección y centralización de la información.

La cantidad de información que actualmente poseen las dependencias aún no es suficiente para generar un análisis con minería de datos confiable; por tanto, es importante implementar el plan para iniciar con el proceso de recolección de datos.

INTRODUCCIÓN

La Universidad de San Carlos de Guatemala es la única universidad estatal del país. Representa un papel importante para la sociedad y es subsidiada mayormente por el presupuesto nacional; por tanto, el mayor gasto económico es cubierto por el Estado y no directamente por los estudiantes.

Actualmente es la universidad más representativa de Guatemala, dado que alberga la mayor cantidad de estudiantes y tiene presencia tanto a nivel nacional como internacional. Por tal razón, los aspirantes deben cumplir algunos requisitos y realizar exitosamente un proceso para ingresar a la unidad académica de su elección.

A lo largo del proceso de ingreso y durante toda la carrera, los estudiantes generan una gran cantidad de información relacionada con sus características, intereses, estatus socioeconómico, desempeño académico y habilidades cognitivas, entre otras. Toda esta información es recolectada y almacenada por dependencias específicas de la universidad.

Algunas universidades internacionales han realizado investigaciones en donde se utilizan datos similares a los citados para mejorar el sistema educativo, implementar mejoras al pensum de estudios, detectar gastos innecesarios y ubicar a los estudiantes que necesitan un apoyo extra para culminar con éxito sus carreras académicas.

Estas universidades han utilizado herramientas derivadas de la inteligencia artificial. La mayoría ha aplicado la minería de datos para realizar los análisis. Esta es una rama de la inteligencia artificial que se especializa en algoritmos para determinar patrones en bancos de datos de gran tamaño, según explica Russell & Norvig (2004).

En la Universidad de San Carlos de Guatemala se genera gran cantidad de información por cada estudiante, por lo cual es viable y factible la aplicación de una herramienta de minería de datos que auxilie en el análisis de la misma.

La información que generan los estudiantes de la universidad recae en tres de las cuatro áreas más importantes de factores que, según Kuh, Bridges, & Hayek (2006) pueden determinar el éxito de un estudiante en su ámbito de estudios. Son el área económica, psicológica y cultural.

Al enlazar estos datos es posible crear un sistema que utilice toda la información generada por los estudiantes en las dependencias, para luego analizarla con la minería de datos y, de esta manera, mejorar la calidad del sistema educativo y reducir la deserción de los estudiantes.

1. CONTEXTO DE LA INVESTIGACIÓN

La Universidad de San Carlos de Guatemala es la única universidad estatal. En la actualidad posee 20 unidades académicas y tiene presencia en todo el país con 22 centros regionales, además de la sede central.

Al 2010, según datos de la UNESCO, poseía el 42 % de la población de estudiantes universitarios del total del país.

1.1. Reseña histórica de la Universidad San Carlos de Guatemala

Fue fundada el 31 de enero de 1676 por el monarca español Carlos II. Aunque la Real Cédula de Fundación llegó a Guatemala en octubre de ese mismo año, el primer año que se impartieron cátedras fue en 1681, con 60 estudiantes inscritos.

En 1700 se realizó la primera investigación en el campo de la medicina, a cargo del doctor en medicina Manuel Trinidad. Generó así un antecedente y con base en este fue fundada el área investigativa en la Universidad San Carlos de Guatemala.

Luego, en 1777, debido al terremoto que ocurrió en la ciudad de Santiago de los Caballeros de Guatemala (el cual dejó destruida la ciudad que en ese momento era la capital y ahora es conocida como la Antigua Guatemala), fue trasladada a Guatemala de la Asunción.

En 1823 la universidad, por medio de sus egresados, fue un pilar importante en la Primera Asamblea Constituyente Centroamericana; en 1825, en la primera Constitución Política del Estado de Guatemala y, en 1832, en la primera Declaración de Derechos Humanos en Guatemala.

Para 1898 se creó el desfile de la Huelga de Dolores, en donde se manifiesta el sentido de lucha contra la tiranía en Guatemala. Un acontecimiento de gran importancia fue la creación de la Facultad de Ingeniería, en 1882, junto a otras facultades que aún están vigentes.

El gobierno del general Jorge Ubico fue una época de represión y privación de libertades para la universidad y los estudiantes. Retiró el derecho a elegir autoridades máximas, ya que eran designadas por el presidente de la República.

Fue hasta el 9 de noviembre de 1944, luego de la Revolución del 20 de Octubre que, a través de una Junta Revolucionaria de Gobierno, se aprobó el decreto número 12. Este establece la autonomía de la Universidad de San Carlos de Guatemala.

La universidad, a lo largo de la historia, se ha caracterizado por tener presencia social que continúa en la actualidad, con gran importancia en los procesos sociales y políticos.

También representa el eje de desarrollo académico con mayor influencia; conserva su autonomía y es representativamente la mayor institución de estudios superiores en Guatemala.

Es por la importancia histórica y contemporánea de la universidad, que obtiene la mayoría de recursos para su funcionamiento de un porcentaje del Presupuesto General de Ingresos Ordinarios del Estado, que por ley le corresponde.

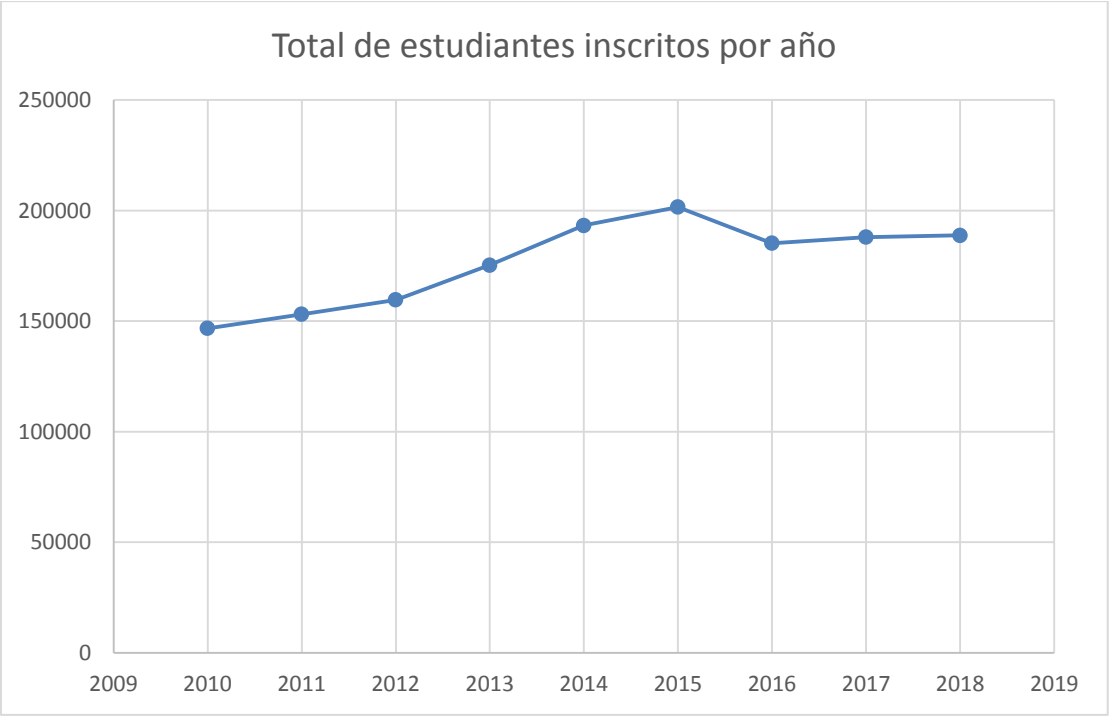
Para el año 2018, según el Departamento de Presupuestos de la USAC (2019), el total de ingresos fue de Q 2 214 626 002,00, del cual el 96 % provino del presupuesto nacional. La mayor parte de este se encuentra constituido por los impuestos que la población tributa al Estado de Guatemala, y es recolectado por la Superintendencia de Administración Tributaria.

Durante el 2018, la universidad operó con un déficit presupuestario de aproximadamente Q400 millones para funcionamiento, según publicación del periódico USAC (2018). En el año en curso se prevé que el presupuesto asignado no sea suficiente para cubrir las operaciones.

1.2. Población de estudiantes de la Universidad San Carlos de Guatemala

Durante los últimos años, la cantidad de estudiantes que se inscriben ha mostrado un comportamiento ascendente desde el 2010 hasta la actualidad, como se muestra en la figura 1.

Figura 1. **Cantidad de estudiantes inscritos por año (2010 a 2018)**



Fuente: Universidad San Carlos de Guatemala. Informe 2018. www.registro.usac.edu.gt. Elaboración propia. Consulta: 15 marzo de 2019.

En la USAC, en el año 2018, según el informe oficial de Registro y Estadística (2018) se inscribieron 23 795 de primer ingreso y 126 118 de reingreso (estudiantes ya inscritos), para un total 181 514 estudiantes activos. En 2017 se inscribieron 26 177 estudiantes de primer ingreso y 117 718 de reingreso, en total, 143 895 activos.

Los datos descritos son únicamente de estudiantes regulares, dado que son el principal interés para la investigación. Se excluye estudiantes pendientes de exámenes público y general, además de posgrado.

Figura 2. **Proceso de ingreso**

Fuente: Universidad San Carlos de Guatemala. Proceso de ingreso. www.usac.edu.gt. Consulta: diciembre de 2018.

1.3. Proceso de inscripción para estudiantes de primer ingreso

Toda persona aspirante a estudiar en la Universidad de San Carlos de Guatemala debe realizar una serie de pasos, como se muestra en la figura 2. Durante el proceso para que un estudiante sea oficialmente inscrito, existen diferentes dependencias encargadas de cada etapa.

El primer paso que un estudiante debe realizar para iniciar su proceso de inscripción es la prueba de orientación vocacional en Bienestar Estudiantil. Consiste en una evaluación psicológica y pedagógica del nivel de habilidades de los estudiantes en las diferentes inteligencias, además de la identificación de los intereses profesionales con base a las carreras que existen en la universidad.

El segundo paso son las pruebas de conocimientos básicos (es necesario que ya tenga su carné de orientación vocacional). La evaluación está a cargo del Sistema de Ubicación y Nivelación. Varía el tipo de evaluación según la carrera a la que el estudiante aspire.

El tercer paso son las pruebas específicas (es necesario que tenga resultado satisfactorio en las pruebas básicas requeridas), las cuales son realizadas por cada unidad académica. En ellas se evalúan habilidades y conocimientos específicos mínimos que los estudiantes debe tener, según su carrera.

Por último, como cuarto paso, luego que el estudiante obtiene resultados satisfactorios de las pruebas específicas, la institución encargada de inscribir a los estudiantes es Registro y Estadística. Allí, los estudiantes que llenan los

requisitos obtienen un número de registro académico, el cual los identifica como estudiantes de la unidad académica correspondiente.

1.4. Dependencias involucradas en el proceso de inscripción de un estudiante de primer ingreso

En el proceso que los aspirantes deben completar para estar formalmente inscritos en una unidad académica, las diversas dependencias de la universidad que intervienen son:

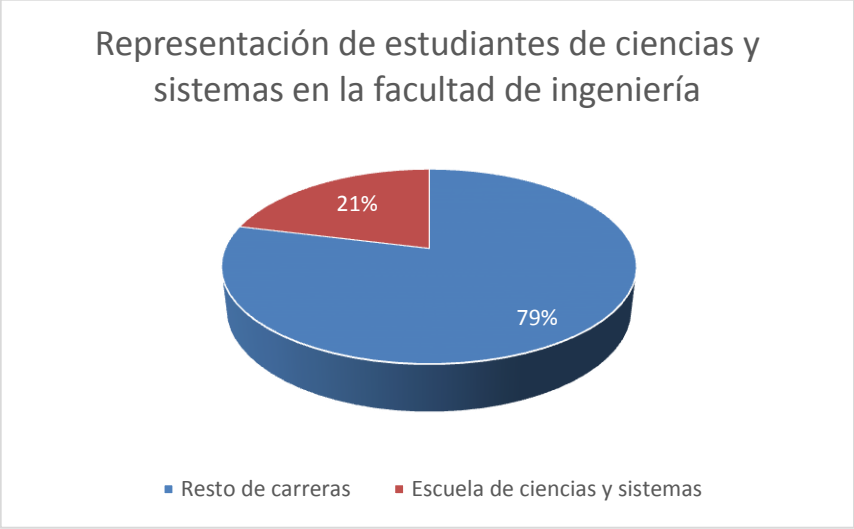
- Bienestar Estudiantil
  - Brinda orientación vocacional a los aspirantes a estudiante en la Universidad de San Carlos de Guatemala, sede central y extensiones departamentales (USAC, 2014).
- Sistema de Ubicación y Nivelación
  - Se encarga de definir el perfil cognitivo básico que los aspirantes deben tener para ingresar a las diferentes unidades académicas y evaluar como satisfactorio o insatisfactorio (USAC, 2017).
- Registro y Estadística
  - Centraliza todos los procesos estadísticos en relación con los estudiantes (Registro y Estadística, 2013).
- Unidad académica
  - Define el pensum de estudios, realizar tareas administrativas y acompaña a los estudiantes como responsable hasta la culminación de los estudios profesionales.
- Centro de Cálculo
  - Se encarga del registro y administración de los datos de los estudiantes de cada unidad académica.

1.5. Situación de los estudiantes de Ingeniería en Ciencias y Sistemas

La facultad de Ingeniería posee seis edificios en donde se imparte clases magistrales y laboratorios prácticos para 10 escuelas, entre las cuales está la de Ciencias y Sistemas. Se encarga de acompañar a los estudiantes durante todo el desarrollo del área profesional de la carrera con el mismo nombre, según se describe en la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala (2014).

Al 2018 posee 13 208 estudiantes activos, que representa el 7 % del total de toda la población estudiantil de la universidad; son 1 347 de primer ingreso y el resto de reingreso. La escuela de Ciencias y Sistemas actualmente tiene una población de 2 789 estudiantes, que representa el 21 % del total de la facultad de Ingeniería.

Figura 3. **Gráfica de porcentaje de estudiantes de Ciencias y Sistemas**



Fuente: Facultad de ingeniería. Registro y Estadística. Elaboración propia.
Consulta: mayo de 2019.

La población de Ciencias y Sistemas es la segunda más grande dentro de la facultad de Ingeniería, superada únicamente por la de Ingeniería Industrial. El dato más importante es que ingeniería en Ciencias y Sistemas posee la mayor cantidad de estudiantes de primer ingreso, con 443 en el 2018, lo cual representa el 33 %. Casi triplica las otras escuelas y es la población que más aumenta actualmente, según datos de Registro y Estadística (2018).

En contraste con los datos anteriores, la cantidad de estudiantes que se gradúan de la carrera de Ingeniería en Ciencias y Sistemas es reducida en relación a los estudiantes inscritos anualmente de primer ingreso, tal como se muestra en la tabla 1.

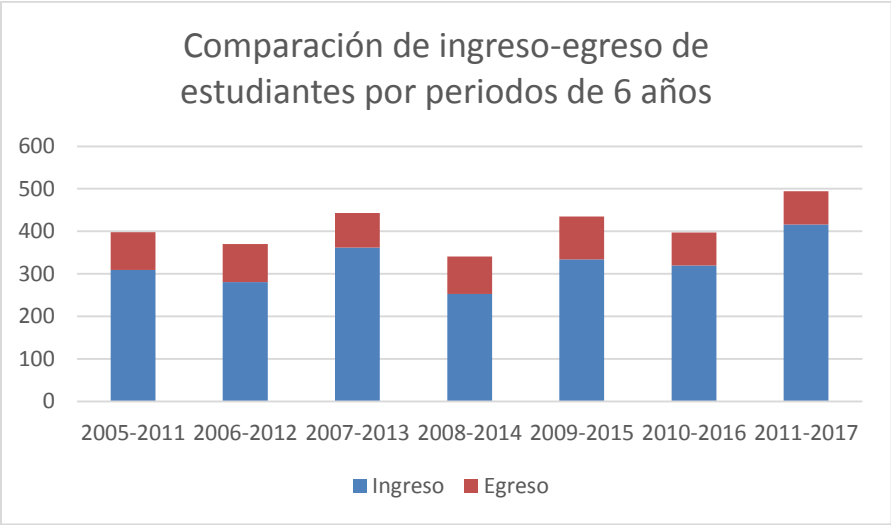
Tabla I. **Porcentaje ingreso-egreso Ciencias y Sistemas**

Año ingreso-egreso	Ingreso	Egreso	Porcentaje
2005-2011	309	89	29 %
2006-2012	281	89	32 %
2007-2013	362	81	22 %
2008-2014	253	88	35 %
2009-2015	334	101	30 %
2010-2016	320	77	24 %
2011-2017	416	78	19 %

Fuente: Facultad de ingeniería. Centro de Cálculo. Elaboración propia.
Consulta: mayo de 2019.

La carrera de Ingeniería en Ciencias y Sistemas tiene una duración curricular de 5 años divididos en dos semestres cada uno. Para optar al título de licenciado se agrega un año, debido al examen privado, público y el trabajo de investigación como primera opción o el ejercicio profesional supervisado. Se completan así seis años mínimo para egresar formalmente.

Figura 4. **Gráfica comparativa de ingreso-egreso Ciencias y Sistemas**



Fuente: Facultad de ingeniería. Centro de Cálculo. Elaboración propia.
Consulta: mayo de 2019.

Debido a la elevada cantidad de estudiantes inscritos en Ciencias y Sistemas, al comportamiento ascendente de estudiantes inscritos y la alta deserción, como se muestra en la gráfica, surge la necesidad de hacer un planteamiento investigativo para utilizar y aprovechar los datos que se generan en las entidades de la universidad sobre esta población. El objetivo principal es auxiliar a las autoridades en la detección temprana de los alumnos con más riesgo de no finalizar con éxito sus estudios universitarios, y de esta manera brindarles apoyo extra para evitar la deserción.

2. PROCESO DE INSCRIPCIÓN Y REINSCRIPCIÓN DE LOS ESTUDIANTES

2.1. Proceso de inscripción para estudiantes de primer ingreso

Los aspirantes a estudiantes deben completar una lista de pasos. Como requisito inicial, deben haber completado una carrera a nivel medio y tener un título que lo acredite, que esté avalado por la Contraloría General de Cuentas.

La universidad acepta estudiantes de todo el mundo, pero únicamente brinda subsidio a los estudiantes de origen guatemalteco, debido a que son los impuestos que paga el pueblo los que conforman la mayor parte del presupuesto de funcionamiento de la universidad, tal como se detalla en el subcapítulo 1.1. Los extranjeros deben pagar los gastos completos según la carrera académica que decidan cursar.

Luego de cumplir con los requisitos anteriores, los aspirantes deben completar una serie de pasos según la carrera académica, como se muestra en el subcapítulo 1.3.

Los pasos están a cargo de diferentes dependencias de la universidad, encargadas de guiar al aspirante durante el proceso, indicarle los requisitos para iniciar y finalizar cada paso, así como la papelería necesaria que necesita completar.

2.1.1. Prueba de orientación vocacional

El primer paso para ingresar a la universidad, luego de completar los requisitos iniciales, es la evaluación de orientación vocacional.

Bienestar Estudiantil es el encargado de proveer el servicio de orientación vocacional a los aspirantes a la Universidad de San Carlos de Guatemala en la sede central y en las extensiones departamentales. Los pasos para la asignación de la evaluación se pueden encontrar en el enlace siguiente: https://vocacional.usac.edu.gt/conte_documentos_a_presentar.php.

La evaluación está dividida en dos partes, para un total de 4 pruebas:

- Prueba de habilidades
  - Lógico-matemática
  - Verbal
  - Abstracta
- Intereses profesionales de estudio

Para la prueba de habilidades, los resultados son dados en cuartiles, divididos en bajo, medio-bajo, medio-alto y alto. Para la prueba de intereses profesionales de estudio los resultados son dados en percentiles, divididos en las 21 unidades académicas de la universidad.

Esta evaluación no puede ser reprobada, dado que es únicamente un instrumento para orientar al aspirante en la elección de la carrera académica.

2.1.2. Pruebas de conocimientos básicos

El segundo paso son las pruebas de conocimientos básicos. Son instrumentos para medir los conocimientos mínimos necesarios en cada unidad académica.

El Sistema de Ubicación y Nivelación (SUN) es la dependencia encargada de realizar las pruebas a todos los aspirantes en la sede central y en las extensiones departamentales.

También brinda al Ministerio de Educación las tendencias de los resultados con la intención de corregir las deficiencias en los estudiantes de nivel medio, a nivel nacional. Los pasos para la asignación de la prueba de conocimientos básicos están en el enlace http://nuevos.usac.edu.gt/proceso_pcb.html.

Las pruebas de conocimientos básicos han sido realizadas desde el año 2000; fueron estandarizadas el 2003 y desde el 2009 tiene una vigencia indefinida.

La prueba está estructurada en una serie de preguntas de respuesta múltiple que permite detectar si el aspirante posee un nivel mínimo satisfactorio para las competencias que la unidad académica requiere. Por tanto, el resultado de las pruebas puede ser satisfactorio o insatisfactorio.

Las pruebas están divididas en las siguientes asignaturas:

- Biología
- Física
- Lenguaje
- Matemática
- Química

Actualmente, existe un simulador para que los aspirantes conozcan la estructura y forma de las pruebas. Para ingresar deben contar con un número de orientación vocacional y una clave de acceso, dados por Bienestar Estudiantil en el paso anterior. Se puede ingresar al simulador en el enlace http://sitios2.usac.edu.gt/simuladorpcb/.

2.1.3. Pruebas de conocimientos específicos

La última prueba que los aspirantes deben realizar es la de conocimientos específicos. Su fin es determinar si tienen el perfil mínimo adecuado para iniciar y culminar la carrera académica.

La prueba es realizada por cada unidad académica y en cada uno de los centros universitarios; cada una es exclusiva y distinta; por lo tanto, no existe una estructura estándar para las mismas.

Cada facultad y escuela tiene sus propios mecanismos para evaluar los conocimientos específicos. Los pasos para la asignación de las pruebas específicas para ingresar a la carrera de Ciencias y Sistemas de la facultad de Ingeniería, se encuentran en el enlace siguiente: https://primeringreso.ingenieria.usac.edu.gt/iniciodeespecificas.

Las pruebas para ingresar a esta facultad son dos:

- Prueba específica de computación
- Pruebas específicas de matemática

Los temarios y guías de estudio pueden ser encontrados en el mismo enlace, para auxiliar al aspirante en la preparación, además del Programa Académico Preparatorio.

2.1.4. Inscripción

Luego de cumplir con los tres pasos anteriores, el aspirante ya puede realizar el proceso de inscripción para ser formalmente estudiante de la universidad y de la unidad académica correspondiente.

El departamento de Registro y Estadística es la dependencia encargada del proceso de inscripción. Los pasos para realizarla se encuentran en el enlace http://registro.usac.edu.gt/index.php?ping=9. Luego de dicho proceso, Registro y Estadística genera un registro académico para cada nuevo estudiante, con lo cual culmina el proceso de inscripción en la unidad académica correspondiente.

Otra tarea importante de la dependencia es utilizar los datos recolectados para investigaciones o procedimientos estadísticos. Actualmente realiza anualmente estudios estadísticos tales como reportes demográficos, étnicos, sociales y económicos, entre otros.

2.1.5. Reinscripción

Cuando los estudiantes ya cursaron el primer y segundo semestre en el primer año de la carrera académica, Registro y Estadística realiza un proceso de reinscripción para que los estudiantes sigan activos el siguiente año, y así sucesivamente.

Como parte del proceso, los estudiantes deben pagar la matrícula anual, llenar un cuestionario socioeconómico y reinscribirse en el enlace https://registro.usac.edu.gt/index.php?ring=4, durante las fechas asignadas.

Este proceso es realizado anualmente mientras el estudiante está cursando la carrera. Es obligatorio para que puedan asignarse a los cursos correspondientes durante el año y tengan validez.

Mientras los estudiantes cursan su carrera, existe una dependencia en cada facultad que se encarga de llevar el control y dar validez a los cursos y notas. En la facultad de Ingeniería, es el Centro de Cálculo.

Este centro es donde los catedráticos, cada semestre, cargan las notas de los cursos impartidos. Allí también se realizan procesos correctivos, permisos especiales y estadísticas de los cursos de los estudiantes.

3. INVESTIGACIONES REALIZADAS EN UNIVERSIDADES CON SITUACIONES ANÁLOGAS

En los últimos años, la utilización de los datos para generar conocimiento y mejorar procesos y sistemas ha tomado relevancia. Muchas instituciones internacionales, especialmente educativas, ha realizado investigaciones orientadas a la utilización de los datos para mejorar los procesos internos, con técnicas y herramientas modernas de las ciencias de la computación, tal como lo realizo el Politécnico de Nueva Zelanda, según explica Kovačić (2010).

3.1. Contexto de la investigación realizada en el Politécnico de Nueva Zelanda

Para las instituciones educativas, uno de los temas prioritarios es el éxito académico de sus estudiantes; es decir, la finalización en el tiempo esperado. Es importante porque los alumnos que finalizan su carrera universitaria tienen una mayor probabilidad de ser más exitosos en el ambiente laboral y, consecuentemente, ser mejor remunerados.

Según estadísticas de la investigación del 2010 realizada en el Politécnico de Nueva Zelanda, la retención de estudiantes y la consecuente finalización de sus estudios universitarios muestra una tendencia a la baja. Para esta casa de estudios, es relevante debido a que el Estado brinda subsidio con base en el porcentaje de alumnos que finaliza exitosamente su carrera universitaria.

Por tales factores, el personal de la universidad estableció como prioridad elevar el porcentaje de alumnos que finaliza con éxito. Por tanto, se realizó una investigación para detectar los factores que identifican a los alumnos con tendencia a la deserción, para brindarles un apoyo extra antes de que ocurra.

3.2. Datos relevantes de la investigación

La investigación mencionada estuvo basada en investigaciones previas realizadas en el Reino Unido en la Universidad Simpson y en los Estados Unidos de América en la Universidad de Arizona. En ambos casos, se determinaron exitosamente los factores implicados en el éxito de los estudiantes con base en una serie de características previas.

Con este fundamento, la universidad vio una oportunidad en los datos que se recolectan de los alumnos nuevos al ingresar al centro de estudios en su primer año.

Para ingresar al centro de estudios, los estudiantes deben completar, entre otros requisitos, formularios con datos tales como las características sociodemográficas como la edad, el género, la etnia, la educación, el trabajo, el estatus económico y si tienen discapacidades.

Otra consideración fue la tecnología empleada para manejar los datos de la mejor manera. Según las investigaciones previas, la herramienta que favorece este tipo de análisis es la minería de datos.

3.3. Solución propuesta de la investigación

Con base en las investigaciones previas en el tema realizadas por otras universidades y a partir de la oportunidad que ofrece la información recolectada de los estudiantes de primer ingreso, se realizó la investigación con diferentes modelos para análisis, derivados de la minería de datos.

Como variable dependiente fue tomada la nota obtenida en el curso más importante del primer semestre, Sistemas de Información, para tener un punto de referencia medible cuantitativamente.

El proceso consistió en aplicar diferentes modelos de la minería de datos para detectar la correlación entre los datos recolectados de los estudiantes de primer ingreso y las notas obtenidas en el curso mencionado.

3.4. Resultados obtenidos

Según los resultados generados por la investigación, las conclusiones a las que llegaron son las siguientes:

- Es posible detectar con antelación a los estudiantes con más riesgo de deserción.
- Los datos más significativos para separar a los alumnos exitosos de los no exitosos fueron el género, la etnia y la edad.
- Con más información recolectada, el análisis eleva su exactitud.

3.5. Relación con la Universidad San Carlos de Guatemala

La Universidad de San Carlos recolecta gran cantidad de información por medio de sus dependencias, tal como muestra el proceso de ingreso de los estudiantes, descrito en segundo capítulo.

Actualmente, la información recolectada no es utilizada al máximo. Como se dijo, incluye datos como la edad, género, etnia, educación, trabajo, estatus económico y discapacidades.

Como se plantea en la investigación realizada en el Politécnico de Nueva Zelanda, los datos que la universidad recolecta mediante sus dependencias pueden ser utilizados para diversos estudios y, de esta manera, auxiliar a las autoridades en la detección de estudiantes con mayor riesgo de deserción.

Los datos que las instituciones recopilan y generan de los estudiantes puede tener un mayor uso que el que actualmente se le da, por medio de técnicas y herramientas que la tecnología actual brinda.

También es posible relacionar entre sí las instituciones y los datos, con la finalidad de deducir comportamientos y validar hipótesis para mejorar el sistema educativo de la universidad y reducir la deserción de los estudiantes.

3.6. Datos relevantes en las dependencias para la investigación

Según la investigación citada, realizada en el Politécnico de Nueva Zelanda, y otras relacionadas como la realizada por Kuh, Bridges & Hayek (2006), los datos más relevantes que pueden determinar el éxito de un estudiante en sus estudios son el ámbito económico, psicológico y cultural.

Además, es importante definir un parámetro que sea utilizado para determinar cuantitativamente el éxito de los estudiantes en la carrera académica, como las notas en un curso específico, promedio, cantidad de cursos perdidos, entre otros.

3.6.1. Datos del ámbito psicológico de los estudiantes

En este ámbito, la herramienta que más se ajusta al contexto de la universidad es la evaluación de orientación vocacional, tal como se describe en el capítulo 1.

Esta evaluación se realiza desde 2002 a todos los aspirantes, pero no fue sino hasta el 2008 que se estandarizaron las evaluaciones a la forma que aún conservan. Tiene vigencia indefinida.

Actualmente, la prueba es realizada con un folleto para las preguntas y un cuadernillo de selección múltiple para las respuestas.

Tomando como base el año 2008, la evaluación de orientación vocacional realizada durante 10 años ha acumulado aproximadamente 25 millones de registros, y suma cada año más de 3 millones de registros nuevos.

3.6.2. Datos del ámbito cultural y económico de los estudiantes

Tanto para el ámbito cultural como el económico, la herramienta que más se ajusta es el cuestionario socioeconómico, descrito en el capítulo 1. El primero se ha realizado desde el año 2010 a todos los estudiantes de reingreso, tanto a nivel de licenciatura como de postgrado.

Actualmente, las preguntas de los cuestionarios son manejadas con un software de encuestas. Las respuestas son almacenadas en una base de datos administrada por los ingenieros del centro de tecnología del departamento de Registro y Estadística.

El cuestionario socioeconómico está constituido aproximadamente por 175 preguntas que cada estudiante de reingreso debe contestar. Durante 8 años se han acumulado aproximadamente 215 millones de registros, y cada año suma más de 35 millones de registros nuevos.

3.6.3. Parámetro clave en el éxito de los estudiantes

Para evaluar el índice de éxito de los estudiantes, los parámetros más determinantes son recolectados por el Centro de Cálculo. En este se tiene la información académica de cada estudiante. Los datos más relevantes son los cursos aprobados, créditos obtenidos, cursos repetidos, cursos perdidos y promedio.

Con base en los datos mencionados se define el éxito de un estudiante de Ingeniería en Ciencias y Sistemas como una composición de cuatro factores. La fórmula aplicada para obtener el índice es la siguiente:

$$\left[\frac{\text{promedio}}{100} * 0{,}4 + \frac{\text{créditos}}{250} * 0{,}3 + \frac{\text{cAprobados}}{60} * 0{,}3 - \frac{\text{cRepedidos}}{\text{cAprobados}} * 0{,}05\right] * 100$$

Anualmente, el Centro de Cálculo realiza actualizaciones al sistema con las notas de los estudiantes durante el semestre y la escuela de vacaciones. Para la escuela de Ciencias y Sistemas se actualiza la información de 4 000 estudiantes aproximadamente, tomando en cuenta todos los cursos asignados de cada estudiante.

La cantidad de estudiantes de Ingeniería en Ciencias y Sistemas, tal como se mostró en el primer capítulo, describe un comportamiento ascendente más acelerado que el resto de las carreras; por lo tanto, la cantidad de información tiende al aumento.

3.6.4. Relaciones entre dependencias

El estado actual de la administración de la información en las tres dependencias descritas, limita establecer una estructura que facilite el flujo de los datos y sea posible recopilar, almacenar, relacionar y consolidar los datos que cada dependencia genera.

Para reestructurar la administración de los datos y facilitar la cooperación de las dependencias, es necesario que se genere una normativa en la Ley Orgánica de la Universidad de San Carlos de Guatemala, en donde se establezca que estas dependencias deben cooperar con la cargar los datos anualmente a un sistema centralizado.

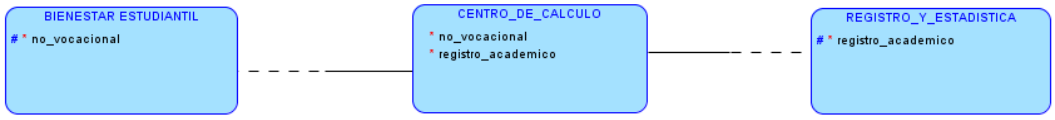
3.6.5. Relación entre los datos

Para relacionar los datos de las tres dependencias; es decir, el cuestionario socioeconómico, los resultados de examen vocacional y los datos académicos de los estudiantes, se debe enlazar con base en un registro único que identifique al estudiante en las tres instituciones.

Después de analizar la organización de los datos en cada dependencia se determinó que no es posible utilizar un solo registro que identifique a los estudiantes: en Bienestar Estudiantil se les asigna un número de orientación vocacional, pero queda sin utilidad cuando los estudiantes se inscriben en Registro y Estadística, ya que se les asigna como única identificación el registro académico.

Dada la organización de los datos, se determinó que las dependencias relacionan a los estudiantes mediante llaves foráneas como el número de orientación vocacional y el registro académico, como se muestra en la figura 5, para poder relacionar los datos.

Figura 5. **Relación lógica entre dependencias**



Fuente: elaboración propia.

Tabla II. **Relación de datos ejemplo entre dependencias**

Bienestar Estudiantil	**Centro de Cálculo**		**Registro y Estadística**
Núm. vocacional	**Núm. vocacional**	**Núm. registro**	**Núm. registro**
2014000000	2014000000	201500000	201500000
2015000000	2015000000	201600000	201600000
2016000000	2016000000	201700000	201700000
2017000000	2017000000	201800000	201800000

Fuente: elaboración propia.

4. MARCO LEGAL APLICABLE AL MANEJO DE LA INFORMACIÓN DE LOS ESTUDIANTES

4.1. Importancia de la protección de la información

Debido a que la información que se plantea recolectar para implementar el plan que propone esta investigación contiene datos de carácter sensible de los estudiantes, se define un proceso para proteger estos datos. Estos pueden ser mal utilizados tanto por personal interno como entidades externas sin autorización.

Teniendo en cuenta que el fin de la investigación es determinar un comportamiento global como grupo (y no identificar a los individuos) con características compartidas, los datos sensibles relevantes en esta investigación identificados son dos: el registro académico y el número de orientación vocacional, debido a que exponen la identidad de los estudiantes.

La principal responsabilidad de la universidad y las personas que implementen el sistema, es proteger la privacidad de los estudiantes. Por tanto, en el plan se debe definir un método efectivo que asegure la privacidad.

4.2. Anonimización como método de protección a la privacidad de las personas

La anonimización es el proceso por el cual se limita e impide que los datos de estudio sean relacionados y vinculados con una identificación individual, en donde se pone en explícito la identidad de la persona, explica el Ministerio de Salud Colombiana (2016).

En la actualidad existen muchos casos, tanto a nivel nacional como internacional, en los cuales la falta de protección de los datos ha afectado la privacidad de las personas y, en consecuencia, también ha afectado a las organizaciones y responsables a nivel económico, social, público y legal.

4.2.1. Marco legal nacional

En Guatemala actualmente no existe un marco legal que contemple delitos informáticos. Las organizaciones que se ven afectadas por algún ataque cibernético carecen de instrumentos legales para actuar en su defensa; por lo tanto, recurren a otros recursos legales para auxiliarse.

4.2.1.1. Caso Infornet, Digidata y Trans Union

El 10 de febrero de 2015, la Corte de Constitucionalidad confirmó la sentencia en relación con tres empresas guatemaltecas, Infornet, Digidata y Trans Union Guatemala, en la cual se establece que la información personal no puede ser comercializada sin el consentimiento voluntario de la persona.

El amparo fue puesto por el Procurador de los Derechos Humanos, Jorge De León Duque. La Corte de Constitucionalidad dictaminó:

“Normas violadas: citó los artículos 1º, 2º, 4º, 5º, 12, 14, 30, 31, 43, 44, 46, 47, 51, 53, 101, 102, 152 y 153 de la Constitución Política de la República de Guatemala; 12 de la Declaración Universal de Derechos Humanos.” Corte de Constitucionalidad (2015)

Dado que Guatemala actualmente no tiene un marco legal robusto para tomar acción o regular las medidas y precauciones que se deben tener en cuenta en el manejo de los datos sensibles de las personas y la protección de su privacidad, es importante respaldarse en el contexto internacional.

4.2.2. Marco legal internacional

Guatemala está sujeta a tratados y convenios internacionales con organizaciones mundiales, como la Organización de las Naciones Unidas y la Declaración Universal de Derechos Humanos. Así lo expone el dictamen de la Corte de Constitucionalidad en el caso de Infornet, Digidata y Trans Union Guatemala.

Distintas organizaciones a nivel internacional definen el proceso para la anonimización como la protección de la privacidad de los datos y el uso de los mismos exclusivamente para el propósito estadístico.

Como primera referencia, la División de Estadísticas de Naciones Unidas explica:

“Los datos que reúnan los organismos de estadística para la compilación estadística, ya sea que se refieran a personas naturales o jurídicas, deben ser estrictamente confidenciales y utilizarse exclusivamente para fines estadísticos.” Ministerio de Salud Colombia (2016)

Por otro lado, el Parlamento Europeo y el Consejo de la Unión Europea contemplan la privacidad de la información sensible de las personas en medios electrónicos:

“El presente Reglamento protege los derechos y libertades fundamentales de las personas físicas y, en particular, su derecho a la protección de los datos personales.” Reglamento general de protección de datos.” (2018, art. 1)

El 25 de mayo de 2018, entró en vigor en la Unión Europea el reglamento general de protección de datos, según la Comisión Europea (2018). En él se regula a las empresas para que quienes utilicen sus servicios conserven el control sobre su información.

4.2.2.1. Caso Facebook

En abril del 2018, abogados británicos y estadounidenses iniciaron un proceso legal en contra de Facebook, Cambridge Analytica, Global Science Research Limited y SCL Group Limited debido al uso de información personal sensible de un millón de británicos y 70 millones de estadounidenses.

Explica Bowcott & Hern (2018) que en el caso se plantea que la información fue proporcionada por Facebook a Cambridge Analytica y fue utilizada para desarrollar campañas de propaganda política.

En el Reino Unido la información se utilizó para el referéndum de la Unión Europea y en Estados Unidos de América fue usada en la campaña política de 2016 por el jefe de campaña Steve Bannon, de Donald Trump.

Entre la información que fue extraída se incluye nombres, apellidos, números de teléfono, correos electrónicos, direcciones, afiliaciones políticas y religiosas. El resultado fue que Facebook se comprometió a notificar a todas las personas que fueron afectados por la extracción de datos.

En el Reino Unido le fue impuesta una multa en octubre del 2018 por 500 000 libras. En Estados Unidos de América a la fecha aún no se ha resuelto la multa que se impondrá a Facebook por la violación a la privacidad de la información de los usuarios.

4.3. Consideraciones importantes para el plan

Con base en el marco legal nacional e internacional planteado como referencia para la correcta anonimización de datos en esta investigación, se debe realizar un proceso con las tres fuentes de datos y mantener la relación entre cada una.

En la recopilación de la información de los estudiantes, los datos sensibles que necesitan un proceso de anonimización son los dos ya mencionados: el registro académico y el número de orientación vocacional. Estos identifican a los estudiantes y pueden ser relacionados con los datos recopilados.

5. MARCO TEÓRICO

> En el estudio de la inteligencia artificial han influido diversas ramas de la ciencia, biología, psicología, matemática, neurología, estadística, economía, ciencia de la información y computación. Cada una ha aportado diversas técnicas y ayudado a la compresión de cómo funciona el cerebro humano, la base de la inteligencia artificial. Han generando teorías que en conjunto la han desarrollado como una rama de la ciencia e impactado en el uso de la tecnología para su aplicación.[1]

5.1. Inteligencia artificial

La inteligencia artificial es un campo de la ciencia donde se estudia el desarrollo de máquinas con capacidad para tomar decisiones y actuar de acuerdo con las decisiones tomadas.

5.1.1. Historia

En 1950, Allan Turing inició el desarrollo de la inteligencia artificial como una rama de la ciencia. La definió a través de la Prueba de Turing, basada en una entrevista entre una máquina y un ser humano. “Si el humano no es capaz de determinar que habla con una máquina, entonces la máquina es realmente inteligente”.[2]

En 1958, John McCarthy desarrolló el lenguaje Lisp, el cual se convirtió en el lenguaje dominante para la inteligencia artificial.

[1] RUSSELL, Stuard, y NORVIG, Peter. *Inteligencia Artificial Un Enfoque Moderno.* p. 562.
[2] Ibíd.

Luego, en 1961, Allen Newel y Herbert Simon desarrollaron la ciencia cognitiva. Les interesó el proceso de razonamiento humano y su aplicación a una máquina que siguiera el mismo patrón. Así surgió el Sistema de Resolución General de Problemas (SRGP). Después, en 1976, con base en este, desarrollaron la hipótesis del sistema de símbolos físicos.

Actualmente una máquina es considerada poseedora de inteligencia artificial si está dotada de las siguientes capacidades:

- Razonamiento automático
- Procesamiento de lenguaje natural
- Aprendizaje automático
- Representación del conocimiento

5.1.2. Los dos enfoques de la inteligencia artificial

La inteligencia artificial, como cualquier otra rama de la ciencia, aloja dos grandes corrientes de pensamiento que se han enfrentado a lo largo de la historia, detallan estos enfoques y los subdividen en dos, con base en la acción y el proceso mental.

El primer enfoque encapsula los sistemas que piensan como humanos; se basa en la automatización de decisiones y resolución de problemas como lo haría un humano. En la misma línea están los sistemas que actúan como humanos. Se definen como máquinas que desarrollan funciones aplicando inteligencia.

El segundo enfoque se orienta a los sistemas que piensan racionalmente; es decir, que utilizan cálculos matemáticos para razonar. Para finalizar se definen los sistemas que actúan racionalmente como artefactos que expresan conductas inteligentes.

5.1.3. Aplicaciones notables

Google es una empresa que se ha convertido en uno de los líderes en la aplicación de inteligencia artificial en la vida cotidiana. El asistente de traducción de lenguaje ha sido recreado en una red neuronal, conocida como Google Neural Machine Translation System o GNMT.

GNMT ofrece una manera diferente para traducir los textos. A diferencia de los traductores típicos que lo hacen palabra por palabra, GNMT toma el texto holísticamente y lo introduce en la red neuronal.

El equipo Google Brain resalta que en el estadío inicial del traductor, los resultados que generaba eran, en promedio, iguales a las otras técnicas. Sin embargo, luego de dos años de entrenar la red neuronal, se incrementó el desempeño, la velocidad y exactitud de las traducciones, y sobrepasó a su competencia, como se muestra en el anexo 1.

5.2. Minería de datos

La minería de datos es vista y utilizada como una forma de análisis exploratorio en el cual se extraen patrones y relaciones de una inmensa cantidad de datos en un proceso automático.

5.2.1. Historia

La minería de datos es una subárea de dos ciencias principales, la estadística y la inteligencia artificial, específicamente del aprendizaje automático.

En 1925 se subdividió la estadística en tres ramas principales: la primera se encarga del estudio poblacional; la segunda se enfoca en la variabilidad y modelización y la tercera engloba los métodos de síntesis de datos.

En 1937, Fisher trabajó en el análisis discriminante para la predicción de una variable nominal. Luego, en 1991, Friedman construyó el método de regresión adaptativa multivariante.

Actualmente se divide en dos grandes áreas:

- Estadística inferencial, encargada del modelamiento de datos.
- Estadística exploratoria, enfocada en el análisis de los datos.

5.2.2. El método de la minería de datos

En la mayoría de los procedimientos donde se aplica la minería de datos se utiliza la validación cruzada, debido al sobreajuste al que tiende un análisis

exploratorio sobre un conjunto de datos. Luego que se realiza el procedimiento sobre varios conjuntos de datos de prueba, el modelo es revisado y ajustado.

Existe un problema denominado indeterminación de la teoría con base en los datos, en donde existe más de un modelo que se ajusta al conjunto de datos. Para resolverlo se emplean los algoritmos genéticos, que generan un conjunto de ecuaciones estimadas para los datos. Luego, de manera aleatoria, se combinan para crear una nueva generación de ecuaciones y mejoran la calidad de las predicciones.

5.2.3. Técnicas de la minería de datos

A continuación, se presentan las técnicas que se utilizan en la aplicación de minería de datos:

- Escalabilidad
- Algoritmos genéticos
- Árboles de decisión
- Redes neuronales
- Redes bayesianas
- Análisis factorial descriptivo

5.2.4. Algoritmos más utilizados

En un estudio realizado por la Universidad de Vermont, se determinó una lista de los algoritmos más utilizados en la minería de datos. Son los siguientes, ordenados de mayor a menor relevancia:

- C4.5
- El algoritmo *k-means*
- Máquinas de soporte vectorial
- El algoritmo *Apriori*
- El algoritmo expectativa-maximización
- Ranking de páginas
- El algoritmo *AdaBoost*
- El algoritmo *k-nearest neighbor*

5.2.5. Descripción del algoritmo Relief

El algoritmo de Relief fue diseñado y propuesto por Kira y Rendell para determinar pesos de los atributos y su importancia en el conjunto de datos en relación con la variable dependiente.

Básicamente, el algoritmo Relief construye un vector de pesos para cada atributo. Utiliza como base el algoritmo de *k-nearest neighbor* (vecino más cercano) y lo aplica para encontrar los dos vecinos más cercanos: el primero perteneciente a su misma clase y el otro, a una clase impostora.

5.2.6. Importancia en esta investigación

El proceso de la minería de datos inicia con la exploración de datos y luego la validación mediante el remuestreo. La minería de datos también utiliza técnicas de remuestreo, tales como la validación cruzada para evitar el sobreajuste.

La minería de datos utiliza técnicas correctivas para manejar valores faltantes y valores atípicos. Impacta en la reducción total de la limpieza de

datos, mientras que en el marco de los métodos estadísticos se debe aplicar transformaciones a los datos que resultan en costo extra de procesamiento.

Para el objetivo que se plantea en esta investigación los métodos estadísticos colapsan, mientras que la minería de datos se especializa en manejar una gran cantidad de datos.

Con base en los conceptos anteriores, el método que más beneficia el enfoque de esta investigación y la implementación que se plantea, es el uso de una herramienta de minería de datos.

6. PLAN PARA IMPLEMENTAR UNA HERRAMIENTA DE MINERÍA DE DATOS QUE DETERMINE LOS FACTORES CLAVES EN EL ÍNDICE DE ÉXITO PARA ESTUDIANTES DE INGENIERÍA EN CIENCIAS Y SISTEMAS

6.1. Descripción del proceso de recolección de los datos

En cada una de las dependencias se recolecta la información de manera específica a través de procesos internos. El proceso en general inicia cuando cada dependencia recolecta la información, específicamente en:

- Bienestar Estudiantil
  - Evaluación de orientación vocacional

- Registro y Estadística
  - Cuestionario socioeconómico

- Centro de Cálculo
  - Datos académicos de los estudiantes

La primera fuente de información es Bienestar Estudiantil, en donde anualmente son generados y tabulados los resultados de las evaluaciones vocacionales realizadas durante el año en curso. El encargado de esta tarea es el personal de TI de la dependencia.

Luego, como segunda fuente de información, está el departamento de Registro y Estadística. Al finalizar el año tabula los datos generados por los estudiantes que llenan el cuestionario socioeconómico. Los encargados de realizar el proceso son el personal de TI junto con el estadístico del departamento.

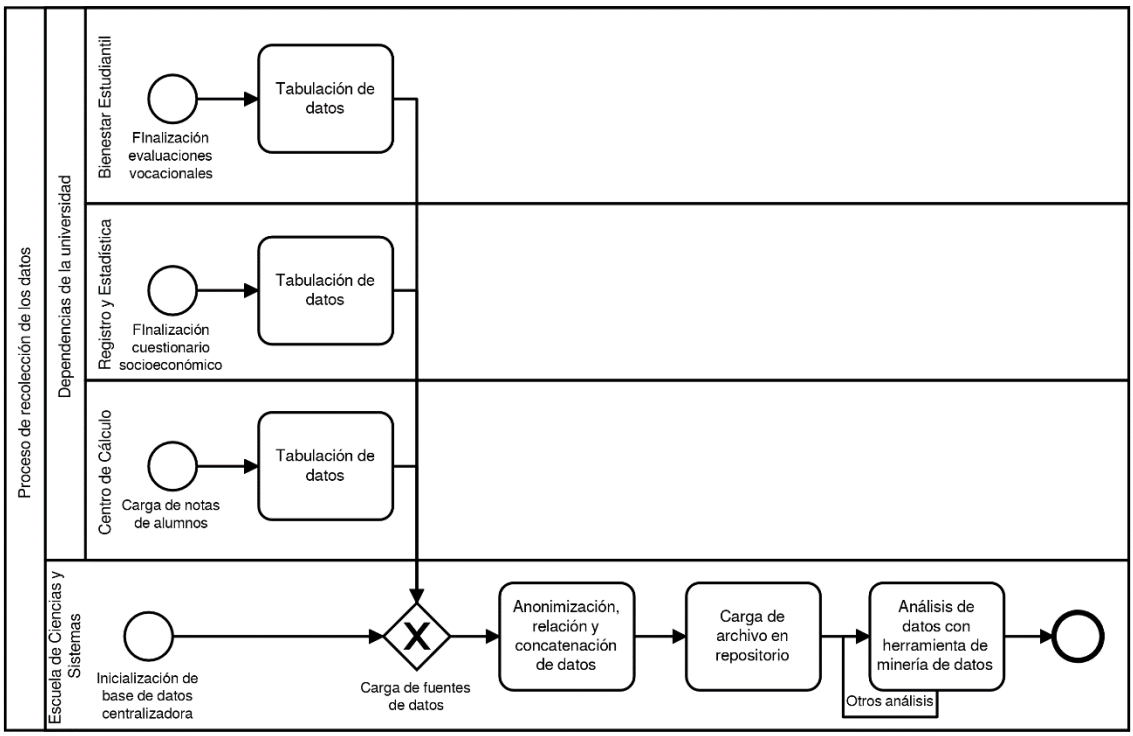
La última fuente de información considerada en el plan es el Centro de Cálculo, en donde dos veces al año es actualizada la información (notas y cursos) de los estudiantes. Los encargados del proceso son el personal de TI de la dependencia junto con la Secretaría Académica.

Tal como se explicó en el capítulo 3, es necesario vincular los datos generados por las dependencias; por lo tanto, los datos de las tres fuentes deben ser centralizados en una base de datos mediante un proceso de carga masiva de archivos en formato csv, tal como se muestra en el apéndice 4. El trabajo debe ser realizado por el personal interno de las dependencias, para luego aplicar tres procesos a la información: la anonimización de datos sensibles, la construcción de la relación entre datos y la concatenación de toda la información segmentada por año.

Por último, el analista debe cargar el archivo concatenado generado en un repositorio para ser consumido por la herramienta de minería de datos. Con esta se realiza el análisis propuesto, o algún otro análisis.

Actualmente, la universidad no cuenta con un sistema automatizado para obtener y vincular la información necesaria para ejecutar análisis con una herramienta de minería de datos. Por tal motivo, el plan considera que el personal interno de las dependencias cargará los datos anualmente al sistema central.

Figura 6. **Diagrama de proceso**



Fuente: elaboración propia.

Considerando que los datos son cargados en el servidor centralizador de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería, es necesaria una serie de elementos y procesos tecnológicos plasmados en un diseño de arquitectura de sistema, además de los recursos humanos y económicos para implementar la solución.

6.2. Diseño de arquitectura

Para plasmar de manera gráfica el plan propuesto en esta investigación se utilizó el modelo de arquitectura 4+1 de Philippe Kruchten, basado en cinco vistas concurrentes: vista lógica, de desarrollo, de procesos, física y escenarios. Todas pretenden mostrar gráficamente un conjunto de intereses.

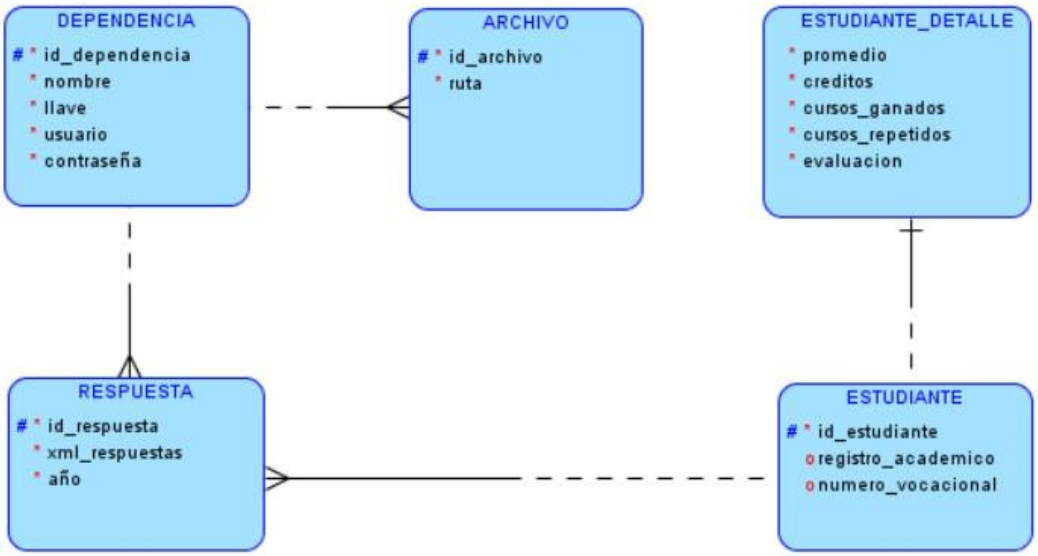
Para representar las vistas propuestas por el modelo mencionado se utilizan diferentes diagramas del Lenguaje Unificado de Modelado (UML). Para cada tipo de vista, son:

- Vista lógica
  - Diagrama de modelo de datos
  - Diagrama de clases
- Vista de desarrollo
  - Diagrama de componentes
- Vista de procesos
  - Diagrama de actividad
- Vista física
  - Diagrama de despliegue
- Escenarios
  - Diagrama de casos de uso

6.2.1. Vista lógica

Son diagramas para mostrar el modelo de datos y la relación entre los mismos, además de la representación de los datos en el software.

Figura 7. **Modelo de datos**



Fuente: elaboración propia.

Como puede observarse en la figura 7, el sistema estará compuesto por cinco entidades que tendrán que ser implementadas a nivel de base de datos y mapeadas a nivel de clases. Las entidades involucradas son:

- Dependencia: almacenamiento de información de las dependencias y credenciales para acceso a plataforma web.

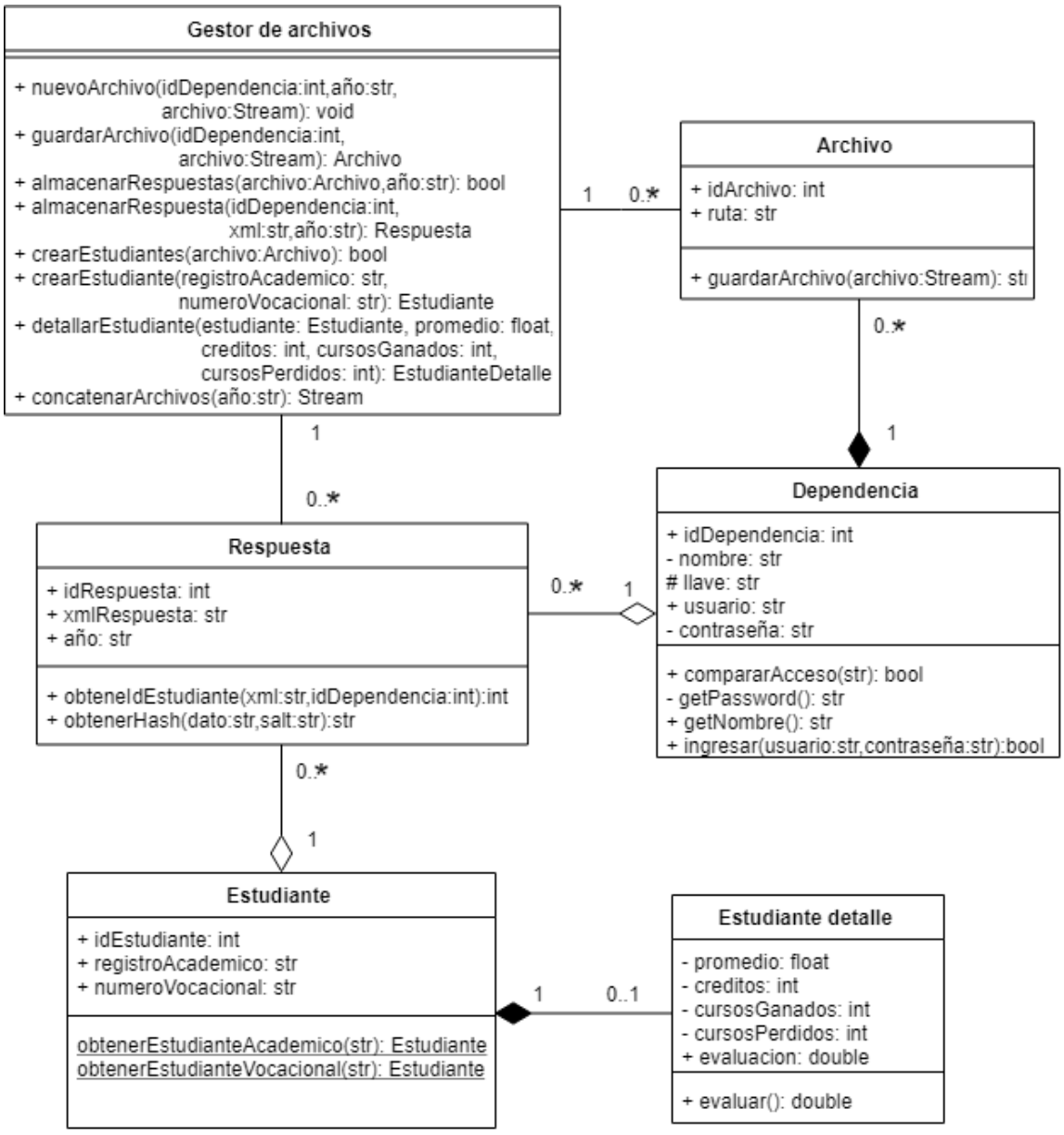
- Archivo: ruta de almacenamiento de archivo cargado por encargados en las dependencias.

- Respuesta: para almacenar en formato xml las respuestas contenidas en los archivos cargados, luego de la extracción, ordenadas por año.

- Estudiante detalle: información de los detalles académicos de cada estudiante y la evaluación obtenida del proceso.

- Estudiante: la relación entre número de registro académico y de orientación vocacional.

Figura 8. **Diagrama de clases**



Fuente: elaboración propia.

Usando como partida las entidades definidas a nivel de base de datos, se mapean las clases para su implementación en lenguaje de programación y posterior utilización como objetos en la ejecución de la solución. Las clases por implementar son:

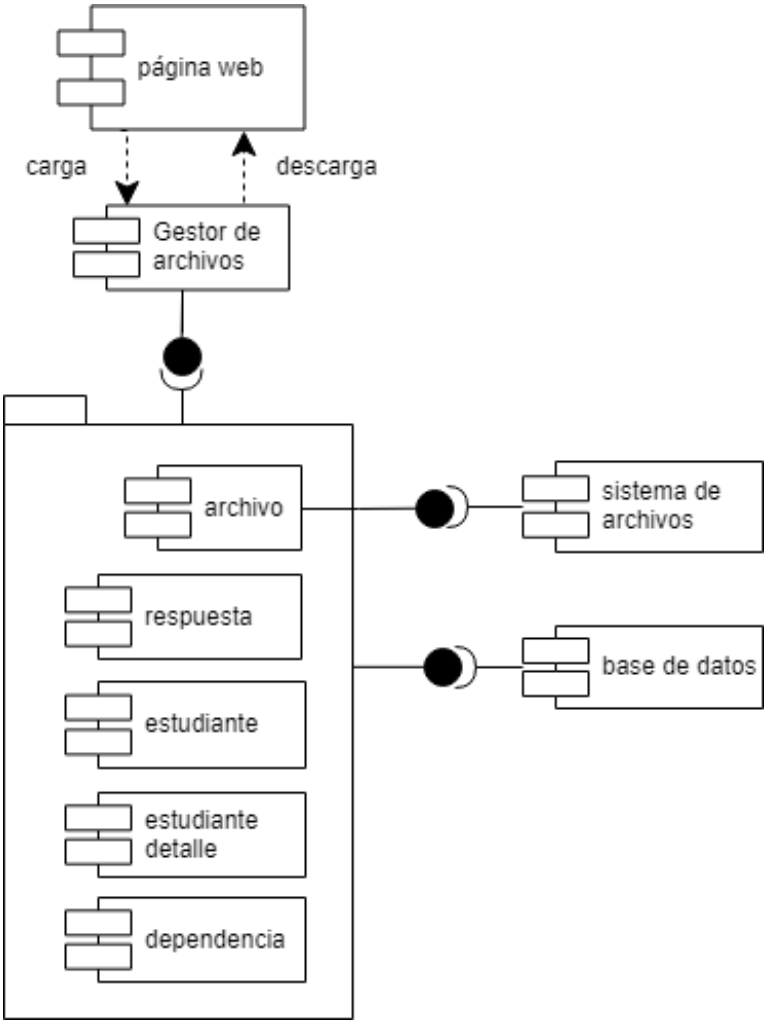
- Gestor de archivos
  - Carga los archivos de las dependencias y almacenarlos en el sistema correspondiente.
  - Genera un xml cuando el archivo es de las dependencias Registro y Estadística o Bienestar Estudiantil, para almacenarlo en la entidad dependencia.
  - Genera un nuevo estudiante o lo actualiza cuando el archivo cargado es de Centro de Cálculo.
  - Gestiona la concatenación y descarga de archivos para analizar con la herramienta.

- Archivo: gestiona el almacenamiento en el sistema de archivos del servidor.
- Respuesta: se encarga de las respuestas en xml y años a los que pertenecen. Realiza la función *hash* para anonimizar los datos.
- Dependencia: gestiona el acceso de las dependencias al sistema y la relación para identificar la pertenencia de los archivos con las respuestas.
- Estudiante: para relacionar el número de registro académico y de orientación vocacional con los detalles académicos de los estudiantes.
- Estudiante detalle: gestiona los datos académicos y evalúa al estudiante cada vez que se actualizan los datos.

6.2.2. Vista de desarrollo

Muestra la organización y ambiente estático del software y los módulos que compondrán el sistema y subsistemas.

Figura 9. **Diagrama de componentes**



Fuente: elaboración propia.

El diagrama de la figura 9 muestra las partes que componen el sistema. Este consta un paquete de software con 5 componentes y 4 componentes externos. Son:

- En el paquete de software
  - Componente archivo, encargado de comunicarse con el sistema de archivos para almacenar estos en el servidor.
  - Componente respuesta, lógica para almacenamiento y recuperación de archivos.
  - Componente estudiante, enlace entre datos académicos y repuestas.
  - Componente estudiante detalle, encargado de los datos académicos y la evaluación de los estudiantes.
  - Componente dependencia, manejo de credenciales, autenticación e información de las dependencias.

- Componente sistema de archivos: sistema nativo en el servidor destinado a la administración lógica de los archivos.
- Componente base de datos: motor de base de datos para preservar la información de los estudiantes y de las dependencias. Se comunica con todos los componentes del paquete de software .
- Componente página web: interfaz gráfica para comunicar al usuario con el sistema, por medio del gestor de archivos.
- Componente gestor de archivos: encargado de extraer la información de los archivos cargados, de concatenar las respuestas y servir los archivos al usuario. Se comunica con todos los componentes del paquete de software.

6.2.3. Vista de procesos

Esta presenta el ambiente y organización dinámica y de comunicación y sincronización entre actividades.

Figura 10. **Diagrama de actividad (carga de archivos de dependencias)**

Fuente: elaboración propia.

El diagrama de actividad acerca de la carga de archivos de dependencias, inicia cuando un usuario de una dependencia sube un archivo al sistema. Finaliza cuando ingresa los datos a la base de datos.

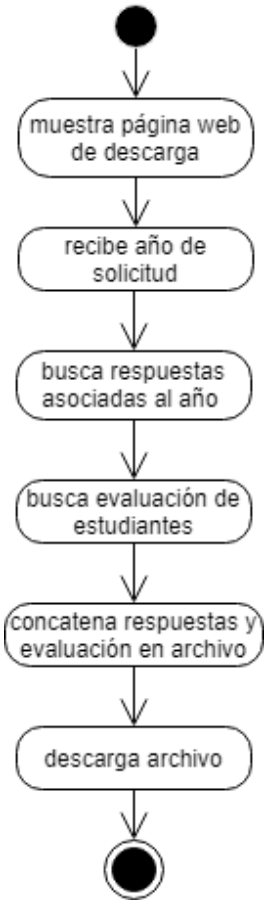
El primer paso lo realiza el usuario desde una dependencia mediante un navegador web. Ingresa a la plataforma alojada en el servidor web; se supone que el usuario ya ingresó al sistema con sus credenciales. Luego se muestra la página de carga de archivo, lo selecciona, así como el año al que corresponde y la dependencia encargada.

Luego de cargar el archivo al servidor, se almacena en el sistema y luego se procesa dependiendo a qué dependencia pertenece.

- Si pertenece a Centro de Cálculo, se crean los nuevos estudiantes y se anonimizan los números de registro académico (de acuerdo con el proceso explicado más adelante, en el subcapítulo 6.3). Si ya existe el estudiante se actualiza los datos y se ejecuta la función de evaluación, para generar la calificación según los datos.

- Si pertenece a Registro y Estadística o a Bienestar Estudiantil, obtiene el identificador del archivo, según el formato especificado en el apéndice 4. Luego genera el *hash* y obtiene el estudiante al que pertenece, para después almacenar las respuestas en formato xml, tal como se define en el apéndice 5.

Figura 11. **Diagrama de actividad (descarga de archivo concatenado)**

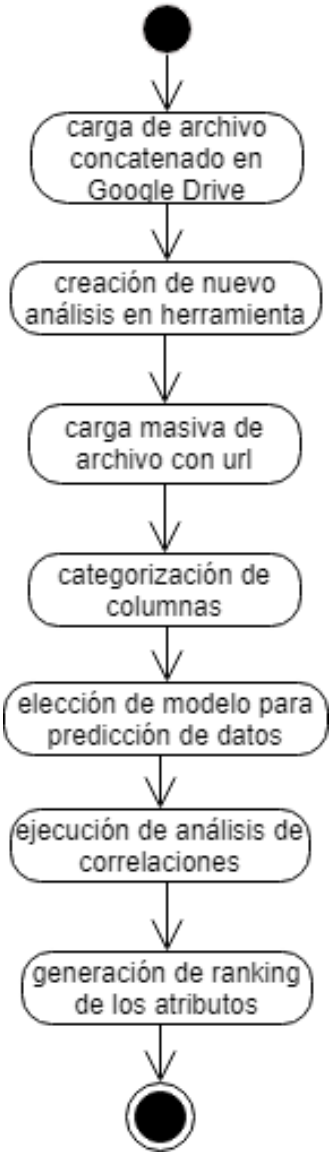


Fuente: elaboración propia.

El diagrama de actividad para la descarga de archivos concatenados inicia cuando el usuario de una dependencia, tras ingresar al sistema, despliega la página con un selector de año para indicar a cuál pertenecen los estudiantes que serán analizados.

Con el año del carné se buscan todas las respuestas de la base de datos asociadas a cada uno de los estudiantes; luego se concatenan en formato csv y se devuelve como respuesta el documento concatenado.

Figura 12. **Diagrama de actividad (análisis con minería de datos)**



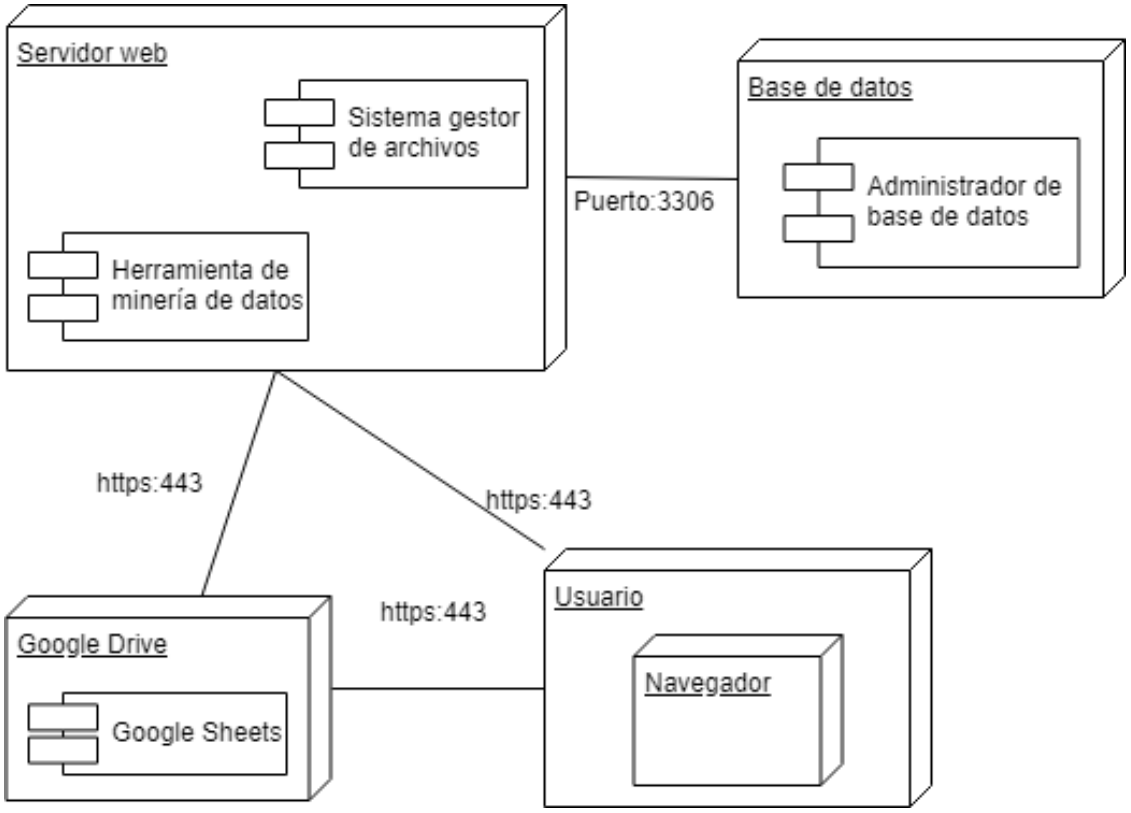
Fuente: elaboración propia, empleando DrawIO.

El diagrama de actividad para el análisis con minería de datos inicia con la carga del archivo concatenado a Google Drive, para luego realizar el análisis. El proceso es explicado a detalle más adelante en el subcapítulo 6.10.

6.2.4. Vista física

En este diagrama se muestra el despliegue de nodos físicos en donde se ejecutará el sistema, además de las configuraciones de red e infraestructura.

Figura 13. **Diagrama de despliegue**



Fuente: elaboración propia.

El sistema tiene un despliegue compuesto de tres capas, la de cliente, web y para persistencia, de la manera siguiente:

- En la capa de cliente los usuario utilizan un sistema de navegación web, tanto móvil como en computadora. Su función es ser la interfaz entre el usuario y el servidor web.
- La capa web está contenida en el servidor web; es la encargada de servir las interfaces para que el cliente pueda cargar y descargar los archivos, además de desarrollar los procesos internos definidos anteriormente.
- La capa de persistencia está compuesta por tres partes: el sistema de archivos, el repositorio en Google Drive y la base de datos relacional. Su objetivo es almacenar los archivos de las dependencias, los archivos concatenados y la información de los estudiantes y demás datos, respectivamente.

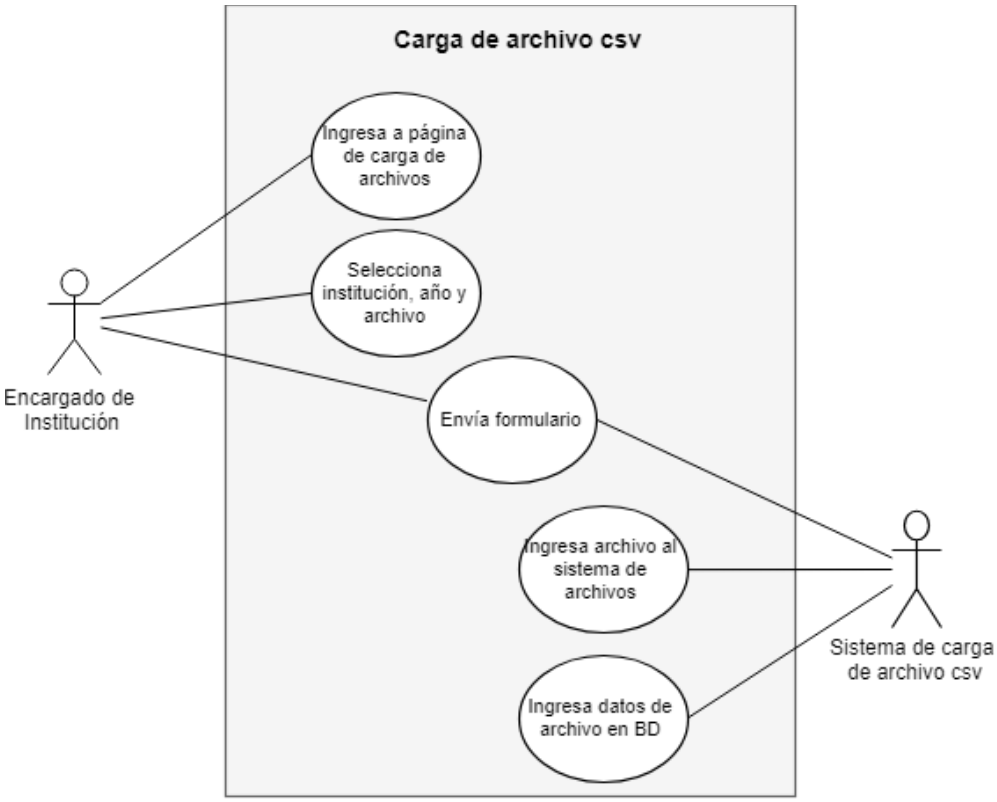
La capa web se conecta con la base de datos por el puerto 3 306 mediante el protocolo tcp; con la capa cliente y con el repositorio en Google Drive a través de https, para consumir los archivos concatenados desde la herramienta en el servidor. Además, internamente, está interconectado al sistema de archivos.

La capa de cliente, utilizada por los usuarios por medio de un navegador web, se conecta con el repositorio en Google Drive para cagar los archivos concatenados listos para el análisis.

6.2.5. Escenarios

Los escenarios son los casos de uso de mayor importancia e impacto para los interesados en el sistema. Resuelven las necesidades y proveen el servicio esperado por el usuario final.

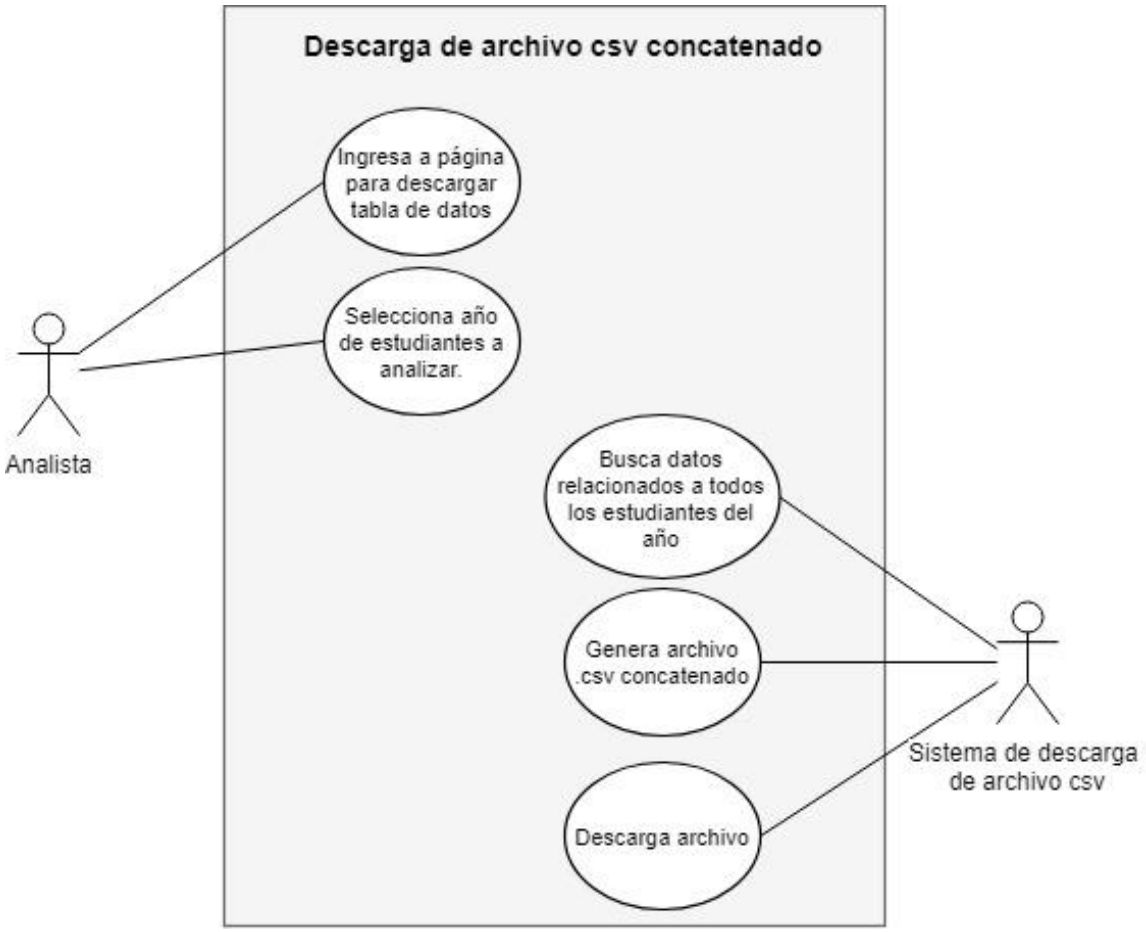
Figura 14. **Caso de uso 1**



Fuente: elaboración propia.

En el caso de uso 1 se presenta las interacciones que el encargado de la institución debe realizar con el sistema de carga de archivos csv, para ingresar la información de los estudiantes y las repuestas al sistema.

Figura 15. **Caso de uso 2**



Fuente: elaboración propia, empleando DrawIO.

El caso de uso 2 presenta las interacciones del analista de datos con el sistema de descarga de archivos csv, para obtener los datos concatenados de los estudiantes respecto al año solicitado en la selección.

6.3. Proceso de anonimización de la información

Con base en la importancia, consideraciones y objetivos planteados en el marco legal tanto nacional como internacional en esta investigación, se utilizará la anonimización en las fuentes de datos principales de la solución.

Toda variable que identifica como persona individual o que puede ser relacionada con la identidad del estudiante en cada una de las fuentes de datos, debe ser anonimizada.

- Cuestionario socioeconómico de Registro y Estadística
  - Registro académico
- Evaluación de orientación vocacional de Bienestar Estudiantil
  - Número de orientación vocacional
- Información académica de estudiantes en Centro de Cálculo
  - Registro académico
  - Número de orientación vocacional

Luego, se ingresa cada uno de los datos detectados por un algoritmo de *hash*, definido como proceso clave en la carga de datos, para luego obtener el dato cifrado que sustituya el dato sensible.

Después de este procedimiento los datos ya pueden ser utilizados sin riesgo de exponer la identidad de los estudiantes.

6.4. Recursos humanos necesarios para la solución

Para la implementación de la solución es necesario que existan tres roles principales en la escuela de Ciencias y Sistemas, si el personal interno de las dependencias carga los archivos de datos al servidor central.

- Analista de sistemas o estadístico
  - Ejecutar herramienta para descarga de archivo.
  - Verificar la integridad de los datos.
  - Ejecutar el proceso de minería de datos y análisis de resultados.
- Programador *Senior*
  - Implementar la base de datos.
  - Efectuar la rutina de software para anonimización de datos.
  - Diseña e implementar aplicación web para carga de datos.
  - Implementar la rutina de software para carga y almacenamiento de datos.
  - Diseñar e implementar la aplicación web para concatenación y descarga de archivo csv concatenado.
  - Implementar la rutina de software para concatenación y descarga de archivo csv concatenado.
- Administrador de sistemas
  - Inicializar y estructurar servicios de hardware.
  - Velar por el correcto funcionamiento del hardware.
  - Instalar herramientas y complementos.
  - Velar por el correcto funcionamiento del software.
  - Evaluar y aprovisionar para el desempeño de la herramienta.
  - Actualizar librerías de software.

6.5. Elección de herramienta para minería de datos

En la minería de datos existen múltiples herramientas que permiten implementar una solución de manera efectiva y para seleccionar la adecuada es necesario realizar el proceso de conocer a detalle cada una.

6.5.1. Herramientas disponibles

En la minería de datos existen múltiples herramientas que permiten implementar una solución de manera efectiva.

A continuación, se lista las herramientas más populares utilizadas para la minería de datos y una breve descripción.

- RapidMiner: desarrollado en lenguaje Java, ofrece análisis predictivo, modelos estadísticos, visualización de datos y preprocesamiento.
- Weka: también construido en lenguaje Java, ofrece clúster, clasificación, regresión, visualización y selección de características.
- Orange: su principal característica es una interfaz gráfica que permite diseñar el flujo de trabajo visualmente.
- KNIME: desarrollado en lenguaje Java y cuenta con *plugins* para extender su funcionalidad.

6.5.2. Descripción de herramientas

Para elegir una herramienta que se adecue a las necesidades específicas de la universidad, es necesario conocer a detalle las funcionalidades, ventajas y desventajas utilizando como referencia las fuentes oficiales.

6.5.2.1. RapidMiner

Toda la información de la herramienta fue obtenida a través de la página oficial de la misma, https://rapidminer.com/.

Figura 16. **Logo de RapidMiner**

Fuente: *Página oficial*. https://rapidminer.com/. Consulta: 10 de marzo de 2019.

Este software tiene cuatro soluciones enfocadas en la minería de datos.

- RapidMiner Studio es un ambiente para diseñar procesos analíticos que utilizan aprendizaje automático, análisis de texto, predictivo y de negocios.
- RapidMiner Server es el ambiente que se utiliza para desplegar las soluciones diseñadas de procesos de análisis en aplicaciones tipo cliente.
- RapidMiner Radoop es un ambiente utilizado para diseñar soluciones avanzadas de análisis de procesos especializados en despliegue usando clúster Hadoop.
- RapidMiner Cloud es un servicio que ofrece RapidMiner para usuarios de RapidMiner Studio *Community* y *Professional* para el despliegue de soluciones.

RapidMiner es una plataforma del campo de la ciencia de los datos que auxilia a los equipos en la preparación de datos, aplicación de aprendizaje automático y despliegue de modelos predictivos.

Las principales características de la herramienta son:

- Unificación de plataforma, ofrece una sola interfaz para todo el sistema.
- Diseño visual del flujo de trabajo, plataforma desarrollada con técnicas de fácil aprendizaje.
- Múltiples funciones de aprendizaje automático e integraciones con aplicaciones externas.
- Mejoras continuas.
- Conexión con múltiples fuentes de datos.

Ofrece soporte para los siguientes sistemas operativos:

- Windows 32Bits y 64Bits
- Mac OS
- Linux

6.5.2.2. Weka

Toda la información de la herramienta fue obtenida a través la página oficial de la herramienta, www.cs.waikato.ac.nz.

Figura 17. **Logo de Weka**

Fuente: *Página oficial*. https://www.cs.waikato.ac.nz. Consulta: 10 de marzo de 2019.

Es un software desarrollado por la Universidad de Waikato de Nueva Zelanda bajo *GNU General Public License.* Witten & Frank (2016) describe que está basado en una colección de algoritmos de aprendizaje automático enfocados en la minería de datos.

Ofrece un catálogo de cursos gratuitos para aprender el uso de la herramienta, permite la carga masiva de datos y una amplia gama de archivos, entre ellos, csv, json, m, dat, bsi.

Su principal función es modificar y ajustar el conjunto de datos, la clasificación y regresión con base en los algoritmos de aprendizaje automático, la aplicación de reglas a los datos, selección de atributos más relevantes y visualización de gráficas interactivas.

Tiene dos editores importantes, el SQL y el de redes Bayesianas. Debido a que está desarrollado en el lenguaje Java, el único requerimiento específico para el correcto funcionamiento de la herramienta es la instalación de la máquina virtual del lenguaje java 1.8.

Ofrece soporte para los siguientes sistemas operativos:

- Windows 32Bits y 64Bits
- Mac OS
- Linux

6.5.2.3. Orange

Toda la información de la herramienta fue obtenida a través de la página oficial de la herramienta, https://orange.biolab.si/.

Figura 18. **Logo de Orange**

Fuente: *Página oficial*. https://orange.biolab.si/. Consulta: 10 de marzo de 2019.

Es una herramienta de software libre para uso de aprendizaje automático y visualización de datos. Permite el análisis interactivo de procesos datos con una extensa cantidad de herramientas.

Permite la visualización y exploración de distribuciones estadísticas representadas en gráficas de barras, árboles de decisión, mapas de calor y proyecciones lineales.

Tiene interfaz gráfica con estilo de pizarra interactiva para diseñar el proceso de carga, transformación y análisis de datos. Para la carga de datos en la herramienta se utiliza los formatos csv, Microsoft Excel, tsv y una conexión mediante url con archivos de *Google Sheets*. Cuando se finaliza la carga de datos y se generan los modelos, es posible realizar predicciones con nuevos datos de entrada.

Ofrece un catálogo de tutoriales en formato video y cursos prácticos que llevan paso a paso al usuario. Además, ofrece soporte para los siguientes sistemas operativos:

- Windows 32Bits
- Windows 64Bits
- Mac OS
- Linux

6.5.2.4. KNIME

Toda la información de la herramienta fue obtenida a través de la página oficial de la herramienta, https://orange.biolab.si/.

Figura 19. **Logo de KNIME**

Fuente: *Página oficial*. https://www.knime.com. Consulta: 10 de marzo de 2019.

Es una herramienta que permite el análisis guiado por datos; su uso es libre bajo *GNU General Public License.* Ofrece cursos presenciales de pago y tutoriales en formato video gratuitos. Es posible integrar análisis de grandes cantidades de datos mediante el uso de *Apache Hadoop*.

La plataforma permite integraciones de código con funciones construidas en los lenguajes C, C++, R, Java y Python. Es utilizado en varias áreas de desarrollo e investigación. Las principales son:

- Inteligencia para el cliente
- Medios sociales
- Finanzas
- Farmacias y cuidado de la salud
- Ventas
- Gobierno

También permite el uso de aprendizaje automático con la extensión de librerías externas. Para la carga de datos acepta formatos json, XML y texto plano.

6.6. Comparación de herramientas

Para determinar la mejor herramienta para la investigación, se toman en cuenta cinco características clave que inciden directamente en la elección:

- Función para correlacionar atributos en predicción
    - Es la función principal que debe tener la herramienta, dado que el propósito del plan es encontrar la correlación de los atributos con el éxito académico de los estudiantes.

- Soporte para gran cantidad de datos
    - La motivación principal para usar una herramienta de minería de datos es la diferencia en el soporte de datos en relación con los métodos estadísticos tradicionales.

- Licencia de la herramienta
    - Desde la perspectiva financiera y técnica el tipo de licencia es importante, debido a la relación entre el costo económico y el esfuerzo de implementación.
- Curva de aprendizaje
    - Es importante dado que las personas que se proyecta que usarán la herramienta, pueden no tener un conocimiento previo en su uso.
- Soporte técnico
    - Todo software en algún momento de su ciclo de vida tiende a fallar, por lo que es importante que exista un método para buscar soluciones a las fallas. Aquí también se tiene en cuenta las actualizaciones y parches.

Tabla III. **Comparación de características de herramientas**

Característica	RapidMiner	Weka	Orange	KNIME	Peso
Correlación de atributos en predicción	1	0	1	1	30
Soporte de gran cantidad de datos	1	0	1	1	25
Tipo de licencia	0,5	1	1	1	25
Curva de aprendizaje	1	0,3	0,5	0,3	10
Soporte técnico	1	0,5	0,5	0,5	10
Total	87,5	33	90	88	100

Fuente: elaboración propia.

Detalle de los valores dados a cada característica:

- Función para correlacionar atributos en predicción
  - 1, posee la característica
  - 0, no posee la característica
- Soporte para gran cantidad de datos
  - 1, posee la característica
  - 0, no posee la característica
- Tipo de licencia de la herramienta
  - 1, es software de licencia libre
  - 0.5, es software de licencia propietaria
- Curva de aprendizaje
  - 1, define una complejidad simple
  - 0,5, define una complejidad media
  - 0,3, define una complejidad alta
- Soporte técnico
  - 1, soporte de comunidad y por pago a la organización
  - 0,5, soporte de comunidad o por pago a la organización

6.7. Justificación de la elección de herramienta

La herramienta seleccionada para implementar la solución en la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería es Orange, según la cualificación obtenida con base en las características de mayor importancia.

Las principales cualidades que hacen de esta herramienta la elección más efectiva, son: la curva de aprendizaje es media, está desarrollada bajo la licencia de software libre *GNU General Public License* y ofrece los procesos necesarios para determinar los factores clave.

La curva de aprendizaje es media debido a la variedad de tutoriales y extensa documentación; además, el uso de la herramienta es intuitivo e informado.

Es importante que esté desarrollada bajo *GNU General Public License,* ya que permite a la institución el uso libre de costos por licenciamiento.

Para finalizar, la herramienta ofrece el proceso necesario para implementar la solución. Inicia con la carga de datos, visualización y exploración de distribuciones estadísticas; análisis de datos, creación de modelos de predicción y determinación de factores clave con base en el comportamiento del conjunto de datos.

6.8. Recursos tecnológicos necesarios para implementar la herramienta

Para la correcta ejecución de la herramienta es necesaria una serie de elementos tecnológicos que la apoyen.

Los detalles técnicos recomendados para la implementación y los mínimos según especificaciones en la página oficial de la herramienta, son:

- Hardware
  - Servidor con procesador mayor a 2GHZ
  - 2GB de memoria
- Software
  - Ubuntu 16,04 LTS
  - Orange 3,1
  - Python 3,5

- Pip3

6.9. Instalación de herramienta

Para instalar la herramienta Orange en el sistema operativo Ubuntu 16,04 LTS es necesaria una secuencia de pasos que se detallan a continuación:

- Instalar sistema operativo Ubuntu 16,04 LTS
  - La fuente se encuentra para su descarga libre y gratuita en la página https://www.ubuntu.com/download/desktop
  - Seguir entorno gráfico de instalación
- Instalar pip3 con versión de Python 3,5
  - Instalar desde la terminal de Linux y ejecutar los comandos
    - *apt-get install python3-pip*
    - *pip3 install –upgrade pip*
- Instalar Orange 3,1
  - Instalar desde la terminal mediante el manejador de paquetes de pip3, con los comandos siguientes:
    - *pip3 install –no-binary orange3 orange3*
  - Ejecutar Orange 3,1 desde la terminal.
    - Orange-canvas

6.10. Entrada de datos a la herramienta

El formato de los datos de entrada debe ser estandarizado y mantenerse así durante todo el proceso, de esta manera el sistema capta sin errores todos los datos para su procesamiento.

6.10.1. Formato de datos de entrada

Para iniciar con el análisis de los datos el proceso de carga masiva es indispensable, ya que una gran cantidad de datos aumenta la probabilidad de éxito en la determinación de los factores clave.

La herramienta tiene la capacidad de manejar archivos nativos de Excel (.xlsx o .xls) y archivos delimitados por coma o tabulador (.csv); también posee su propio formato de datos y lector de hojas de cálculo de *Google Sheets* por medio de una url.

La estructura del archivo es una tabla compuesta por atributos y sus valores. Los atributos pueden ser de tipo continuo, discreto, temporales o cadenas de texto y debe ser definidos en el encabezado.

6.10.2. Proceso de carga masiva

La carga masiva de datos es un proceso que provee Orange y puede ser realizado con dos procedimientos diferentes: el primero es un archivo almacenado localmente y el segundo, por medio de *Google Sheets* con una url que indica el origen de los datos.

Para la carga masiva de los datos en esta solución se sugiere utilizar la herramienta remota y almacenar en Google Drive los archivos de datos concatenados para alimentar el análisis, generados en el proceso de descarga anteriormente definido.

Para la carga masiva a través de la herramienta remota, se ejecutan los siguientes pasos dentro del programa Orange 3,1:

- Se coloca el elemento File dentro del canvas de Orange 3,1, como se visualiza en la figura 20.

Figura 20. **Objeto tipo *File***

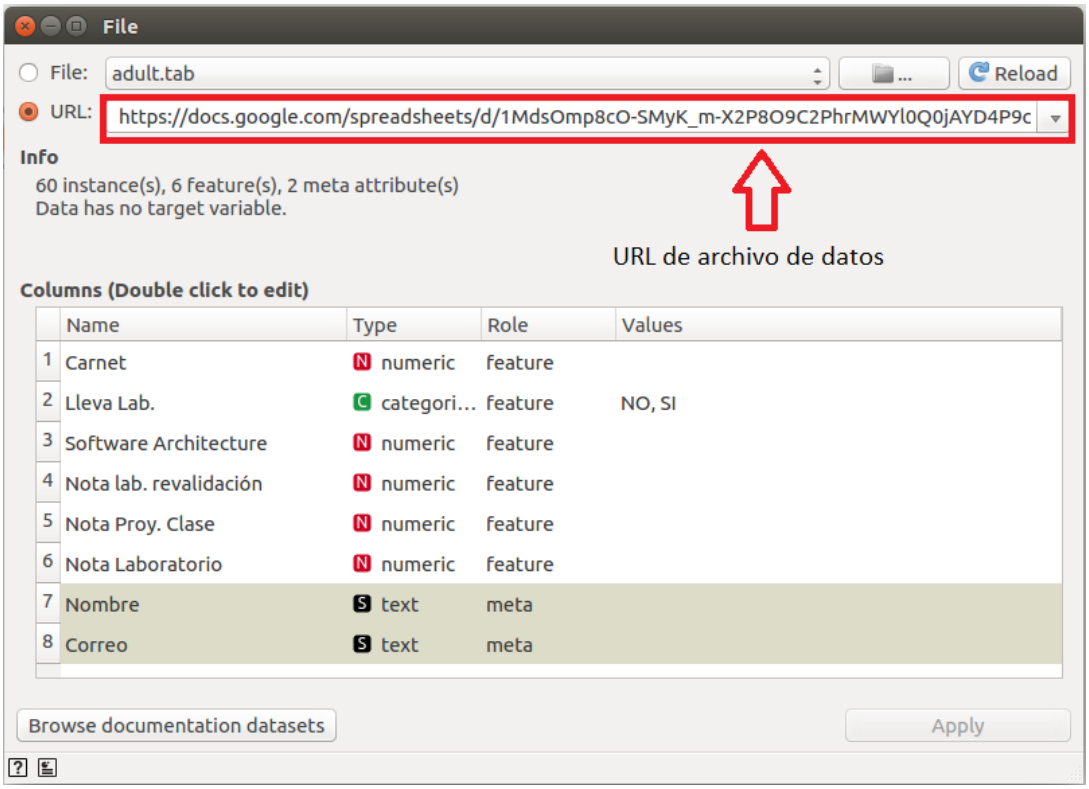
Fuente: *Página oficial*. https://orange.biolab.si/. Consulta: 3 de junio de 2018.

- Se dan dos clics sobre el objeto y esto abre una ventana en donde se selecciona la opción para ingresar la dirección url del archivo de datos. Como muestra la figura 21, luego se despliega la lista de columnas con los elementos siguientes:

    - Nombre: identificación de la columna
    - Tipo: dominio de datos válidos
        - Categoría
        - Numérico
        - Texto
        - Fecha
    - Rol: describe qué función cumple la columna dentro del contexto
        - Atributo
        - Valor por clasificar (clase)
        - Meta (descriptivo)
        - Ignorado

- Valores: posibles valores que pueden tomar los datos de la columna

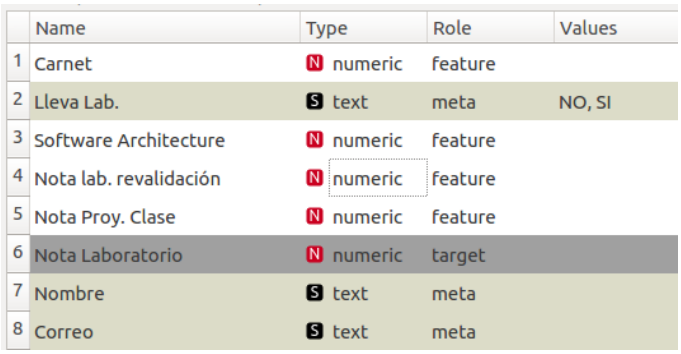
Figura 21. **Ventana para carga de datos**



Fuente: elaboración propia.

- Luego se selecciona un valor objetivo para clasificar, además de un tipo y un rol para cada una de las columnas restantes.

Figura 22. **Selección de valor por clasificar**



	Name	Type	Role	Values
1	Carnet	numeric	feature	
2	Lleva Lab.	text	meta	NO, SI
3	Software Architecture	numeric	feature	
4	Nota lab. revalidación	numeric	feature	
5	Nota Proy. Clase	numeric	feature	
6	Nota Laboratorio	numeric	target	
7	Nombre	text	meta	
8	Correo	text	meta	

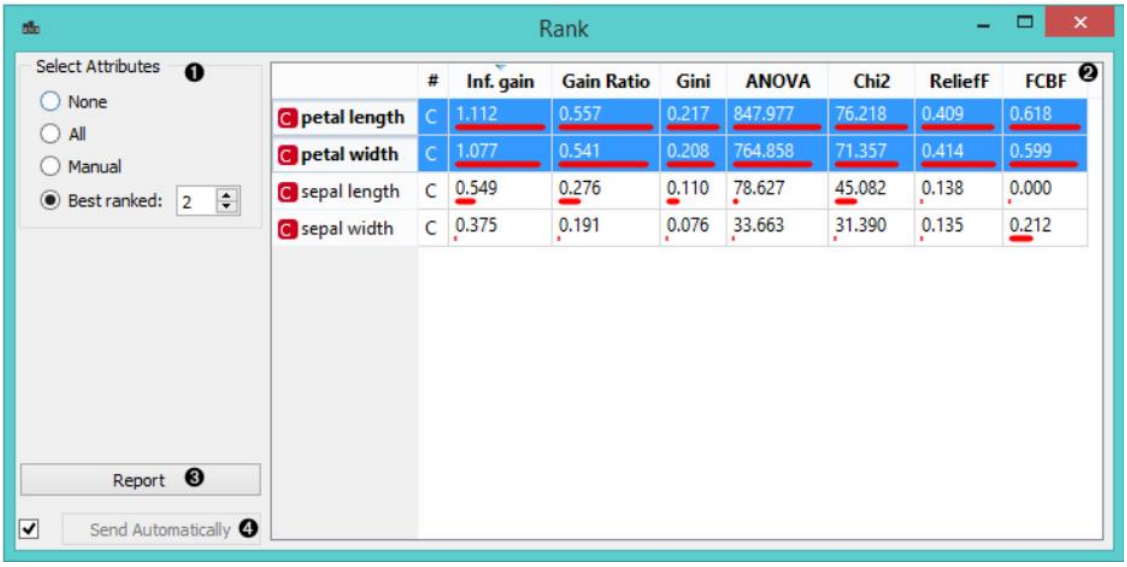
Fuente: elaboración propia.

6.11. Configuración de la herramienta

En el contexto de la herramienta Orange 3.,1 existen múltiples métodos de ranking que indican los atributos de la muestra que más determinan el resultado objetivo. Es decir, dentro de este estudio, muestran cuáles son las preguntas del cuestionario y las habilidades más importantes con base en la correlación del índice de éxito de un estudiante para Ingeniería en Ciencias y Sistemas, según fue definido en el capítulo 4.

Para determinar los factores claves se utiliza la herramienta interna *Rank* de Orange 3,1. Esta espera una entrada de datos y un conjunto de modelos de predicción que le ayudan a clasificar efectivamente, y devuelve un ranking de los atributos. Cada atributo posee siete descriptores, además de los que tengan según los modelos que se utilicen, como se muestra en la figura 23.

Figura 23. **Descriptores para atributos**



Fuente: elaboración propia.

La herramienta interna *Rank* permite utilizar modelos de predicción para alimentar el proceso de clasificación. También funciona sin ayuda de modelos de predicción, ya que internamente genera un indicador por defecto con el algoritmo de Relief.

Los modelos de predicción permitidos para alimentar la herramienta de ranking son 4 y su selección queda a discreción del estadista.

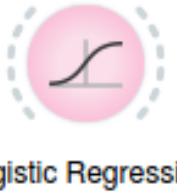
- Regresión lineal: es una aproximación que describe la relación que existe entre una variable dependiente y las variables independientes. El icono de la herramienta se muestra en la figura 24.

Figura 24. **Herramienta de regresión lineal**

Fuente: elaboración propia.

- Regresión logística: es un tipo de regresión utilizado para predecir el resultado de una variable dependiente, dado que es posible clasificarla discretamente, con base en las variables independientes. El icono de la herramienta se muestra en la figura 25.

Figura 25. **Herramienta de regresión logística**



Fuente: elaboración propia.

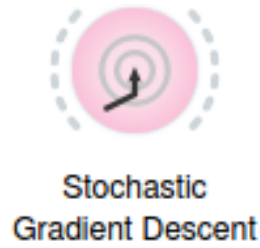
- Bosque de decisión aleatorio: está basado en construir árboles de decisión que, en conjunto, promedian la clasificación. El icono de la herramienta se muestra en la figura 26.

Figura 26. **Herramienta de bosque de decisión aleatorio**

Fuente: elaboración propia.

- Gradiente descendiente estocástico: es un método que trata de encontrar el mínimo y el máximo a través de iteraciones. El icono de la herramienta se muestra en la figura 27.

Figura 27. **Herramienta de gradiente descendiente estocástico**

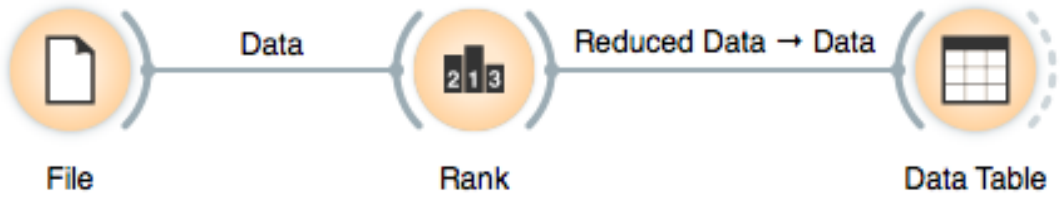


Fuente: elaboración propia.

Para alimentar la herramienta de ranking se selecciona dos modelos y se ingresa los datos seleccionados, etiquetados y con el valor objetivo determinado.

El modelo final para determinar los factores clave que determinan el éxito se muestra en la figura 28.

Figura 28. **Modelo de proceso en herramienta**



Fuente: elaboración propia.

6.12. Estimación de costos para implementación

Teniendo en cuenta las actividades y recursos necesarios descritos para implementar el plan, los costos asociados fueron divididos en dos partes: una de recursos tecnológicos y la otra de recursos humanos.

- Tecnológicos

Para el plan de implementación se evaluaron los costos estimados asociados al hardware si se utilizaran servicios de un proveedor estándar de plataforma como servicio, según datos obtenidos el 20 de mayo de 2019 en la página https://aws.amazon.com/es/pricing/.

Tabla IV. **Costos estimados por hardware**

Hardware	**Costo anual**
1 servidor 2GHZ con 2GB RAM	Q1 500,00
30 GB espacio de almacenamiento no volátil	Q100,00
IP Publica	Q-
Nombre de dominio	Q150,00
Encriptación de protocolo de internet https	Q-
Total	**Q1 750,00**

Fuente: elaboración propia.

- Humanos

Para estimar los costos asociados a los recursos humanos necesarios para implementar el plan se evaluó la cantidad de tiempo promedio de desarrollo para cada tarea que realiza cada uno de los recursos; luego se generó un estimado con el costo promedio por hora de cada uno, con base en datos de referencia de las empresas Tigo Guatemala, Conduent y Telus.

Tabla V. **Horas estimadas por tarea para programador**

Tarea	Horas	Periodicidad
Implementación de base de datos	4	1 vez
Rutina de software para proceso anonimización de datos	5	1 vez
Diseño e implementación de aplicación web para carga de datos	15	1 vez
Rutina de software para carga y almacenamiento de archivos	20	1 vez
Diseño e implementación de aplicación web para concatenación y descarga de archivo csv concatenado	15	1 vez
Rutina de software para concatenación y descarga de archivo csv concatenado	20	1 vez
Soporte de usuario	2	A requerimiento
Corrección de errores	3	A requerimiento
Total	**84**	

Fuente: elaboración propia.

Tabla VI. **Horas estimadas por tarea para analista o estadista**

Tarea	Horas	Periodicidad
Ejecución de herramienta para descarga de archivo	1	cada análisis
Verificación de la integridad de los datos	5	cada análisis
Ejecución de minería de datos y análisis de resultados	8	cada análisis
Total	**14**	

Fuente: elaboración propia.

Tabla VII. **Horas estimadas por tarea para administrador de sistemas**

Tarea	Horas	Periodicidad
Inicialización y estructuración de servicios de hardware	5	1 vez
Instalación de herramienta y complementos	5	1 vez
Evaluación de correcto funcionamiento del hardware	1	cada análisis
Evaluación de correcto funcionamiento del software	1	cada análisis
Evaluación y aprovisionamiento según desempeño de la herramienta	2	cada análisis
Actualización de librerías	1	cada análisis
Total	**15**	

Fuente: elaboración propia.

Tabla VIII. **Costos estimados iniciales**

Recurso	Horas	Costo hora	Total
Programador	84	Q50	Q4200
Analista	14	Q60	Q840
Administrador de sistemas	15	Q60	Q900
Total			**Q5 940**

Fuente: elaboración propia.

Tabla IX. **Costos estimados en cada análisis**

Recurso	Horas	Costo hora	Total
Programador	5	Q50	Q250
Analista	15	Q60	Q900
Administrador de sistemas	5	Q60	Q300
Total			**Q 1 450**

Fuente: elaboración propia.

7. IMPLEMENTACIÓN DEL PLAN PARA ESTUDIANTES DE CIENCIAS Y SISTEMAS

Para demostrar la factibilidad, a manera de guía se implementó el plan propuesto con los datos recolectados en las dependencias descritas en los capítulos anteriores. Se utilizó la herramienta seleccionada para ejemplificar cómo determinar los factores clave que afectan el éxito en el cierre de pensum de los estudiantes de Ingeniería en Ciencias y Sistemas.

Al realizar todo el proceso descrito en el plan, el resultado es una tabla con un grupo de factores en orden de mayor a menor correlación. Son datos que posteriormente se pueden utilizar para otro tipo de análisis e investigación.

El plan puede ser aplicado a cualquier carrera de la Universidad de San Carlos de Guatemala; sin embargo, se debe tener en cuenta que el proceso para obtener la información académica de los estudiantes y la manera en cómo está dispuesta varía en cada facultad.

7.1. Proceso realizado para obtener los datos en las dependencias

Luego de determinar los datos más relevantes y de más impacto, se realizó un proceso en cada dependencia para obtener una muestra de los datos.

7.1.1. Evaluación de orientación vocacional de Bienestar Estudiantil

Para obtener los resultados de los exámenes vocacionales, se ingresó por medio de una carta (mostrada en el apéndice 2) la solicitud de los datos dirigida a la jefa de bienestar estudiantil, la licenciada Dora Leticia Quiñónez, con el visto bueno del director de la Escuela de Ciencias y Sistemas. Se solicitó los datos de los resultados de las evaluaciones vocacionales de los estudiantes del año 2007 al 2017.

Luego se envió la solicitud por correo electrónico y se adjuntó la lista de los números de orientación vocacional de interés. También se envió la carta de autorización de la jefa de Orientación Vocacional (mostrada en el anexo 2) al encargado del manejo de los datos de las evaluaciones vocacionales, Johnny Barrios. Él realizó la consulta y consolidación de los datos solicitados, para luego entregarlos.

7.1.2. Cuestionario socioeconómico de Registro y Estadística

Para obtener los resultados del cuestionario se ingresó una solicitud por medio de una carta (mostrada en el apéndice 1) dirigida a la jefa del Departamento de Estadística, la licenciada Erica Marroquín, con el visto bueno del director de la Escuela de Ciencias y Sistemas, el ingeniero Marlon Pérez Turk.

Luego se realizó una solicitud verbal (con el visto bueno de la jefa del Departamento de Estadística) al encargado del proceso, el licenciado Armando Guzmán, quien formuló la solicitud al centro de tecnología del departamento.

7.1.3. Información académica de estudiantes en Centro de Cálculo

Para obtener los datos académicos de los estudiantes de la carrera de Ciencias y Sistemas de la Facultad de Ingeniería, se dirigió una carta (mostrada en el apéndice 3) a la Secretaría Académica, con el visto bueno del director de la Escuela de Ciencias y Sistemas.

Los datos requeridos fueron el número de registro académico y de orientación vocacional; cursos aprobados, créditos, promedio, total de cursos repetidos y desasignados para alumnos con carné del año 2008 en hasta el año 2017.

Luego, Secretaría Académica transfirió la solicitud a Centro de Cálculo, para el procesamiento de los datos solicitados. Después los envió junto con una carta de aprobación de Secretaría Académica (mostrada en el anexo 3), dentro de un CD en formato Microsoft Excel.

7.2. Datos obtenidos en la fase de investigación de campo

Los datos de los estudiantes en cada dependencia fueron obtenidos en el proceso de investigación de campo realizado en esta investigación, con el objetivo de mostrar un ejemplo de cómo se debería utilizar la herramienta y cómo se verían los resultados finales.

7.2.1. Bienestar Estudiantil

Los resultados de las evaluaciones vocacionales de los años 2007 al 2016 fueron entregados en formato csv, con un total de 3 285 registros de estudiantes, compuestos por 24 datos cada uno, para un total de 78 840 datos durante 10 años.

Figura 29. **Muestra de datos de evaluaciones vocacionales del año 2016**

APRE_NUMERICA	APRE_ABSTACTA	APRE_VERBAL	MEDICINA	FARMACIA	ODONTOLOGI	VETERINARI
MEDIO ALTO	ALTO	MEDIO BAJO	6	13	0	13
MEDIO ALTO	MEDIO ALTO	MEDIO ALTO	0	6	0	0
MEDIO ALTO	ALTO	MEDIO BAJO	6	20	0	0
MEDIO ALTO	MEDIO ALTO	MEDIO BAJO	0	0	0	0
ALTO	ALTO	MEDIO BAJO	0	0	0	0
MEDIO ALTO	ALTO	MEDIO BAJO	33	47	0	13
MEDIO BAJO	MEDIO BAJO	MEDIO BAJO	0	0	0	0
MEDIO ALTO	ALTO	MEDIO BAJO	20	20	0	6
MEDIO BAJO	MEDIO ALTO	MEDIO BAJO	20	0	0	33
ALTO	ALTO	MEDIO BAJO	0	0	0	0
MEDIO ALTO	MEDIO BAJO	MEDIO ALTO	0	0	0	0
MEDIO ALTO	ALTO	ALTO	33	13	80	0
ALTO	ALTO	MEDIO BAJO	0	0	0	0
ALTO	ALTO	MEDIO ALTO	0	0	0	0
ALTO	ALTO	MEDIO BAJO	0	0	0	0
MEDIO ALTO	MEDIO ALTO	MEDIO BAJO	6	27	0	0
MEDIO ALTO	ALTO	MEDIO BAJO	0	6	0	0
MEDIO ALTO	ALTO	MEDIO ALTO	0	13	0	0
ALTO	MEDIO ALTO	MEDIO BAJO	0	0	0	0
MEDIO ALTO	MEDIO BAJO	BAJO	0	6	0	0
MEDIO ALTO	ALTO	MEDIO BAJO	0	6	0	6
ALTO	MEDIO ALTO	ALTO	0	0	6	0
MEDIO ALTO	MEDIO ALTO	MEDIO BAJO	0	0	0	0
MEDIO ALTO	ALTO	MEDIO ALTO	0	0	0	0
ALTO	MEDIO ALTO	MEDIO BAJO	0	6	0	0
MEDIO BAJO	MEDIO ALTO	MEDIO BAJO	0	13	0	6

Fuente: Orientación Vocacional.

7.2.2. Registro y Estadística

Los resultados de los cuestionarios socioeconómicos fueron entregados en formato csv, de los periodos 2016, 2017 y 2018. Por razones internas del departamento, estos son los únicos periodos disponibles.

El total de registros recolectados para la investigación son 101 047 en el 2016; 85 826 en el 2017 y 123 160 en el 2018, para un total de 310 033, cada uno compuesto por 175 respuestas en cada registro. Lo anterior resulta en 54 millones de datos aproximadamente durante los 3 años.

Figura 30. **Muestra de datos del cuestionario socioeconómico del año 2017**

Personas con quien vive	Estado Civil	La vivienda que habita es:	¿Cómo sostiene sus estudios universitarios?
Padre y/o madre y hermanos	Soltero	De los padres	Ayuda de padre o madre
Sólo madre	Soltero	Propia	Ayuda de padre o madre
Padre y/o madre y hermanos	Soltero	De los padres	Ayuda de padre o madre
Padre y/o madre y hermanos	Soltero	Propia	Ayuda de padre o madre
Padre y/o madre y hermanos	Soltero	De los padres	Ayuda de padre o madre
Padre y/o madre y hermanos	Soltero	De los padres	Ayuda de padre o madre
Sólo madre	Soltero	De los padres	Ingresos propios
Padre y/o madre, hermanos y otros familiares	Soltero	De los padres	Ayuda de padre o madre
Padre y/o madre y hermanos	Soltero	Propia	Ayuda de padre o madre
Cónyuge e hijos	Soltero	De los padres	Otro
Padre y/o madre, hermanos y otros familiares	Soltero	De los padres	Ingresos propios
Padre y/o madre, hermanos y otros familiares	Soltero	De los padres	Ingresos propios
Padre y/o madre y hermanos	Soltero	De los padres	Ayuda de padre o madre
Cónyuge, hijos y otros familiares	Casado	De los padres	Ingresos propios
Padre y/o madre, hermanos y otros familiares	Soltero	Propia	Ingresos propios
Padre y/o madre, hermanos y otros familiares	Soltero	De familiares	Ayuda de padre o madre
Padre y/o madre y hermanos	Soltero	De los padres	Ingresos propios
Padre y/o madre y hermanos	Soltero	Propia	Ingresos propios
Padre y/o madre, hermanos y otros familiares	Soltero	Alquilada	Ayuda de padre o madre
Cónyuge e hijos	Casado	Propia	Ingresos propios
Cónyuge, hijos y otros familiares	Casado	De los padres	Ingresos propios
Cónyuge e hijos	Casado	Propia	Ingresos propios
Padre y/o madre y hermanos	Soltero	De los padres	Ingresos propios
Padre y/o madre, hermanos y otros familiares	Soltero	De los padres	Ayuda de padre o madre

Fuente: Registro y Estadística.

7.2.3. Centro de Cálculo

La información académica de los estudiantes fue entregada en formato Microsoft Excel para el análisis de la investigación, con los registros desde el año 2008 al 2017. El total fue de 3 389 estudiantes de Ingeniería en Ciencias y Sistemas, cada uno con los cursos aprobados, créditos, promedio, cursos desasignados, cursos repetidos, para un total de 16 945 datos en 10 años.

Figura 31. **Muestra de datos de información académica de estudiantes con carné del año 2012, de la escuela de Ciencias y Sistemas de la Facultad de Ingeniería**

CURSOS APROBADOS	CRÉDITOS	PROMEDIO	CURSOS DESASIGNADOS	CURSOS REPETIDOS
69	274	78	0	11
70	265	77	2	8
68	266	78	1	5
67	264	77	1	3
64	253	83	0	1
68	265	76	1	6
64	253	86	0	10
67	261	78	2	6
65	255	78	0	5
66	260	76	1	10
64	254	78	1	5
65	263	75	0	8
68	271	73	0	23
68	262	71	0	11
67	261	74	0	17
68	257	72	0	12
65	253	75	0	7
65	259	74	0	11
67	259	73	1	16
62	248	81	0	12
65	259	76	3	18
66	257	73	0	14
66	256	72	0	11
63	256	75	1	11
66	259	72	0	17
65	252	73	0	11
64	254	74	2	13
62	251	76	1	10

Fuente: Centro de Cálculo.

7.3. Dificultades para obtener los datos en las dependencias

Durante la investigación de campo realizada en las dependencias se encontraron dificultades, diferentes en cada una, debido a que actualmente no se está llevando a cabo ningún proceso estandarizado para la recolección de datos.

7.3.1. Evaluación de orientación vocacional de Bienestar Estudiantil

La dificultad para obtener los datos de evaluación vocacional fue el proceso de consolidación de los resultados de las evaluaciones, debido a la estructura de la prueba.

Según explicó el encargado de la consolidación y responsable de los datos, la evaluación consta de tres secciones encargadas de generar los resultados. El proceso de consolidación determina el nivel de la habilidad según las respuestas. El área de intereses profesionales determina las carreras más afines para el aspirante, con base en las respuestas dadas.

Las ponderaciones internas para cada área toma un tiempo considerable de cómputo para el sistema encargado generar los resultados.

7.3.2. Cuestionario socioeconómico de Registro y Estadística

Una de las dificultades encontradas en el proceso de recolección es que aún no se han estandarizado las preguntas del cuestionario, dado que cada año se agregan y remueven preguntas, con base en el criterio de los analistas.

Tomando en cuenta esta dificultad, la base de datos descrita en el capítulo 6 está diseñada para ser variable respecto a agregar o quitar datos de las fuentes, con una estructura de xml definida en el apéndice 5, para almacenar diferentes variantes sin importar la dependencia o si los campos de datos de una misma dependencia cambian a lo largo del tiempo.

7.3.3. Información académica de estudiantes en Centro de Cálculo

En este centro es muy limitado el acceso a la información académica de los estudiantes, dado que tienen información sensible de los mismos.

Por lo tanto, es importante asegurar el proceso de anonimización de la información explicado en el subcapítulo 6.3 y basarse en un marco legal robusto que la respalde, tomando en cuenta la base expuesta en el capítulo 4, para salvaguardar la privacidad de los estudiantes.

7.4. Proceso aplicado a las fuentes de datos

Para obtener el archivo final con los datos de las tres fuentes se realizó manualmente un proceso de tres pasos. El plan propuesto permite automatizar dicho proceso al aplicar la arquitectura del sistema diseñado.

- Anonimización de los datos de las fuentes.

Luego de obtener los archivos de las dependencias se realizó un proceso de anonimización de los datos sensibles.

Todos los archivos fueron convertidos a formato csv. En cada uno se ubicó las columnas con datos sensibles y se aplicó una función hash para sustituirlos.

- Evaluación de la información académica de los estudiantes.

Tal como se plantea en el capítulo 3, debe existir un parámetro que indique si el estudiante es exitoso o no en el desempeño académico.

La fórmula de la evaluación aplicada se explica en el capítulo 6. Dicha evaluación fue efectuada sobre los datos académicos de cada estudiante. Este es el indicador utilizado como variable dependiente.

- Relación y concatenación de los datos de las fuentes

Al tener los archivos estandarizados en formato csv y con los datos anonimizados, se procedió a concatenar los registros de cada archivo de las dependencias. El registro académico y el número de orientación vocacional fueron utilizados como llave de las relaciones.

Los datos de Bienestar Estudiantil y de Registro y Estadística fueron utilizados como variables independientes; es decir, los factores evaluados por el análisis.

El archivo final fue subido a un repositorio de Google Drive con la aplicación Google Sheets. Luego, con la url que otorga la aplicación, fue cargada en la herramienta Orange.

7.5. Proceso realizado en la herramienta de minería de datos

Para la carga de datos masiva en la herramienta se utilizó el objeto *File;* luego se clasificó las variables independientes con rol atributo y tipo según el dominio del dato, y la variable dependiente con rol objetivo con tipo de dato numérico. Luego se vinculó el objeto *File* hacia el objeto *Rank.* Para este análisis no se implementó ningún modelo de predicción para alimentar el proceso de clasificación; simplemente se utilizó el algoritmo *Relieff*, nativo de la herramienta, para obtener el indicador.

7.6. Resultados obtenidos

Con el indicador obtenido en el paso anterior, se genera un ranking de mayor a menor puntuación. Según la teoría del algoritmo, un mayor índice indica una mayor correlación de la variable independiente respecto de la variable dependiente.

El análisis descrito fue ejecutado sobre los datos de la escuela de Ciencias y Sistemas de la Facultad de Ingeniería de la Universidad de San Carlos de Guatemala para los estudiantes con registro académico con año 2012 y 2013, con datos del formulario de Registro y Estadística del año 2017, y las pruebas vocacionales de los números de orientación vocacional correspondientes a cada alumno.

Según el proceso ejecutado se obtuvo que los 5 factores que más correlación tienen con el éxito de cierre de pensum de los estudiantes de Ingeniería en Ciencias y Sistemas, de mayor a menor importancia, son:

- ¿La vivienda que habita es? (pregunta No. 7 de cuestionario
- socioeconómico 2017)
- Apreciación abstracta (cuartil del examen de orientación vocacional)
- Apreciación verbal (cuartil del examen de orientación vocacional)
- ¿Personas con quien vive? (pregunta No. 5 de cuestionario
- socioeconómico 2017)
- Apreciación numérica (cuartil del examen de orientación vocacional)

CONCLUSIONES

1. La Universidad de San Carlos de Guatemala es de gran importancia para la sociedad guatemalteca, debido a las contribuciones históricas en los procesos políticos, sociales y democráticos desde su fundación. Además, es la universidad que alberga mayor cantidad de estudiantes.

2. La deserción de los estudiantes de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería va desde el 65 % para la generación del 2008 a 2014 hasta el 80 % para la generación del 2011 a 2017, lo cual evidencia la problemática.

3. Las dependencias clave en el proceso de ingreso a lo largo de la carrera son Bienestar Estudiantil, Registro y Estadística y Centro de Cálculo. Además, son las dependencias que más información generan.

4. Se propuso un plan de implementación considerando el marco legal nacional e internacional, y dejó en evidencia la importancia de la privacidad de los estudiantes.

5. Se demostró la factibilidad del plan al implementar la herramienta Orange en los datos recolectados durante la investigación. Hay cinco factores con la mayor correlación respecto al éxito de un estudiante, tal como define esta investigación.

RECOMENDACIONES

1. Implementar el plan propuesto para la Escuela de Ciencias y Sistemas en el resto de unidades académicas de la Universidad de San Carlos de Guatemala con altos índices de deserción; ajustar distintos parámetros e incluir las fuentes acorde al contexto.

2. Al momento de implementar el plan en cualquier unidad académica, actualizar el marco legal acorde a la ley, tanto en el ámbito nacional como internacional.

3. Incrementar la cantidad de fuentes de datos para la investigación, tales como el examen de salud de la Unidad de Salud de Bienestar Estudiantil, resultados de las pruebas básicas realizadas por el Sistema de Ubicación y Nivelación y las pruebas específicas a cargo de cada unidad académica, con el objetivo de aumentar la precisión y confiabilidad de los resultados de los análisis estadísticos.

BIBLIOGRAFÍA

1. ABC. *ABC redes*. [en línea]. <https://www.abc.es/tecnologia/redes/abci-facebook-tendra-pagar-multa-565000-euros-escandalo-cambridge-analytica-201810251951_noticia.html>. [Consulta: 3 de mayo de 2018].

2. ALUJA, Tomás. *La minería de datos, entre la estadística y la inteligencia artificial.* [en línea]. <http://www.raco.cat/index.php/Questiio/article/viewFile/27009/26843>. [Consulta: 3 de mayo de 2018].

3. BOWCOTT, Owen, & HERN, Alex. *The Guardian*. [en línea]. <https://www.theguardian.com/news/2018/apr/10/cambridge-analytica-and-facebook-face-class-action-lawsuit>. [Consulta: 11 de mayo de 2018].

4. CASALBONI, Alex. *Cloud Academy Blog*. [en línea]. <https://cloudacademy.com/blog/aws-machine-learning/>. [Consulta: 11 de mayo de 2018].

5. Comisión Europea. *Comisión Europea*. [en línea]. <https://ec.europa.eu/commission/priorities/justice-and-fundamental-rights/data-protection/2018-reform-eu-data-protection-rules_es/>. [Consulta: 19 de mayo de 2018].

6. Corte de Constitucionalidad. *Corte de Constitucionalidad.* [en línea]. <http://143.208.58.124/Sentencias/826866.3552-2014.pdf/>. [Consulta: 11 de mayo de 2018].

7. FIUSAC. *Facultad de ingenieria USAC*. [en línea]. <https://portal.ingenieria.usac.edu.gt/>. [Consulta: 19 de mayo de 2018].

8. HAND, David, MANNILA, Heikki y SMYTH, Padhraic. *Priciples of Data Mining.* Londres: Massachusetts Institute of Technology, 2001. 322 p.

9. IBRAHIM, Zaidah y RUSLI, Daliela. *researchgate.* [en línea]. <https://www.researchgate.net/profile/Daliela_Rusli/publication/228894873_Predicting_Students'_Academic_Performance_Comparing_Artificial_Neural_Network_Decision_Tree_and_Linear_Regression/links/0deec51bb04e76ed93000000/Predicting-Students-Academic-Performa>. [Consulta: 19 de mayo de 2018].

10. *KDNuggets*. [en línea]. <https://www.kdnuggets.com/software/suites.html>. [Consulta: 3 de mayo de 2018].

11. Kovačić, Z. *citeseerx.* [en línea]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.381.7571&rep=rep1&type=pdf>. [Consulta: 3 de mayo de 2018].

12. KRUCHTEN, Philippe. *IEEE Software.* [en línea]. <http://www.inf.ed.ac.uk/teaching/courses/seoc/2006_2007/resources/Mod_5ViewModel.pdf>. [Consulta: 3 de mayo de 2018].

13. KUH, George; BRIDGES, Brian, y HAYEK, John. *National Postsecondary Education Cooperative.* [en línea]. <https://www.ue.ucsc.edu/sites/default/files/WhatMattersStudentSuccess(Kuh,July2006).pdf>. [Consulta: 3 de mayo de 2018].

14. LEWIS-KRAUS, Gidon. Going Neural. [en línea]. <http://www.gideonlk.com/writing>. [Consulta: 3 de mayo de 2018].

15. MARTÍNEZ, Aida, C. *Usac Virtual Files.* [en línea]. <https://usacvirtual.files.wordpress.com/2013/08/situacion-de-la-educacion-superior-en-guatemala.pdf>. [Consulta: 3 de mayo de 2018].

16. Ministerio de Salud Colombia. *MinSalud.* [en línea]. <https://www.minsalud.gov.co/sites/rid/Lists/BibliotecaDigital/RIDE/VS/ED/GCFI / lineamientos - anonimizacion - sistema - encuestas.pdf>. [Consulta: 19 de mayo de 2018].

17. MUSTAFA, Alí y GAUSSIER, Eric. *Semantics Scholar.* [en línea]. <https://pdfs.semanticscholar.org/3536/df7b196857f1959d033ebc09c19b064c9616.pdf>. [Consulta: 25 de mayo de 2018].

18. Orange. *Un taller práctico en el taller de Genómica Funcional, Liubliana, Eslovenia*. [en línea]. <https://orange.biolab.si/docs/>. [Consulta: 19 de mayo de 2018].

19. Parlamento Europeo. Reglamento (UE) 2016/679 DEL PARLAMENTO EUROPEO. *Diario Oficial de las Comunidades Europeas*, pág. 3.

20. Periodico USAC. *Soy USAC*. [en línea]. <http://soy.usac.edu.gt:http://soy.usac.edu.gt/?p=3370>. [Consulta: 25 de mayo de 2018].

21. Rapid Miner, Inc. *Plataforma de extremo a extremo totalmente transparente*. [en línea]. <https://rapidminer.com/products/>. [Consulta: 3 de mayo de 2018].

22. Real Academia Española. *Diccionario de la lengua española*. [en línea]. <http://www.rae.es/>. [Consulta: 3 de mayo de 2018].

23. Registro y EstadÍstica USAC. *Departamento de Registro y Estadistica*. [en línea]. <http://rye.usac.edu.gt/wiki/index.php/Referencia_Hist%C3%B3rica_del_Departamento>. [Consulta: 11 de mayo de 2018].

24. _____. [en línea]. <https://registro.usac.edu.gt/formularios_rye/AvanceEstad01_2018.pdf>. [Consulta: 11 de mayo de 2018].

25. ROJAS, Alex. *Prensa Libre*. [en línea]. <https://www.prensalibre.com/guatemala/politica/cc-falla-contra-comercializadoras-de-datos-personales/>. [Consulta: 19 de mayo de 2018].

26. RUSSELL, Stuard, y NORVIG, Peter. *Inteligencia Artificial Un Enfoque Moderno.* Madrid: Pearson Educación, S.A., 2004. 1241 p.

27. SAGASTUME, M. *USAC.* [en línea]. <https://www.usac.edu.gt/g/Sintesis_Historica_edicion_2013.pdf>. [Consulta: 3 de mayo de 2018].

28. SHMUELI, Galit; PATEL, Nitin y BRUCE, Peter. *Data Mining for Business Intelligence: Concepts, Techniques and Applications in Microsoft Office Ecel with XLMiner.* [en línea]. <Wiley-Interscience>. [Consulta: 25 de mayo de 2018].

29. USAC. *Bienestar Estudiantil USAC*. [en línea]. <https://vocacional.usac.edu.gt/conte_acercadeov.php>. [Consulta: 25 de mayo de 2018].

30. _____. *Departamento de Presupuestos*. [en línea]. <http://presupuesto.usac.edu.gt/wp-content/uploads/2019/01/Punto-CUARTO-inciso-4.3-Acta-36-2018-aprobaci%C3%B3n-Presupuesto-de-Ingresos-y-Egresos-2019.pdf>. [Consulta: 3 de mayo de 2018].

31. _____. *DIGED USAC*. [en línea]. <https://diged.usac.edu.gt/sun/>. [Consulta: 258 de mayo de 2018].

32. WITTEN, Ian y FRANK, Ernest. *The WEKA Workbench.* [en línea]. <https://www.cs.waikato.ac.nz>. [Consulta: 11 de mayo de 2018].

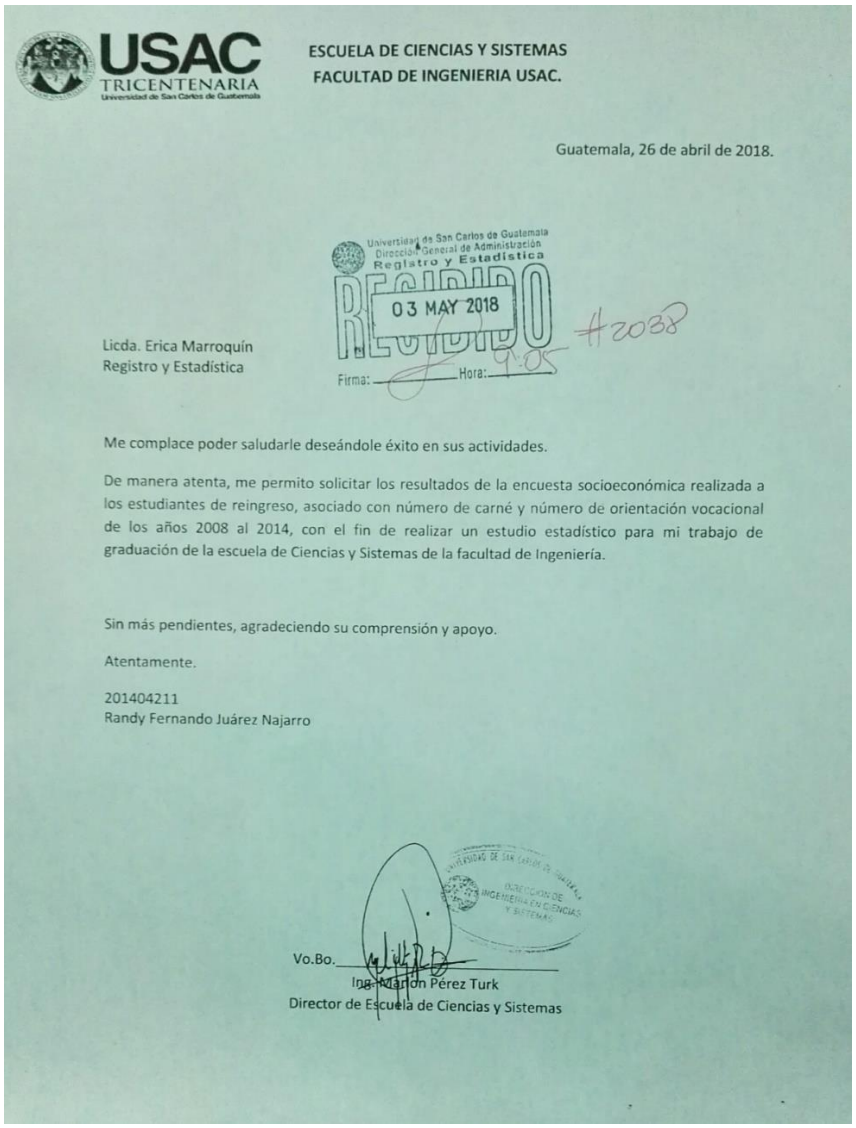
33. WU, X., KUMAR, V., & QUINLAN, R. *Citeserx.* [en línea]. <http://citeseerx.ist.psu.edu/viewdoc/download?doi=10.1.1.587.2765&rep=rep1&type=pdf>. [Consulta: 19 de mayo de 2018].

34. YU Ho, Chong. *Research Gate.* [en línea]. <https://www.researchgate.net/profile/Chong_Ho_Yu/publication/228684382_A_Data_Mining_Approach_for_Identifying_Predictors_of_Student_Retention_from_Sophomore_to_Junior_Year/links/55810ecf08aed40dd8cd39d5/A-Data-Mining-Approach-for-Identifying-Predictors-of>. [Consulta: 25 de mayo de 2018].

35. YUDELSON, M., MEDVEDEVA, O., & LEGOWSKI, E. *Semantics Scholar.* [en línea]. <https://pdfs.semanticscholar.org/6a2b/28215e55d9480d236e543ca8db5ec0d2c809.pdf>. [Consulta: 3 de mayo de 2018].

APÉNDICES

Apéndice 1. **Carta de solicitud de respuestas de cuestionario dirigida a Registro y Estadística**



USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

ESCUELA DE CIENCIAS Y SISTEMAS
FACULTAD DE INGENIERIA USAC.

Guatemala, 26 de abril de 2018.

Universidad de San Carlos de Guatemala
Dirección General de Administración
Registro y Estadística
RECIBIDO
03 MAY 2018
#2038
Firma: Hora: 9:05

Licda. Erica Marroquín
Registro y Estadística

Me complace poder saludarle deseándole éxito en sus actividades.

De manera atenta, me permito solicitar los resultados de la encuesta socioeconómica realizada a los estudiantes de reingreso, asociado con número de carné y número de orientación vocacional de los años 2008 al 2014, con el fin de realizar un estudio estadístico para mi trabajo de graduación de la escuela de Ciencias y Sistemas de la facultad de Ingeniería.

Sin más pendientes, agradeciendo su comprensión y apoyo.

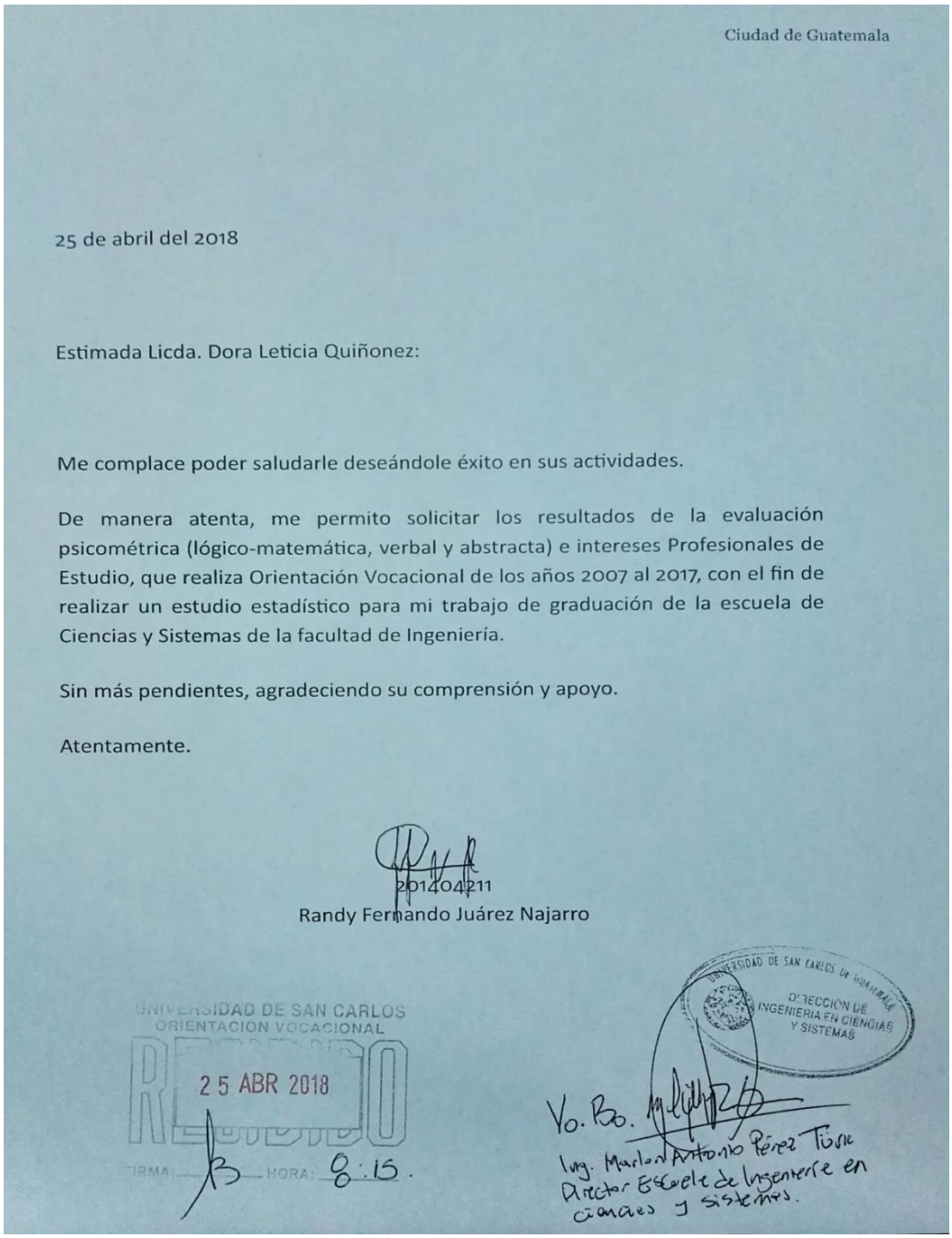
Atentamente.

201404211
Randy Fernando Juárez Najarro

Vo.Bo.
Ing. Marlon Pérez Turk
Director de Escuela de Ciencias y Sistemas

Fuente: elaboración propia.

Apéndice 2. **Carta de solicitud de resultados de examen vocacional dirigida a Bienestar Estudiantil**



Ciudad de Guatemala

25 de abril del 2018

Estimada Licda. Dora Leticia Quiñonez:

Me complace poder saludarle deseándole éxito en sus actividades.

De manera atenta, me permito solicitar los resultados de la evaluación psicométrica (lógico-matemática, verbal y abstracta) e intereses Profesionales de Estudio, que realiza Orientación Vocacional de los años 2007 al 2017, con el fin de realizar un estudio estadístico para mi trabajo de graduación de la escuela de Ciencias y Sistemas de la facultad de Ingeniería.

Sin más pendientes, agradeciendo su comprensión y apoyo.

Atentamente.

201404211
Randy Fernando Juárez Najarro

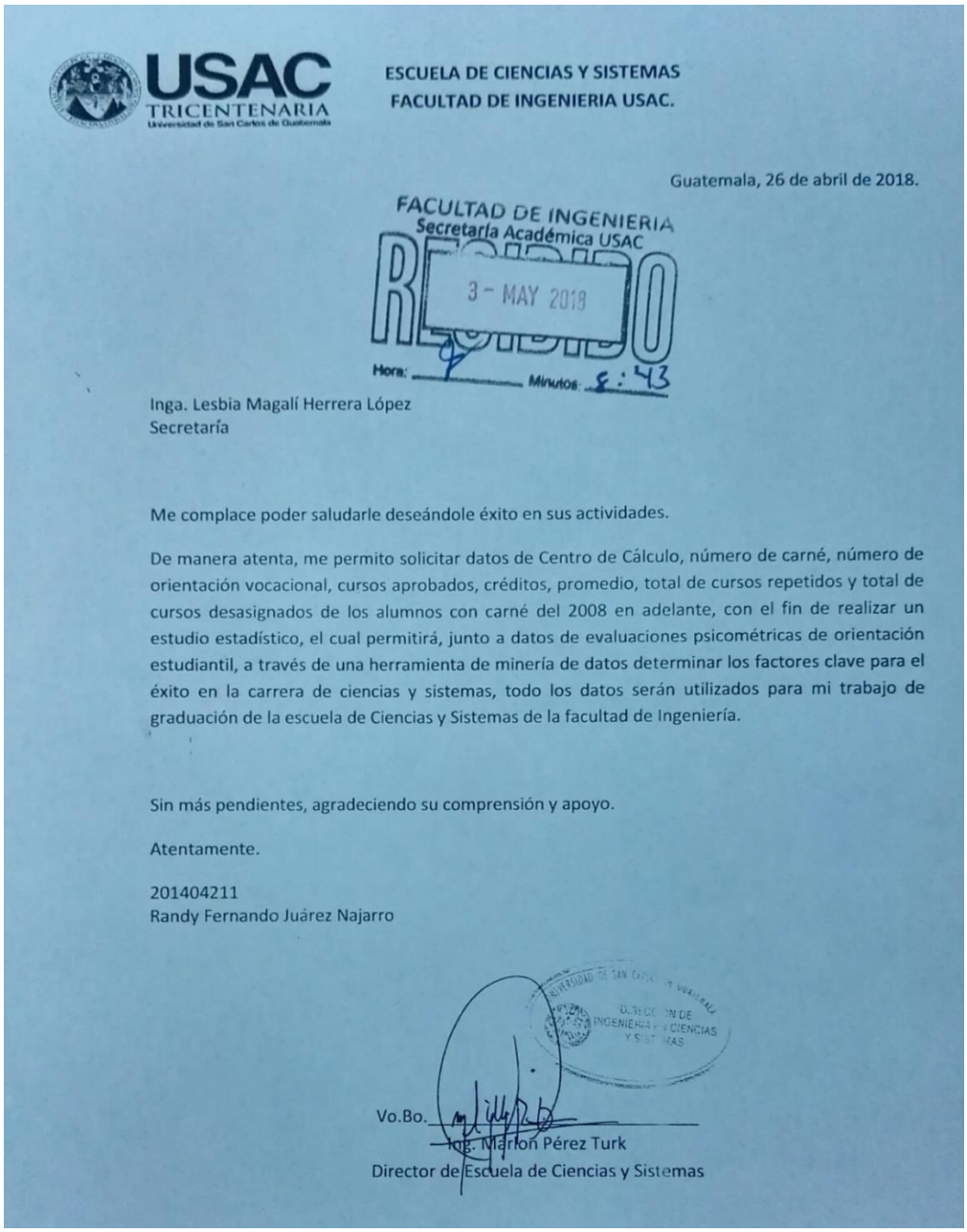
UNIVERSIDAD DE SAN CARLOS
ORIENTACION VOCACIONAL
RECIBIDO
25 ABR 2018
FIRMA: HORA: 8:15.

UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
DIRECCIÓN DE INGENIERÍA EN CIENCIAS Y SISTEMAS

Vo. Bo.
Ing. Marlon Antonio Pérez Türk
Director Escuela de Ingeniería en Ciencias y Sistemas.

Fuente: elaboración propia.

Apéndice 3. **Carta de solicitud de datos académicos dirigida a Secretaría Académica**



USAC
TRICENTENARIA
Universidad de San Carlos de Guatemala

ESCUELA DE CIENCIAS Y SISTEMAS
FACULTAD DE INGENIERIA USAC.

Guatemala, 26 de abril de 2018.

FACULTAD DE INGENIERIA
Secretaría Académica USAC
RECIBIDO
3 - MAY 2018
Hora: 9 Minutos: 8:43

Inga. Lesbia Magalí Herrera López
Secretaría

Me complace poder saludarle deseándole éxito en sus actividades.

De manera atenta, me permito solicitar datos de Centro de Cálculo, número de carné, número de orientación vocacional, cursos aprobados, créditos, promedio, total de cursos repetidos y total de cursos desasignados de los alumnos con carné del 2008 en adelante, con el fin de realizar un estudio estadístico, el cual permitirá, junto a datos de evaluaciones psicométricas de orientación estudiantil, a través de una herramienta de minería de datos determinar los factores clave para el éxito en la carrera de ciencias y sistemas, todo los datos serán utilizados para mi trabajo de graduación de la escuela de Ciencias y Sistemas de la facultad de Ingeniería.

Sin más pendientes, agradeciendo su comprensión y apoyo.

Atentamente.

201404211
Randy Fernando Juárez Najarro

Vo.Bo.
Ing. Marlon Pérez Turk
Director de Escuela de Ciencias y Sistemas

Fuente: elaboración propia.

Apéndice 4. **Estructura de archivo de las dependencias en formato csv para caga de datos**

Identificador de estudiante	Dato 1	Dato 2	Dato n
201700000	Alto	Bajo	Medio
201700001	Medio	Medio	Alto
201700002	Bajo	Bajo	Medio
201700003	Alto	Alto	Alto
201700004	Bajo	Bajo	Alto
201700005	Bajo	Alto	Medio
201700006	Bajo	Bajo	Alto

Fuente: elaboración propia.

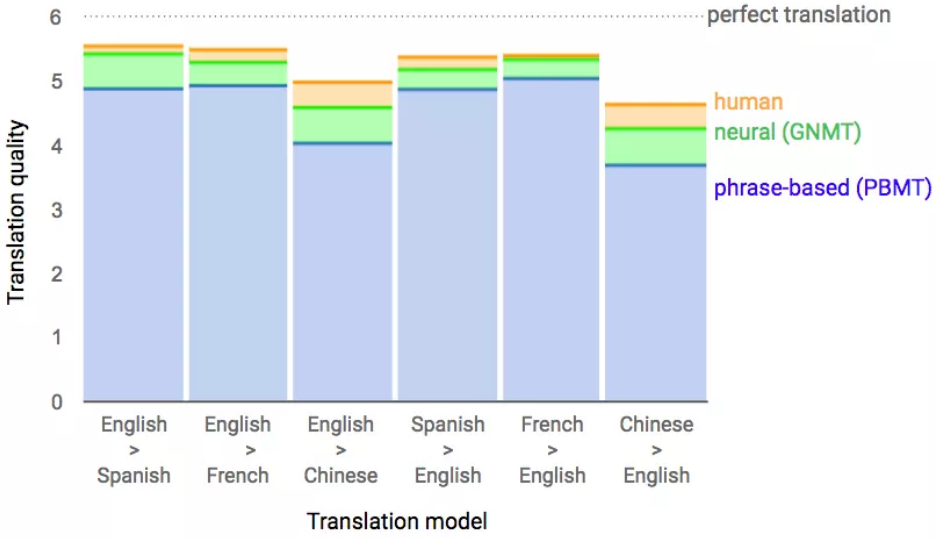
Apéndice 5. **Formato definido para xml de almacenamiento de respuestas de dependencias**

```
<?xml version="1.0" encoding="UTF-8"?>
<datos>
        <dato pregunta="¿Cómo sostiene sus estudios universitarios?" repuesta="INGRESOS PROPIOS">
        <dato pregunta="¿Cuántas personas dependen económicamente de usted?" repuesta="2">
        <dato pregunta="¿Qué número de horas trabaja por día?" repuesta="8">
        <dato pregunta="¿El ingreso que percibe, sirve para su sostenimiento?" repuesta="SI">
        <dato pregunta="¿En qué jornada estudia?" repuesta="NOCTURNA">
        <dato pregunta="¿Qué medio utiliza para trasladarse a la Universidad?" repuesta="TRANSMETRO">
        <dato pregunta="¿Trabaja actualmente?" repuesta="SI">
</datos>
```

Fuente: elaboración propia.

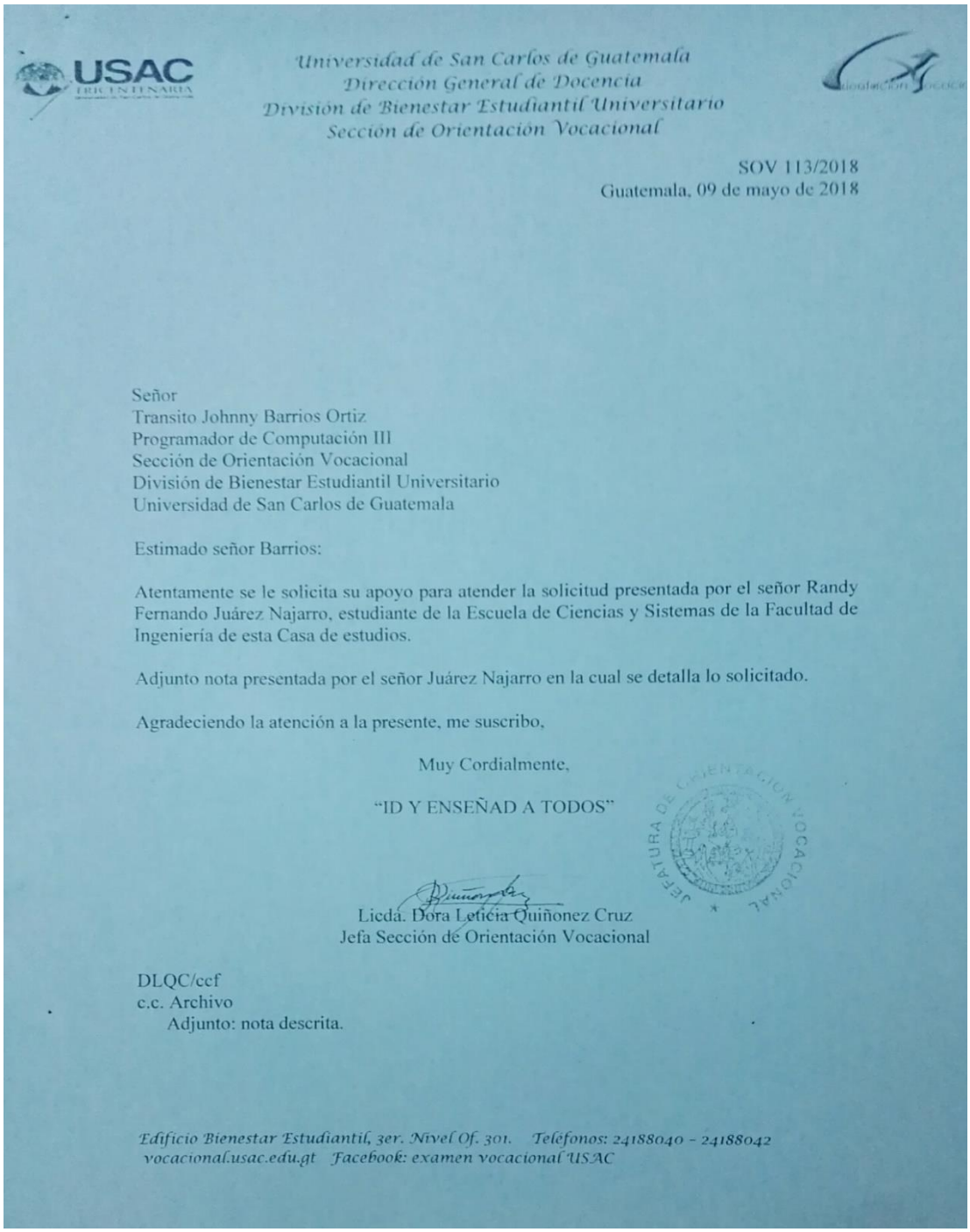
ANEXOS

Anexo 1. **Comparación tipos de traducción**



Fuente: ZDNet Magazine. *Comparación tipos de traducción*. zdnet2.cbsistatic.com. Consulta: marzo de 2018.

Anexo 2. **Carta de aprobación de Bienestar Estudiantil**



Universidad de San Carlos de Guatemala
Dirección General de Docencia
División de Bienestar Estudiantil Universitario
Sección de Orientación Vocacional

SOV 113/2018
Guatemala, 09 de mayo de 2018

Señor
Transito Johnny Barrios Ortiz
Programador de Computación III
Sección de Orientación Vocacional
División de Bienestar Estudiantil Universitario
Universidad de San Carlos de Guatemala

Estimado señor Barrios:

Atentamente se le solicita su apoyo para atender la solicitud presentada por el señor Randy Fernando Juárez Najarro, estudiante de la Escuela de Ciencias y Sistemas de la Facultad de Ingeniería de esta Casa de estudios.

Adjunto nota presentada por el señor Juárez Najarro en la cual se detalla lo solicitado.

Agradeciendo la atención a la presente, me suscribo,

Muy Cordialmente,

"ID Y ENSEÑAD A TODOS"

JEFATURA DE ORIENTACIÓN VOCACIONAL

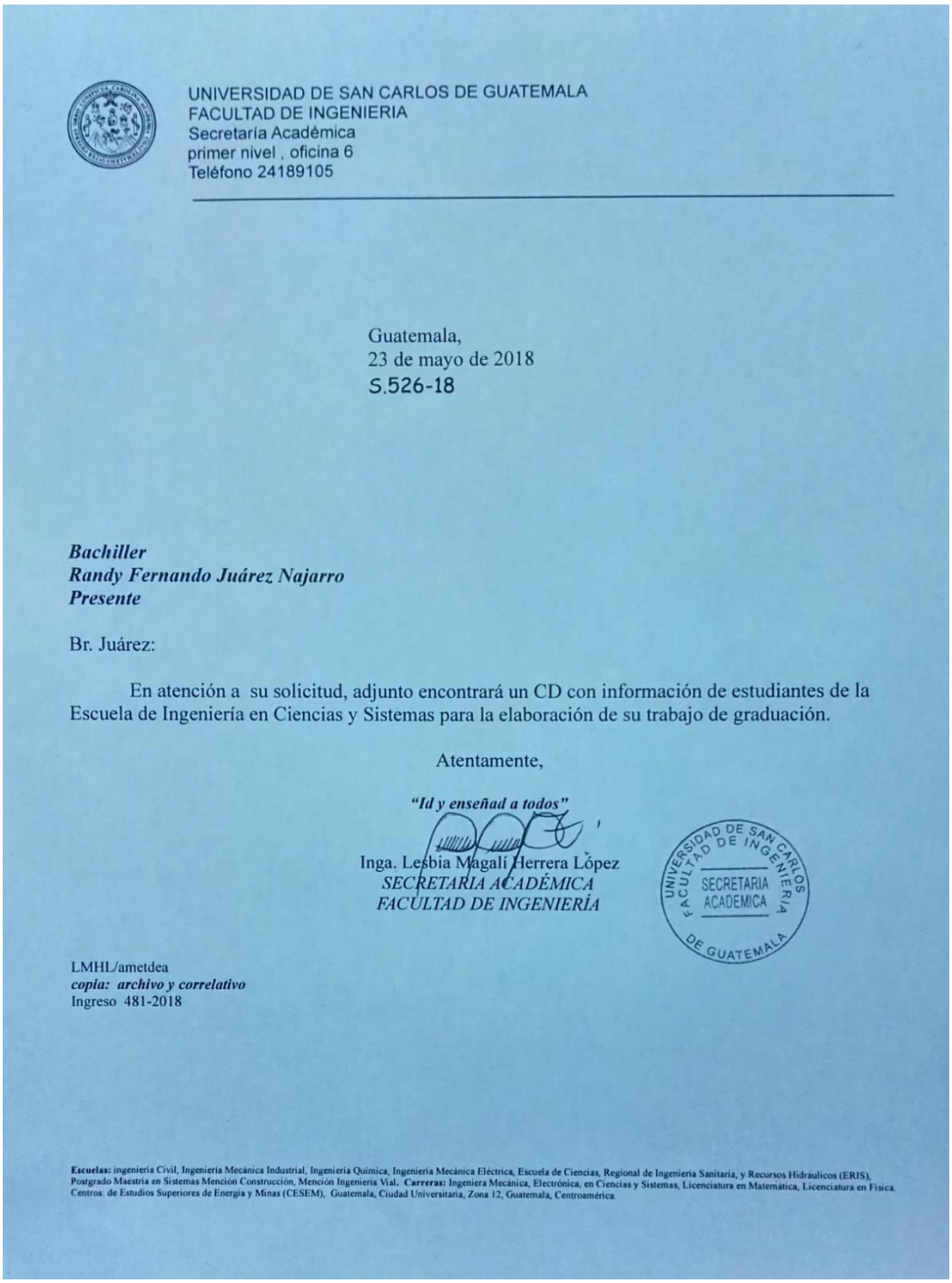
Licda. Dora Leticia Quiñonez Cruz
Jefa Sección de Orientación Vocacional

DLQC/ccf
c.c. Archivo
Adjunto: nota descrita.

Edificio Bienestar Estudiantil, 3er. Nivel Of. 301. Teléfonos: 24188040 - 24188042
vocacional.usac.edu.gt Facebook: examen vocacional USAC

Fuente: Jefatura Orientación Vocacional.

Anexo 3. **Carta de aprobación de Secretaría Académica**



UNIVERSIDAD DE SAN CARLOS DE GUATEMALA
FACULTAD DE INGENIERIA
Secretaría Académica
primer nivel , oficina 6
Teléfono 24189105

Guatemala,
23 de mayo de 2018
S.526-18

Bachiller
Randy Fernando Juárez Najarro
Presente

Br. Juárez:

En atención a su solicitud, adjunto encontrará un CD con información de estudiantes de la Escuela de Ingeniería en Ciencias y Sistemas para la elaboración de su trabajo de graduación.

Atentamente,

"Id y enseñad a todos"

Inga. Lesbia Magalí Herrera López
SECRETARIA ACADÉMICA
FACULTAD DE INGENIERÍA

UNIVERSIDAD DE SAN CARLOS FACULTAD DE INGENIERIA SECRETARIA ACADEMICA DE GUATEMALA

LMHL/ametdea
copia: archivo y correlativo
Ingreso 481-2018

Escuelas: ingeniería Civil, Ingeniería Mecánica Industrial, Ingeniería Química, Ingeniería Mecánica Eléctrica, Escuela de Ciencias, Regional de Ingeniería Sanitaria, y Recursos Hidráulicos (ERIS), Postgrado Maestría en Sistemas Mención Construcción, Mención Ingeniería Vial. **Carreras:** Ingeniera Mecánica, Electrónica, en Ciencias y Sistemas, Licenciatura en Matemática, Licenciatura en Física. Centros: de Estudios Superiores de Energía y Minas (CESEM), Guatemala, Ciudad Universitaria, Zona 12, Guatemala, Centroamérica.

Fuente: Secretaría Académica.